The Effect of Charter Schools on Student Achievement:

A META-ANALYSIS OF THE LITERATURE

Julian R. Betts and Y. Emily Tang

NATIONAL CHARTER SCHOOL RESEARCH PROJECT
center on **reinventing** public education

UNIVERSITY OF WASHINGTON
BOTHELL

The Effect of Charter Schools on Student Achievement:

A META-ANALYSIS OF THE LITERATURE

OCTOBER 2011

Authors:

Julian R. Betts and Y. Emily Tang

Department of Economics, University of California, San Diego

National Charter School Research Project
Center on Reinventing Public Education
University of Washington Bothell
425 Pontius Ave N., Suite 410
Seattle, Washington 98109
www.ncsrp.org

UNIVERSITY OF WASHINGTON
BOTHELL

NATIONAL CHARTER SCHOOL RESEARCH PROJECT
center on **reinventing** public education

The National Charter School Research Project (NCSRP) brings rigor, evidence, and balance to the national charter school debate.

NCSRP seeks to facilitate the fair assessment of the value-added effects of U.S. charter schools and to provide the charter school and broader public education communities with research and information for ongoing improvement.

NCSRP:

- ✓ **Identifies high-priority research questions.**
- ✓ **Conducts and commissions original research to fill gaps in current knowledge or to illuminate existing debates.**
- ✓ **Helps policymakers and the general public interpret charter school research.**

We would like to thank our current and past funders for their generous support:

- **Anonymous**
- **Achelis & Bodman Foundations**
- **Annie E. Casey Foundation**
- **Daniels Fund**
- **Doris & Donald Fisher Fund**
- **Thomas B. Fordham Foundation**
- **Bill & Melinda Gates Foundation**
- **The Heinz Endowments**
- **Ewing Marion Kauffman Foundation**
- **Rodel Charitable Foundation**
- **U.S. Department of Education**
- **Walton Family Foundation**

CONTENTS

TABLES AND FIGURES

Tables

Figures

ABSTRACT

Charter schools are largely viewed as a major innovation in the public school landscape, as they receive more independence from state laws and regulations than do traditional public schools, and are therefore more able to experiment with alternative curricula, pedagogical methods, and different ways of hiring and training teachers. Unlike traditional public schools, charters may be shut down by their authorizers for poor performance. But how is charter school performance measured? What are the effects of charter schools on student achievement?

Assessing literature that uses either experimental (lottery) or student-level growth-based methods, this analysis infers the causal impact of attending a charter school on student performance. Focusing on math and reading scores, the authors find compelling evidence that charters under-perform traditional public schools in some locations, grades, and subjects, and out-perform traditional public schools in other locations, grades, and subjects. However, important exceptions include elementary school reading and middle school math and reading, where evidence suggests no negative effects of charter schools and, in some cases, evidence of positive effects. Meta-analytic methods are used to obtain overall estimates on the effect of charter schools on reading and math achievement. The authors find an overall effect size for elementary school reading and math of 0.02 and 0.05, respectively, and for middle school math of 0.055. Effects are not statistically meaningful for middle school reading and for high school math and reading. Studies that focus on urban areas tend to find larger effects than do studies that examine wider areas. Studies of KIPP charter middle schools suggest positive effects of 0.096 and 0.223 for reading and math respectively. New York City and Boston charter schools also appeared to deliver achievement gains larger than charter schools in most other locations. A lack of rigorous studies in many parts of the nation limits the ability to extrapolate.

.

ACKNOWLEDGMENTS

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

This research was funded by the Walton Family Foundation and the Ewing Marion Kauffman Foundation, and we thank them for their support. However, the contents of this publication are solely the responsibility of the grantee.

ABOUT THE AUTHORS

Julian R. Betts is professor and former chair of economics at the University of California, San Diego, where he is Executive Director of the San Diego Education Research Alliance (sandera.ucsd.edu). He is also a research associate at the National Bureau of Economic Research, and an adjunct fellow and a Bren fellow at the Public Policy Institute of California. He has written extensively on the link between student outcomes and measures of school spending, and he has studied the role that school choice, educational standards, accountability, and teacher qualifications play in student achievement. He has served on three National Academy of Sciences panels, the Charter School Achievement Consensus Panel of the National Charter School Research Project, and various advisory groups for the U.S. Department of Education. He is also principal investigator for the federally mandated National Evaluation of Magnet Schools. He holds a Ph.D. in economics from Queen's University, Kingston, Ontario, Canada and an M.Phil. in Economics from Oxford University.

Y. Emily Tang is a lecturer in economics at the University of California, San Diego. She is a co-author of "Madness in the Method? A Critical Analysis of Popular Methods of Estimating the Effect of Charter Schools on Student Achievement" (in T*aking Measure of Charter Schools: Better Assessments, Better Policymaking, Better Schools*, ed. Julian R. Betts and Paul Hill, 2010) and "Value Added and Experimental Studies of the Effect of Charter Schools on Student Achievements: A Literature Review" (in *Hopes, Fears, & Reality: A Balanced Look at American Charter Schools*, ed. Robin Lake, National Charter School Research Project, Center on Reinventing Public Education, 2008). She obtained her Ph.D. in Economics from UCSD in 2007.

INTRODUCTION

Charter schools are public schools that receive more independence from state laws and regulations than do traditional public schools. Unlike traditional public schools, however, charters can be shut down by their authorizers if they do not perform well. Many view charter schools as a major innovation in the public school landscape because they have more freedom to experiment with alternative curricula and pedagogical methods and different ways of hiring and training teachers.

In Betts and Tang (2008), we surveyed the literature on the effect of charter schools on achievement. In that paper, we focused on two approaches that the Charter School Achievement Consensus Panel (2006) argued were most likely to provide accurate estimates of the causal effect of attending a charter school on achievement. The first approach compares those who win and lose lotteries to attend a given charter school. The second approach uses one of several variations of value-added models. These models follow individual students over a given period of time and examine any improvement in test scores. This second approach is helpful because it takes into account a student's past academic history.

In our earlier paper we found that roughly two-thirds of studies available at the time did not use methods that could obtain estimates of the causal effect of attending a charter school on achievement. Often these studies took a single snap shot of average achievement and used this to compare schools, without taking into account differences in the background of students at different schools.

In spite of a large increase in the number of studies since the initial Betts and Tang (2008) review, it is still the case that the majority of charter school studies take snap shots of student achievement at one point in time, or compare successive cohorts of students in a given grade. Both of these approaches are likely to entail severe omitted variable biases, and thus we exclude them in this newer review.[1] Although the number of rigorous studies has increased in the last three years, the number that use the most rigorous methods is still small. For example, to date only eight papers have used the lottery approach, studying roughly 90 charter schools.

1. 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. Betts, Tang and Zau also attempt to replicate lottery-based evidence for one charter school in San Diego. They find that models that do not take into account students' past achievement produce estimates far off the mark, but that value-added models can approximate the lottery-based findings much more closely.

Betts and Tang (2008) found mixed results, with evidence that charter schools in some studies outperformed traditional public schools in terms of math and reading achievement and underperformed in other studies. The present paper finds similar results, however overall average effects are more strongly in favor of charter schools than in the earlier review.

An equally important finding is the heterogeneity in the estimated effects. In the current paper we emphasize this heterogeneity by adopting formal meta-analytic methods that allow for effect sizes to vary randomly across studies. We find that more than 90% of the variation we observe across studies likely represents true variations, rather than statistical noise.

This finding has important policy implications. If charter schools are intended to be hotbeds of educational innovation, then successes should be identified, studied further, and the replicable parts of those models should be copied in other settings. Conversely, charter schools that consistently underperform their traditional public school counterparts in terms of math and reading achievement will require interventions and support to help them improve, and in cases of persistent underperformance, probably should be closed. Many charter schools may fall in the middle, neither innovating successfully nor innovating and failing, but simply replicating quite closely the standard fare in traditional public schools.

With some exceptions, such as studies of KIPP schools, the empirical literature that we review does not provide estimates of the effects on achievement of individual charter schools. But we see enough variation in average effect sizes across studies of different geographic locations, and across grade spans, to infer that variations in effectiveness across individual charter schools may be quite high. If policymakers were routinely able to obtain rigorous evaluations at the level of individual schools, then the full promise of the charter school movement—as a generator of new ways of teaching—could begin to be realized.[2]

The organization of this paper is as follows. Section 2 outlines the various methods used as well as challenges we faced in putting these methods into practice. It concludes with a discussion of the main empirical methods used by the studies we include, and

2. An alternative and complementary rationale for charter schools is that they induce competition among schools for students, making all schools more effective. The literature on this separate question is not covered in this paper. However, see Betts (2009) for a review of this nascent literature. Betts concludes that there is some promising but as yet inconclusive evidence in favor of the theory that charters induce better education by promoting competition. Zimmer et.al (2009) also conduct tests for competitive effects in seven locations, and find no evidence for competitive effects in six of those seven. In Texas, they found very small positive competitive effects. Note that if such competitive effects exist, any study of the impact of attending a charter school is likely to underestimate the effect because the comparison group, even in lottery studies, will perform better due to the presence of charter schools.

the strengths and considerable problems that can arise in each method, including lottery-based analyses. Section 3 presents statistical evidence on whether there are any negative or positive effects of charter schools. Section 4 presents the results of a formal meta-analysis of effect sizes, which emphasizes tests of significance while also presenting the reader with a sense of the degree of variation in estimated effect sizes. Section 5 illustrates this heterogeneity using histograms of effect sizes and counts of studies finding significant and insignificant, positive and negative effects. Section 6 examines whether the method of analysis is related to the effect sizes estimated. Section 7 provides a brief overview of the small amount of literature that examines the relation between charter school attendance and student outcomes other than math and reading achievement. Finally, we present our conclusions in Section 8.

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 four approaches to summarize the results. First, we test whether we can reject two hypotheses: that the effects of charters are never positive, and that the effects of charters are never negative. In a second approach that Betts and Tang (2008) did not implement, we perform a formal meta-analysis of the estimates in the literature. This tests whether the average effect is zero, and also characterizes the degree to which the variation we see across studies is real variation in the effectiveness of charter schools as opposed to statistical noise. Third, 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 negative or positive), or negative and significant results. This fourth method is transparent and easy to understand. Researchers have rightly criticized this approach because it might wrongly interpret a large number of studies that find "no significant results," when in truth each study has limited statistical power, perhaps due to small sample size. However, as as we will show, charter school studies produce far more significant results than one would expect if small samples were biasing researchers towards 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

Appendix Table A.1 shows the set of papers that are used for at least one of our four research methods, along with information on the geographic location and time span of the study.

Several challenges present themselves. As was true in Betts and Tang (2008), there is still fairly narrow geographic coverage in the studies we review here, although geographic coverage has improved over the last few years. Because there is a risk of overstating the generalizability of results, in Section 4 we report not only an overall effect size, but also the

number of studies and the number of geographic locations underlying a given estimate. The number of studies is quite large for estimates such as the effects for all charter middle schools, but quite small for studies of specific student subgroups, such as the effects for African-American students attending charter elementary schools.

A second challenge is that studies vary in which grade spans they cover. We found by far the greatest number of estimated charter school effects have been produced at the middle school grade span, however another popular approach has been to present results that combine elementary and middle schools together. A third popular approach has been to estimate an overall charter school effect for a given geographic area using grades that include elementary, middle, and high schools. (We refer to this last approach as an "All grade span" study.) Studies focused on the elementary or high school levels were least common.

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 across studies in the true effect size. We assume that 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% 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. Section 5, which shows histograms of actual effect sizes, illustrates this point by showing unweighted results and results which weight in favor of studies with more student observations. These two methods produce somewhat different pictures, particularly in the elementary and high school levels where there are relatively fewer studies.

As reported in Betts and Tang (2008), in some cases we requested information on standard deviation of test scores (within grade) in order to translate results from diverse testing systems into effect sizes, and for the number of charter schools and charter school students included in the analysis. (Effect sizes express the impact of attending a charter school in terms of the proportion of a standard deviation by which a student's test score changes.) We found that many papers do not report the exact number of charter schools being studied or the sample of charter school students, and thus when we provide weighted histograms we instead weight by the number of observations only. Comparisons across papers would be far simpler if authors routinely included these statistics.

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.

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.

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. It seems likely that such schools outperform other charter schools that are less popular, thus the external validity of lottery-based studies may be quite low.

A second potential issue with lotteries is differential attrition among the lottery losers. For instance, suppose highly motivated parents who lose an admission lottery to kindergarten at a popular charter school opt for private school for their child. This would bias the results of the lottery analysis, potentially in favor of finding a positive "effect" of attending a charter school. However, it is straightforward to check for this potential problem.

A third and equally important issue is that lottery-based studies can produce two distinct estimates: "intent to treat" and the impact of "treatment on the treated." The former, intent to treat, refers to the causal impact of winning a lottery. If researchers check that lottery winners, on average, resemble lottery losers at the time of the lottery (to confirm that the lottery was conducted fairly), and that the aforementioned problem of differential attrition is not an issue, then lottery analysis will yield the causal effect on outcomes of winning a school choice 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. The estimated impact of treatment on the treated is usually bigger in absolute value than the corresponding estimate of intent to treat, due to dropout and substitution bias. In the presence of dropout and substitution bias, several strong assumptions must hold true for these estimates to be valid.

Dropout bias refers to the fact that not all students who win a school choice lottery will attend. Suppose that only one in ten students who win a school choice lottery actually enrolls. If lottery winners who do not choose to attend a charter school have zero change to their achievement, and the tenth of lottery winners who do attend the charter school experience a gain of 50 points, then our estimate of the impact of winning a lottery is only one tenth as big as the 50 point "impact of treatment on the treated." That is, the average gain of winning the lottery is only 0.9X0 + 0.1X50=5 points. In this case, the impact of treatment on the treated would be ten times as high as the estimated intent to treat.

Substitution bias, sometimes referred to as crossover bias, refers to a situation in which some of those who are lotteried out of a charter school nonetheless manage to find a substitute school choice program. If some in the control group actually receive treatment, then we must scale up the intent to treat estimate to obtain the impact of treatment on those who are actually treated.

There are two approaches to converting the intent to treat estimate into an estimate of the impact of treatment on the treated, both of which produce identical results. First, one can scale the intent to treat estimate by dividing it by *(b-a)*, where *b* is the proportion of lottery winners who attend and *a* is the proportion of lottery losers who find substitute treatment. Second, researchers can use an instrumental variable (IV) strategy. In this latter approach, a student outcome is regressed on an indicator for attendance at a charter school, which is then instrumented using indicators for whether the student won a school choice lottery.

These estimates of the impact of treatment on the treated are very useful from a policy standpoint, and they have the same goal as non-lottery methods in that they attempt to estimate the causal effects of actually attending a charter school.

However, only under two strong assumptions can we obtain an unbiased estimate of the impact of treatment on the treated. The first assumption is that the impact of treatment (winning a lottery to attend school *s*) is identical for all students, and that this holds for all schools in a choice program. The second is that the impact of treatment to attend a school of choice, designated as school *s*, is identical to the impact of attending any other school of choice *s*. These are very strong assumptions, but they are clearly necessary. For instance, if the impact of attending a school varies by student, then there is likely to be self-selection into a school of choice. The subsample of lottery winners who actually switch to the school and persist will be those who will get the most from the school, and our estimate of the impact of treatment on the treated will be too high. Similarly, if schools of choice are differentially effective, then we cannot simply scale up using the factor *a* in the denominator, as this makes sense only if the effect of attending other schools of choice is identical to the

effect of attending school of choice *s*. (For details, see section 5 of Heckman, Lalonde and Smith, 1999.) Finally, a third implicit assumption is that researchers can estimate the proportions *b* and *a* well. Estimating the parameter *a*, the proportion of students who lose the given lottery but nonetheless enter a school choice program, often proves difficult because of lack of information about the extent to which lottery losers find some alternative form of school choice.

Each lottery-based study must be judged on its own terms. For example, the *National Study of Charter Middle Schools* by Gleason et al. (2010) does a good job of detecting and controlling for substitution bias, finding and adjusting estimates for the fact that 6% of lottery losers later enroll in a given charter regardless of the results of the lottery. They do not control for substitution into other schools of choice. On the other hand, the external validity of this study—that is, its applicability to other charter schools at the middle school level—is probably quite low. Gleason et al. (2010) report 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.

Some lottery-based studies of charter schools present intent-to-treat estimates only (e.g., McClure et al., 2005), and others present only estimates of the impact of treatment on the treated (e.g., Hoxby, Murarka, and Kang, 2009). Beginning with intent-to-treat estimates, and then proceeding to estimates of the impact of treatment on the treated (subject to a discussion of the validity of the underlying assumptions required for the latter estimate), would be one modeling approach that researchers could follow.

Propensity Score Matching

The main weakness of non-lottery-based methods is that they typically compare students who attend and students who do not attend charter schools. 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. Two recent studies of charter schools belonging to the Knowledge is Power Program (KIPP) have used propensity score matching. (See Tuttle et al., 2010, and Woodworth et al., 2008.) These studies match charter school attendees with non-charter attendees who have similar estimated probabilities of attending a charter school. This approach is useful, but is subject to bias because the method cannot control for unobservable variables that might be related to both 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.

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. 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. 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 to the growth for the same student in years not enrolled in a charter school.

However, this method has its own issues. The two primary weaknesses of fixed-effect models stem from the fact that identification comes from students who switch between charter and traditional public schools. In elementary schools, many students start in charters and do not switch, so that 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. Zimmer et al. (pages 35-36, 2009) compare test-score gains of those who switch into or out of charter schools, and who therefore contribute to their fixed-effect estimates, with the gains of students who remain in a charter school for the entire period. They conclude that it is "unclear" whether external validity is an issue, but they do present evidence that in some locales and subject areas test-score gains of charter school students who did not switch in or out during the sample period were higher than the test-score gains of those who switched.

Zimmer et al. (2009) also highlight a similar problem of "reversibility" in most fixed-effect analyses. Fixed-effect analyses assume that charter school estimates can be estimated equally well from comparing past trajectory to current trajectory in charter school entrants (students who switch out of traditional schools into charter schools) as from comparing current trajectory to future trajectory in charter school leavers (students who switch out of charter schools and into traditional schools). Zimmer et al. (page 33, 2009) conduct

tests to examine whether the charter fixed-effect estimates change markedly when they exclude students who switch out. The authors could not confirm that the charter school effects are the same from those switching in as from those switching out. In this case, the threat is internal rather than external, because charter school effect estimates obtained for charter school entrants may not apply to charter school leavers and vice versa (both of whom comprise the sample study population). Because the effects may be different for entrants than for leavers, and because elementary school estimates are primarily derived from leavers, Zimmer et al. argue that elementary school fixed-effect estimates should especially be interpreted with caution.[3]

Second, fixed-effect models can control for unobserved heterogeneity among students only to the extent that the heterogeneity is fixed over time. 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.[4] 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. (page 33, 2009) test whether student trajectories, in the year preceding switches into charter schools, are significantly different from trajectories in earlier years in the locations in which they had sufficient data to do so. They find no evidence that pre-transfer dips may be biasing estimates in San Diego or 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.

In short, none of the methods utilized in the papers included in our meta-analysis is entirely reliable.

---

3. Given these concerns, Zimmer et al. (2009) argue that more attention should be paid to estimates derived from switchers into and out of non-primary charter schools only than to estimates derived from switchers into and out of all schools, including primary schools. They offer estimates that are derived from analysis after dropping schools that start in Kindergarten, which they note comprise a large portion of charter schools. The authors note that with the exception of one location, the estimates from the complete sample and the non-primary sample are similar. However, in two cases, positive and significant effects lose their significance (but remain positive), and in two cases, negative and significant effects lose their significance and are nearly zero. Out of 14 estimates of charter school effectiveness (math and reading results in seven locations), seven estimates were smaller in the non-primary sample, while six were larger in the non-primary sample. One was the same. We include their estimates that do not make the non-primary sample exclusion for comparability with other fixed-effect studies and because it is not clear how or whether the larger sample estimates are biased estimates of the samples studied.

4. 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.

TESTING WHETHER CHARTER SCHOOLS IN ANY STUDY UNDERPERFORM OR OUTPERFORM TRADITIONAL PUBLIC SCHOOLS

Even if some studies indicate a negative or positive effect of charter schools on student achievement, we must exercise caution. If the true effect is zero, because of random variations, we should expect some fraction of studies to purport to find non-zero effects that are "statistically significant."

Fortunately, a method exists to test whether *any* of the estimated effects across independent studies are truly positive or negative. Fisher's inverse Chi-squared test allows one to test the hypothesis that all the effects are zero or negative against the alternative, that at least some of the effects are positive.[5] A rejection of this hypothesis signals that at least one study in the sample truly does provide evidence of positive effects. Conversely, we can test the hypothesis that all the effects are zero or positive, against the alternative that at least some of the effects are negative. Rejection of the hypothesis in this instance would support the contention that at least some charter schools underperform traditional public schools.

In order for Fisher's method to be valid, the studies that are used in the test must be statistically independent. For example, in some calculations we excluded the Betts et al. (2005) study of San Diego schools because the time period studied for a given set of grade levels overlapped with that of Zimmer et al. (2003), which covered roughly the same period for a variety of California districts.

Table 1 shows the probability that charter school effects are negative/zero or positive/zero for various combinations of studies. The top row of both panels shows the results when we combined all studies, regardless of whether they studied elementary, middle, and high schools together, one of these three grade spans individually, or combinations such as elementary/middle schools. For both reading and math, the probability that there are no studies with positive charter effects is miniscule, below 0.0001. The same applies to the probability that there are no studies showing true negative effects. These results suggest that in some instances charter school students learn less than they would in traditional public schools, and that in other instances, charter school students learn more. Our analyses of the patterns of statistical significance and of effect sizes will echo this finding of heterogeneity across locations.

5. See for instance Chapter 3 of Hedges and Olkin (1985).

Table 1. Tests for Existence of Positive or Negative Effects of Charters Among All Studies

STUDIES THAT INCLUDE CHARTER SCHOOLS FROM THE GRADE SPANS:	NUMBER OF STUDIES (# STATE STUDIES/ DISTRICT(S)/ SCHOOL(S))	PROBABILITY OF NO POSITIVE EFFECTS	PROBABILITY OF NO NEGATIVE EFFECTS
READING			
Studies of All Grades (A) or largest grade span(s) if an all-grade study not available	34 (16/11/7)	<0.001	<0.001
E (Elementary)	6 (1/5/0)	<0.001	0.987
E, M, and E/M (Elementary, Middle, or Combined Elementary/Middle)	25 (10/8/7)	<0.001	<0.001
M (Middle)	10 (2/3/5)	<0.001	0.994
H (High School)	4 (2/2/0)	<0.001	<0.001
A (Studies that include all three grade spans)	13 (8/5/0)	<0.001	<0.001
MATH			
Studies of All Grades (A) or largest grade span(s) if an all-grade study not available	35 (17/11/7)	<0.001	<0.001
E (Elementary)	7 (2/5/0)	<0.001	<0.001
E, M, and E/M (Elementary, Middle, or Combined Elementary/Middle)	27 (12/8/7)	<0.001	<0.001
M (Middle)	11 (3/3/5)	<0.001	0.978
H (High School)	5 (3/2/0)	0.001	<0.001
A (Studies that include all three grade spans)	14 (9/5/0)	<0.001	<0.001

NOTES: The columns showing probabilities show the p-value, or probability, that there are either no positive effects or no negative effects.

The second row of the top panel of Table 1 strongly suggests that some elementary charter schools outperform in reading, and that no study has produced evidence that charter schools underperform. (More precisely, the probability of no negative effects is 98.7%. Conversely, the probability that none of the studies find a positive effect is less than 0.1%). For math, there is strong evidence that elementary charter schools both underperform and outperform, depending on the time and location, which vary across studies.

Many studies combine elementary and middle schools. To obtain an overall picture of performance in these grades, in the next line we add studies that include elementary and middle school students to studies that focus only on elementary or middle schools, and find that there is strong evidence of both negative and positive charter effects in both math and reading.

When we examine studies that focus on middle schools alone, there is ample evidence for positive but not negative effects on reading and math. In both cases the probability of no positive effects is less than 0.1% and the probability of no negative effects is 99.4% and 97.8% for reading and math respectively.

When we examine high school studies by themselves, we find evidence that charter schools both outperform and underperform relative to traditional public schools.

Overall, this analysis shows that the literature suggests that some charter schools outpace their traditional counterparts while other charter schools trail behind. Notably, there are three cases in which charters do not seem to underperform in any of the studies but outperform in some studies: elementary school reading and middle school math and reading. This of course is not the same as saying that no individual charter schools underperform in these subjects and grades, rather, that the studies taken as a whole support this conclusion based on the specific places and times studied.

META-ANALYSIS OF EFFECT SIZE

A convenient and meaningful way to report results is as effect sizes; that is, the number of standard deviations that attending a charter school is predicted to move test scores. We used individual studies' effect-size estimates or converted them into effect sizes by dividing by the standard deviation of test scores in the given grade as reported by the study. Thus, an effect size of 0.1 indicates that a student's test score rises by one tenth of a standard deviation relative to the comparison population if the student attends a charter school for one year.

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.[6]

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:

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

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 toward 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.

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

Table 2. Effect Sizes and Significance from Meta-Analysis, by Grade Span and Subject Area

GRADE SPAN	READING TESTS	MATH TESTS
E (Elementary)	0.022*	0.049*
	(9-7), 77.7%	(10-8), 94.7%
M (Middle)	0.011	0.055*
	(9-7), 85.7%	(10-8), 92.0%
H (High School)	0.054	-0.015
	(7-5) 98.3%	(8-6), 98.6%
Combined E/M	-0.009	-0.012
	(15-12), 93.4%	(15-12), 97.9%
E, M, and Combined E/M	0.002	0.020*
	(31-17), 90.3%	(33-18), 96.8%
All	0.008	0.014
	(17-14), 98.4%	(18-15), 97.7%

NOTES: Asterisks indicate effect size significantly different from zero at the 5% 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. The percentage refers to the I2 estimate of the percentage of the variation across estimates that reflects true variation in the effect of charter schools, rather than just statistical noise. Thus for example in the reading test result for elementary schools "(9-7), 77.7% " indicates nine estimates covering seven locations (with two studies each of New York City and San Diego schools, and that 77.7% of the variation across estimates in the literature may reflect true variation in the effect of charter schools.

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.

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

One issue faced in this analysis is that there are many studies of individual KIPP schools, which typically have quite large positive effect sizes and relatively small standard errors. If placed into a meta-analysis alongside studies of entire districts or states, the KIPP studies have disproportionate influence. For this reason, our main results in this section, shown in Table 2 (page 15), exclude the KIPP results from both the middle school results and the results that combine elementary studies, combined elementary/middle studies, and middle school studies. 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.

Table 2 shows the main results. For each grade span, results for reading and math appear in the first and second columns respectively, and the first row shows the estimated overall effect size. Effect sizes that are statistically significant (at the 5% level) are indicated with an asterisk. For elementary schools, we conclude that overall there is a positive and significant

effect of charter schools on both reading and math achievement, with estimated effect sizes of 0.022 and 0.049 respectively.

Below these estimates we present 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, "(9-7)" indicates that we found and used 9 separate estimates from 7 geographic areas in calculating the overall effect. It is important to keep in mind that even though the literature has grown robustly in the three years since the Betts and Tang (2008) literature review, there are still surprisingly few rigorous studies that specifically study the impact of charter schools at the elementary level. The same applies to high school studies, although we now have quite a few estimates of the impact of charter schools at the middle school level.

The final number in the second row of results for each grade span shows an estimate of the percentage of the variation across estimates that reflect 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 77.7% and 94.7% of the variation reflects true variations in impact. These are large percentages, which suggests that in attempting to find an "average" or "overall" effect, we must be very careful to recognize that there appear to be important variations in charter school effects across studies, and, implicitly, across 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.

There are relatively few studies that focus specifically on charter high schools. As shown in the third row of Table 2, no significant effect emerges overall in these studies.

A number of studies combine elementary and middle schools together and, as shown in the fourth row of Table 2, overall there is no significant effect of attending a charter school on reading or math achievement found 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 2 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 bottom row of

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

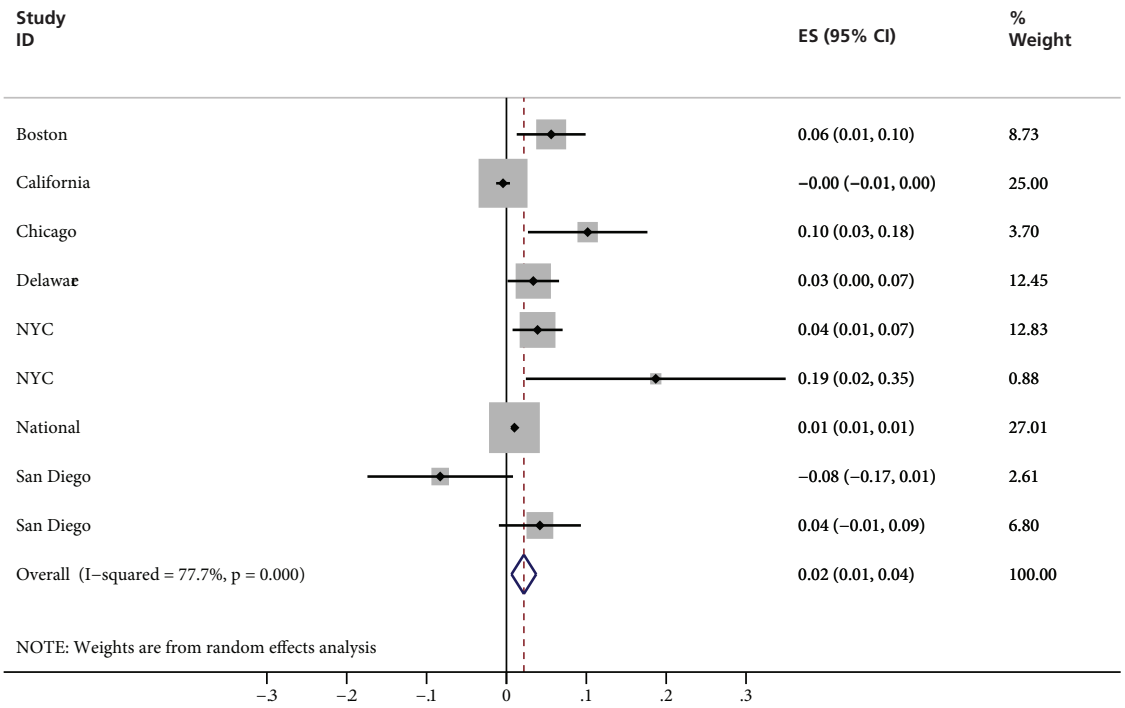


Figure 2. Elementary School Math Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study

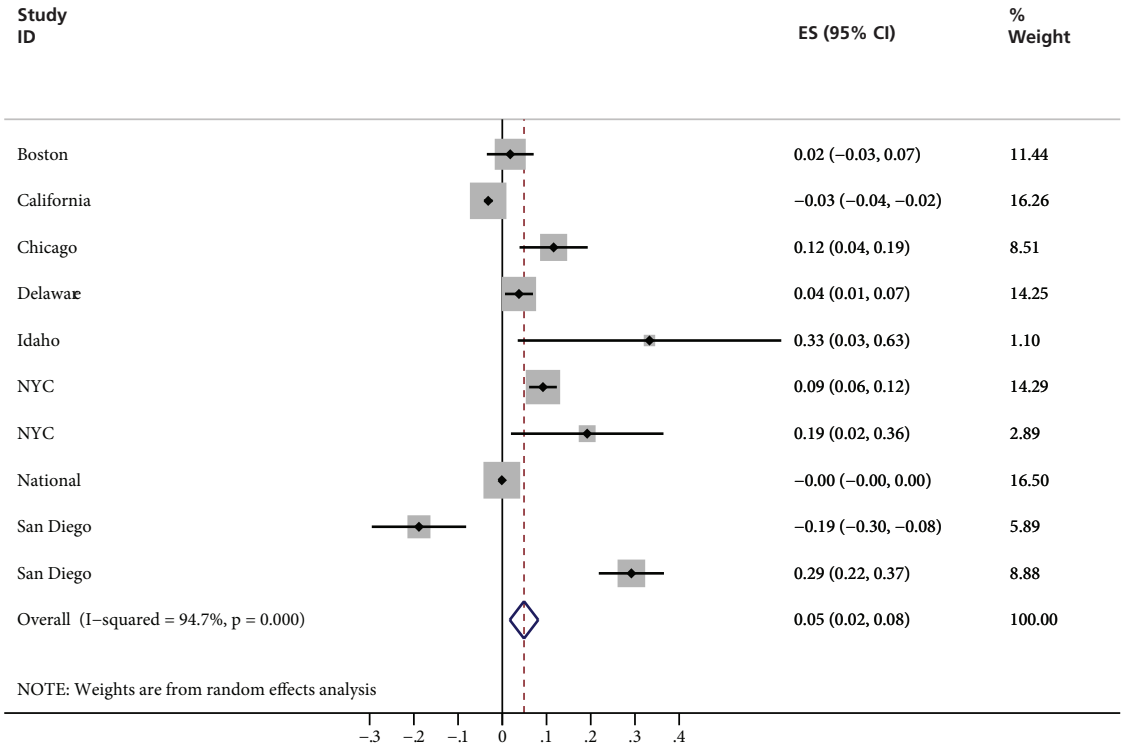


Table 2 shows that the mean effect sizes in reading and math are small and statistically insignificant, but on the other hand almost 100% of the variation across studies appears to be true variation.

It is useful to look at the effect sizes of individual studies and how they contribute to the overall estimates shown in Table 2. 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% 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% confidence interval.

Elementary school studies with the largest estimated effect size for charter school attendance include studies of New York City, Boston, and Chicago. The only study with a large negative (yet 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 timeframe 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 regards to math and reading achievement over time.)

The bottom left of 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 found in all of our analyses. Thus, the notion that we are estimating a homogeneous effect size across studies is roundly rejected.

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 underlying variance in effect sizes that reflects true variation. Smaller, less precise studies get less weight than larger, more precise studies; 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. As we will demonstrate in a later section that shows histograms of effect sizes, how one weights the estimates matters greatly. The weighting scheme here is optimal in that it produces the minimum variance estimate of the overall effect.

Figures 3 and 4 show the estimated effects for middle school studies for reading and math respectively. For reading, estimates lie in a fairly narrow band centered at just above zero. Positive results from Boston exhibit the largest effect size in these studies. Figure 4 shows

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

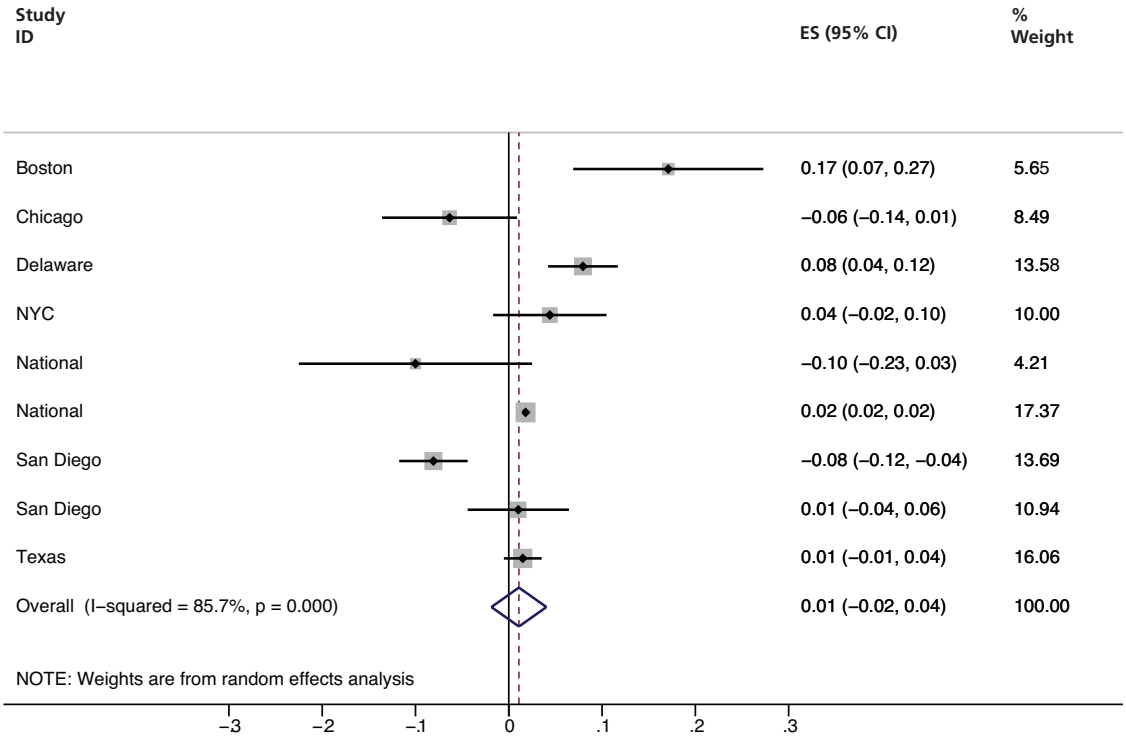


Figure 4. Middle School Math Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study

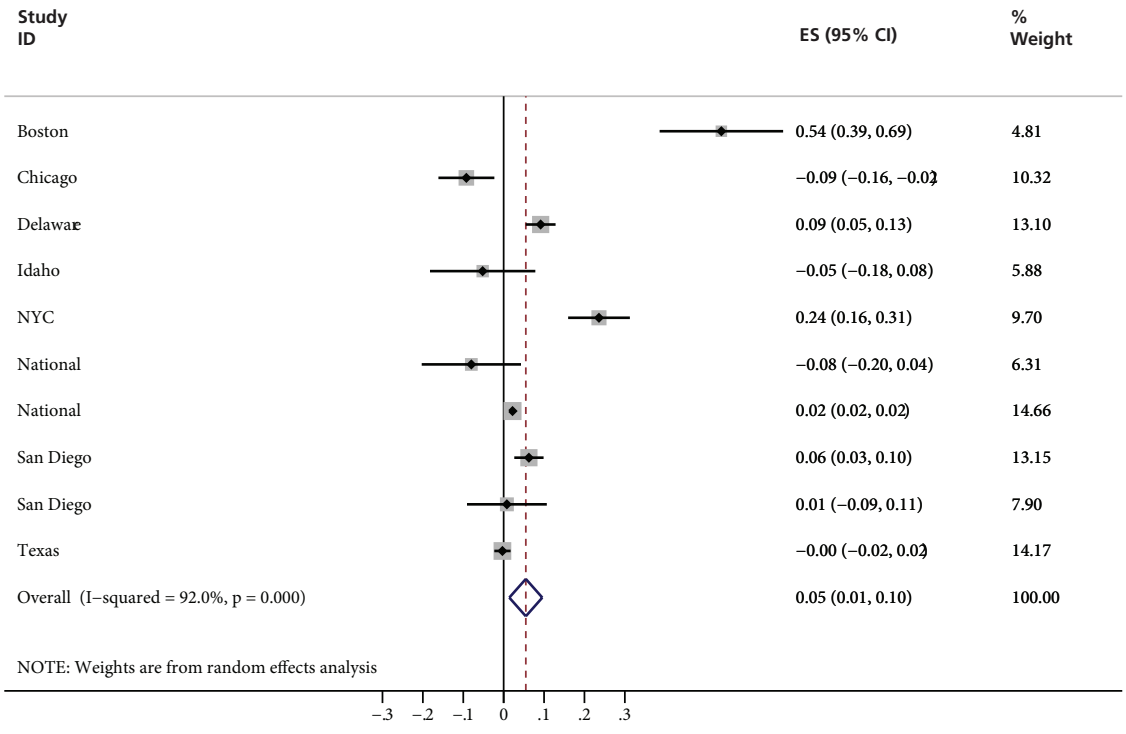


Figure 5. High School Reading Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study

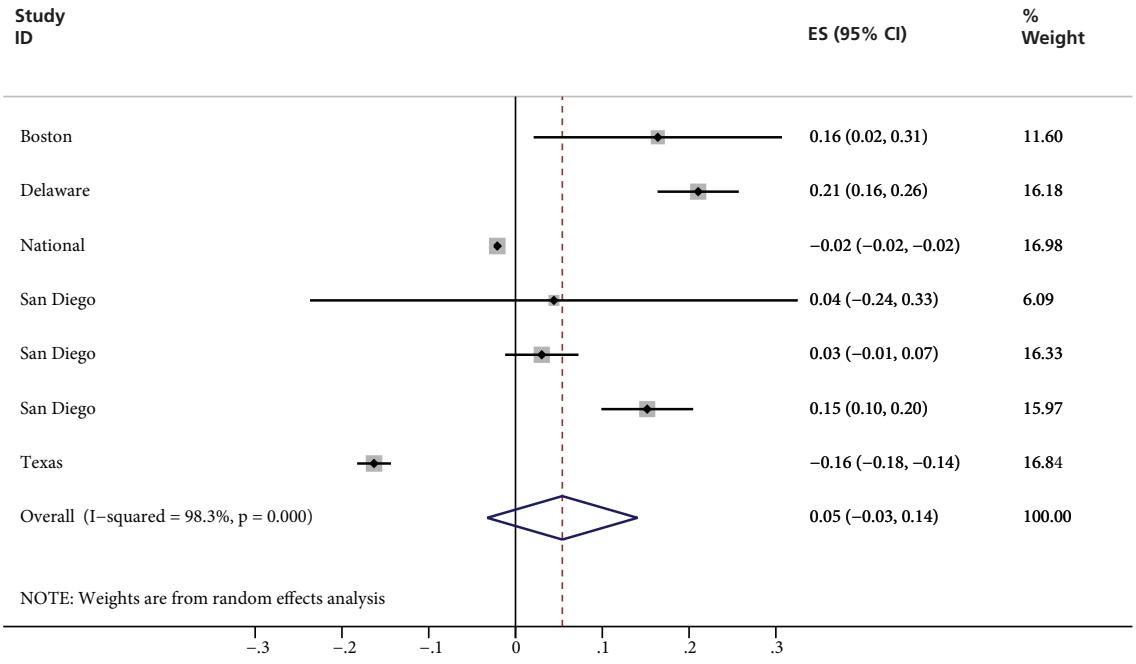


Figure 6. High School Math Effect Sizes by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study

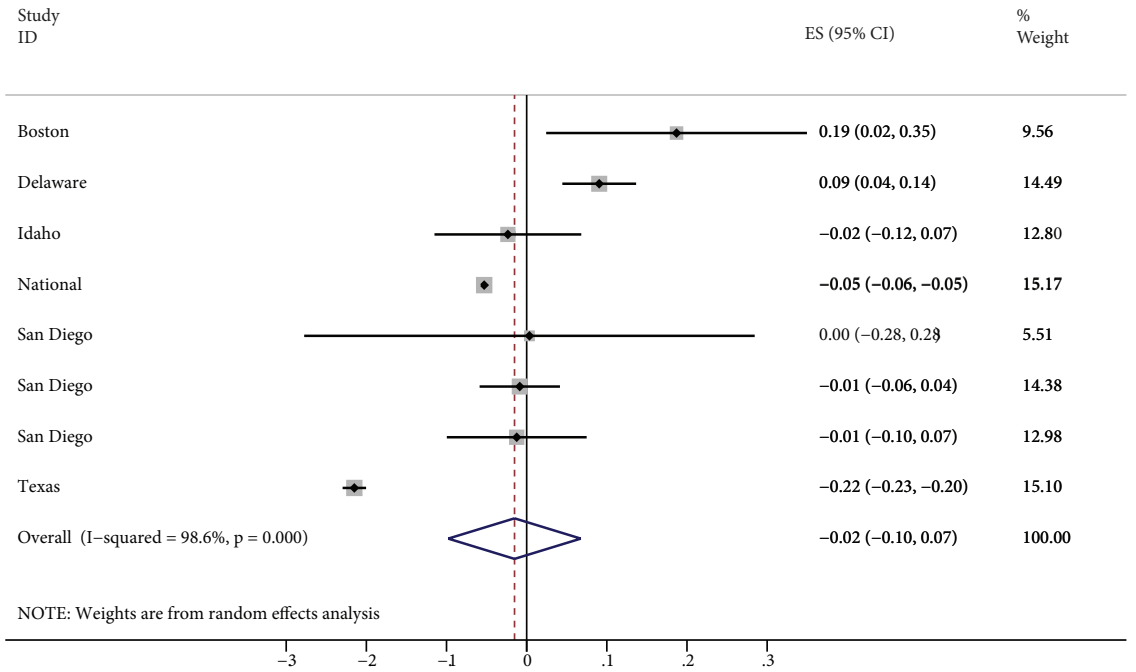


Figure 7. Reading Effect Sizes for Studies that Combine Elementary and Middle Schools by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study

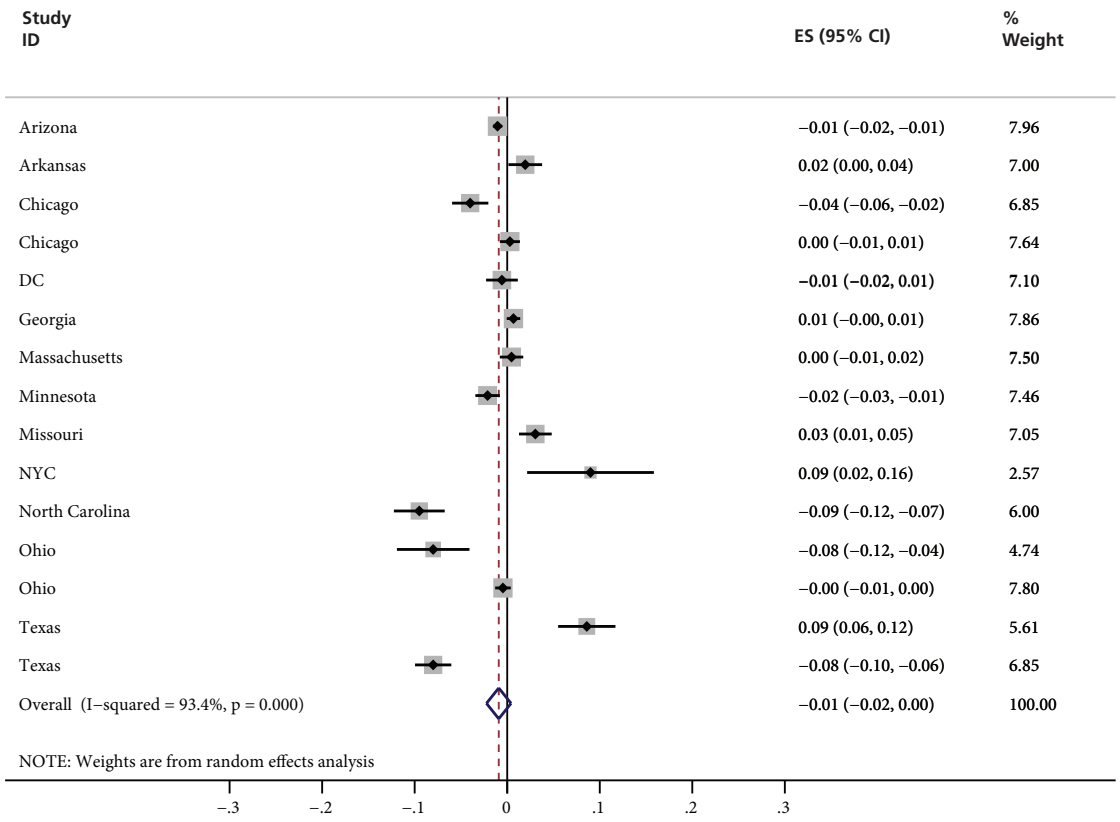
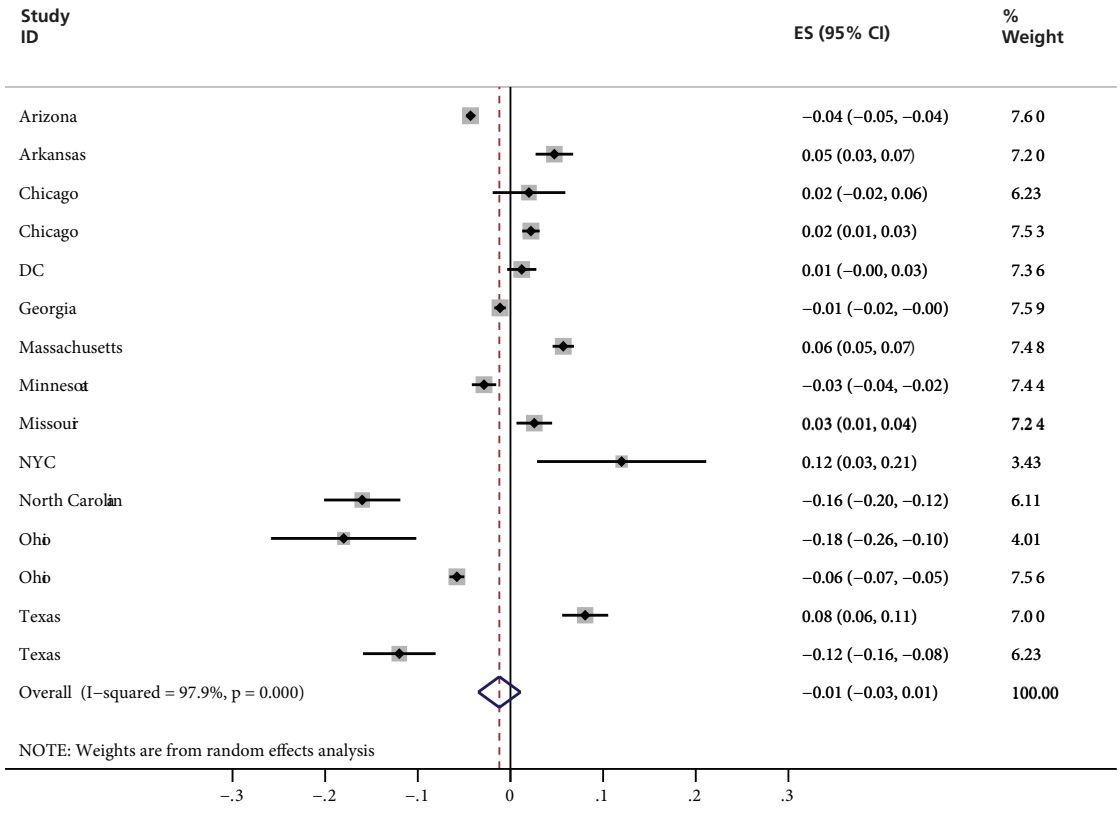


Figure 8. Math Effect Sizes for Studies that Combine Elementary and Middle Schools by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each Study



results for math, and again the biggest outlier is the result from Boston, with an effect size more than double the size of the next biggest estimate (from New York City). Notably in both figures, one of the largest negative estimates derives from the national study by Gleason et al. (2010), which uses a lottery approach. One interpretation is that some of the other studies that do not rely on randomization may be biased upward. On the other hand, results from this national study are estimated quite imprecisely compared to most of the other studies.

Figures 5 and 6 show corresponding figures for high school results. Behind the overall estimates that are insignificantly different from zero, there are a number of studies that find statistically significant positive and negative effects of attending a charter school.

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 Promise Academies in the Harlem Children's Zone in New York City and another study of Texas showing the largest positive effects, and studies of North Carolina 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. This 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 that 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 or 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.[7]

Figure 8 shows estimates for math achievement from studies that combine elementary and middle schools. Again, the overall insignificant estimate masks considerable variation. The studies with the largest estimated positive effects come from 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 produces the differing estimates in the directions outlined above for reading.)

7. The authors 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 estimates 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 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

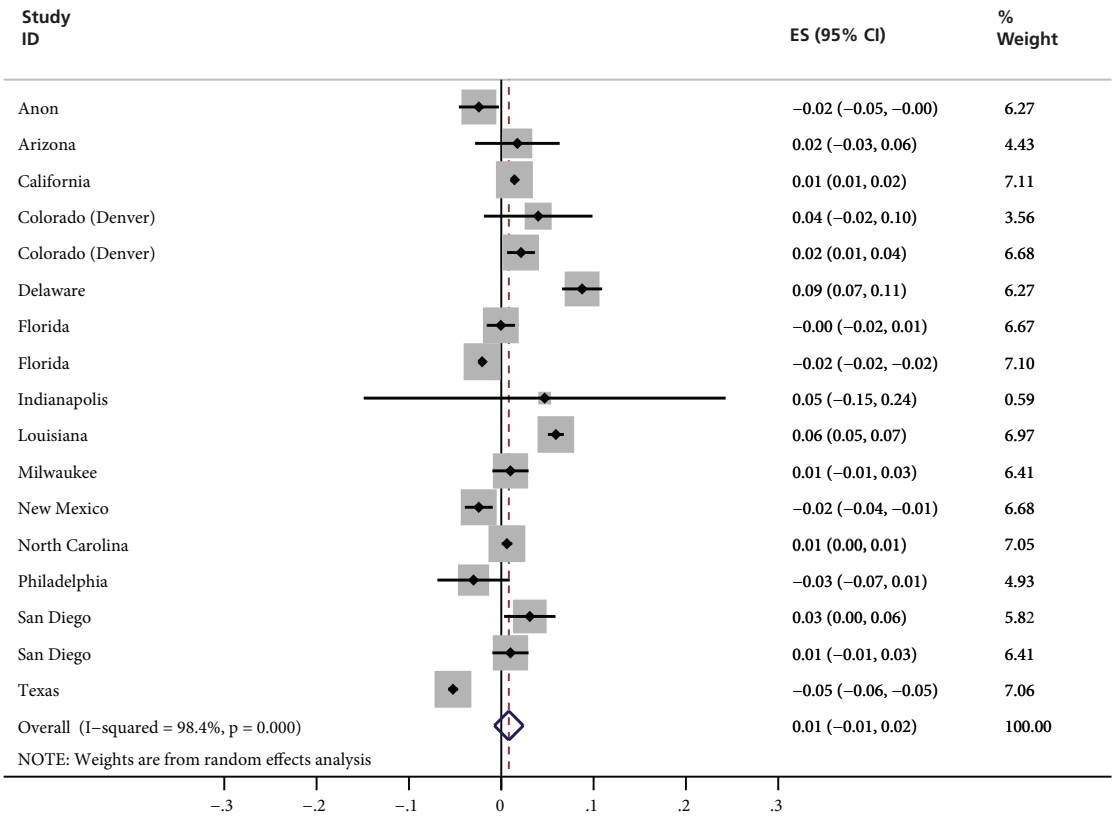


Figure 10. Math Effect Sizes for Studies that Combine Elementary, Middle, and High Schools by Study, Showing Weights Ascribed by Random-Effects Meta-Analysis to Each

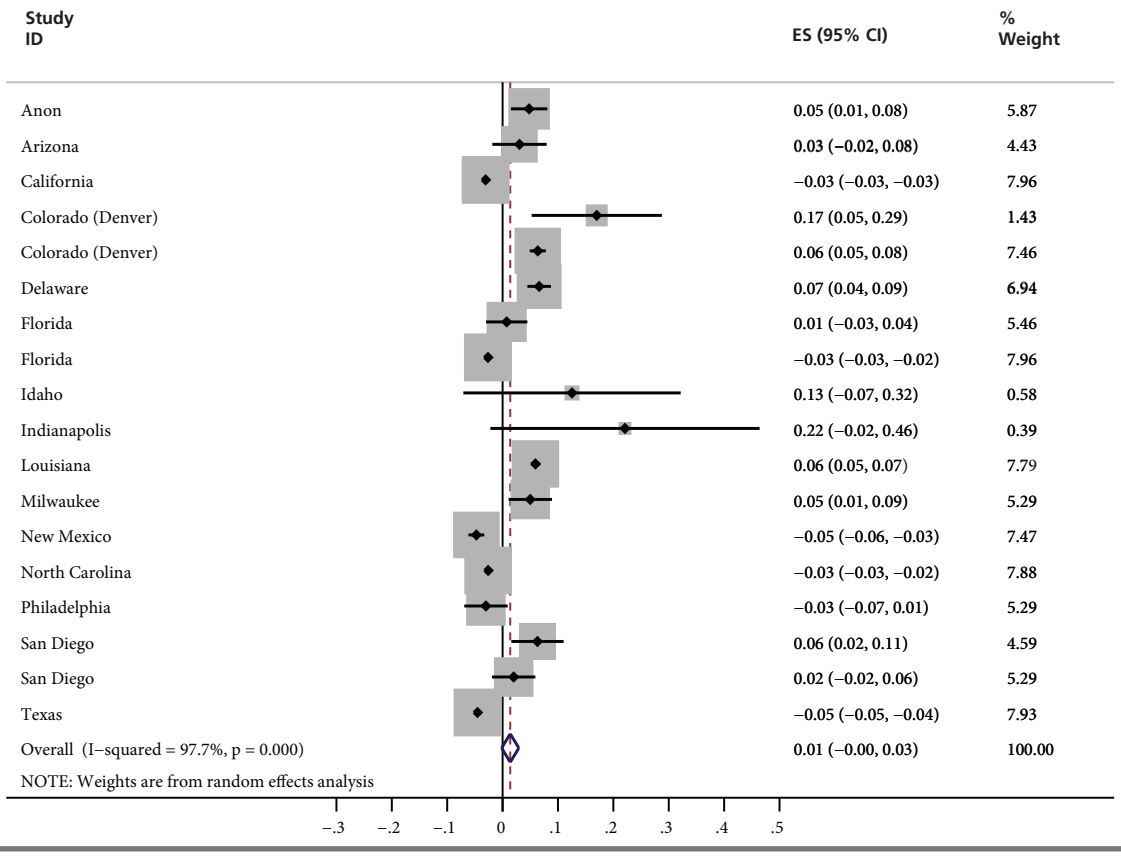


Table 3. Results with KIPP School Estimates Included, and KIPP School Estimates by Themselves: Effect Sizes and Significance from Meta-Analysis, by Grade Span and Subject Area

GRADE SPAN	READING TESTS	MATH TESTS
INCLUDING KIPP SCHOOLS		
M (Middle)	0.070*	0.180*
	(38-33), 88.3%	(39-34), 96.8%
E (Elementary), M (Middle), and Combined E/M	0.034*	0.105*
	(60-43), 90.8%	(62-44), 98.6%
RESULTS INCLUDING ONLY KIPP ESTIMATES		
M (Middle)	0.096*	0.223*
	(29-unknown), 82.7%	(29-unknown), 93.7%

NOTES: Asterisks indicate effect size significantly different from zero at the 5% 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.

Figures 9 and 10 show reading and math results for the "All Grade Span" studies, which in both cases 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 exception is a positive effect size of 0.09 found for Delaware by Miron et al. (2007). For math, as shown in Figure 10, the overall estimate is positive and almost significant at the 5% level. There are three large positive effect size estimates, for Indianapolis, Denver, and Idaho, but each receives 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.

The middle school results presented in Table 2 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, slightly over 50% of the weight in these meta-analyses goes to the studies of individual KIPP schools; yet our estimates suggest that nationwide KIPP schools account for around only 2% 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 first meta-analysis of the KIPP literature. 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.

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

GRADE SPAN	READING TESTS	MATH TESTS
E (Elementary)	0.034*	0.072*
	(8-6), 79.5%	(9-7), 95.2%
M (Middle)	0.010	0.068*
	(8-7), 87.2%	(9-8), 92.8%
H (High School)	0.072	-0.002
	(6-4), 98.5%	(7-5), 97.5%
Combined E/M	-0.023	-0.041
	(6-5), 95.5%	(6-5), 96.9%
E, M, and Combined E/M	0.008	0.038*
	(22-10), 92.0%	(24-11), 95.0%
All	0.016	0.041*
	(10-9), 86.6%	(11-10), 67.7%

NOTES: Asterisks indicate effect size significantly different from zero at the 5% 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 2, we also exclude the KIPP studies from the middle school and combined elementary school, middle school and combined elementary/middle school studies.

Just as the KIPP studies would dominate the middle school analysis had they been included in Table 2, the CREDO studies of individual states and cities also play a major role in the "Combined Elementary/Middle" and the "All Grade Span" studies, contributing 9 of 15 and 7 of 17 studies respectively. CREDO studies contributed only a few estimates to the results for the other grade spans. One can also examine the results once the many results published by CREDO are removed. A concern about these CREDO estimates is that they all use the same non-experimental method that hinges upon how successfully the studies matched charter school students with counterparts at traditional public schools. Because for many charter school students they were matched with non-charter students using their characteristics and test scores once at the charter school, this could bias the results. Table 4 shows the results when the aforementioned grade spans estimated from Table 2 are repeated without the CREDO studies.

Compared to Table 2, the estimated effect sizes are generally slightly higher when we exclude the CREDO studies. In the case of math achievement in studies that combined all three grade spans ("All Grade Span"), the estimated charter school effect becomes positive and significant when we exclude the CREDO studies. Three explanations, which are not mutually exclusive, might account for the higher estimates when we drop the CREDO

studies. First, the method used in the CREDO studies is not experimental and does not use student fixed effects, and therefore is forced to make comparisons between charter school students and observationally similar students at traditional public schools. Many charter school students were matched with non-charter students using the charter school students' characteristics and test scores measured once at the charter school. Because these test scores are endogenous outcomes of the charter school experience, this could bias the results. A second explanation for why the CREDO results are slightly less optimistic than those of other studies is that the other studies may have focused on unrepresentative charter schools that boost achievement more than other charter schools do. In contrast, the CREDO studies attempt to be comprehensive at the state level (or at the city level, in the case of Denver and Washington, D.C.). A third explanation is offered by Hoxby (2009), who notes that the comparison student is in fact an average over several students, and because the regression controls for a lagged achievement score, the charter dummy will be biased downward because there will be more measurement error in the individual charter school student's own lagged score than in the mean lagged test score for that student's control group. CREDO (2009b) provides a rebuttal, pointing out correctly that in theory the sign of the bias is unknown because in a multivariate regression the direction of the bias will depend on the other explanatory variables as well.[8]

Interpreting the Effect Sizes

Elementary school effect sizes reported in Table 2 are 0.022 and 0.049 for reading and math. Perhaps the most nationally representative estimates of effect sizes at the middle school come from the second row of Table 2, in which we drop the many estimates for individual KIPP schools. These effect sizes are 0.011 and 0.055 for reading and math respectively.

Some simple comparisons provide some perspective on whether an effect size of roughly 0.05, that is, an increase of 5% of a standard deviation, is large or small. For comparison purposes, 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% -1.5% of a standard deviation. With an effect size of about 0.05 for math at the elementary and middle school levels, a student with median test scores — ranking at the 50th percentile — would be predicted to move to around the 52nd percentile after one year at the charter school. This is not a large change but over several years of such gains, it could be quite meaningful.

The charter school effect sizes estimated for reading, of about 0.02 and 0.01 at the elementary

8. CREDO (2009c) reports the mean and variance of test scores for the charter school students and the averaged test scores for the "synthetic controls." The report claims that the variances of test scores in the two samples are equal, thus obviating the likely bias pointed out by Hoxby (2009). It is puzzling that the variance of a mean test score, where the median number of students contributing to each mean test score was 6, would be equal to the variance of the test scores of individual charter school students. It should be smaller by a factor of 1/6.

Table 5. Effect Sizes for Studies of Urban Districts and Schools, by Grade Span and Subject Area

GRADE SPAN	READING TESTS	MATH TESTS
E (Elementary)	0.046*	0.085
	(6-4), 61.8%	(6-4), 92.2%
M (Middle)	0.009	0.139
	(5-4), 87.0%	(5-4), 94.8%
H (High School)	0.101*	0.019
	(4-2), 78.2%	(4-2), 42.7%
Combined E/M	-0.003	0.021*
	(4-3), 86.2%	(4-3), 47.7%
E, M, and Combined E/M	0.016	0.077*
	(15-5), 84.1%	(15-5), 92.4%
All	0.008	0.045*
	(8-6), 63.2%	(8-6), 74.8%

NOTES: Asterisks indicate effect size significantly different from zero at the 5% 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.

and middle school levels, are much smaller, but are close to the Clotfelter, Ladd, and Vigdor (2007) estimated effects of reducing class size by five students. These estimated effects will move a student up in the distribution of student achievement rather slowly. After one year at a charter school, a student who started at the 50th percentile would have risen only to percentiles 50.4 to 50.8.

As shown above, the estimates for KIPP middle schools are far higher than our average estimates, with estimated effect sizes for reading and math at 0.096 and 0.223 respectively. These impressive effect sizes are enough to move a student initially at the 50th percentile to the 54th and 59th percentiles in a single year.

Urban Districts and Schools

Table 5 shows the results when we focus on studies of urban districts or on individual schools in urban areas. Although the number of studies is smaller than in the full set of studies shown in Table 2, the effect size estimates are almost always higher in the urban subsample than in the overall sample. There are also several cases in which charter schools had no significant effect in the overall sample, but in the set of studies of urban schools, the estimated effect becomes positive and significant. These include reading achievement in high schools, and math achievement in studies that combine elementary and middle schools, and studies that combine all grade spans. One counterexample is math achievement

Table 6. Effect Sizes for White, Black, Hispanic, and Native American Students and Significance from Meta-Analysis, by Grade Span and Subject Area

RACE/ ETHNICITY	E	M	H	COMBINED E/M	E, M, AND COMBINED E/M	ALL
READING TESTS						
White Students	-0.093	-0.122*	0.088*	-0.029*	-0.033*	-0.007
	(1-1)	(1-1)	(1-1)	(12-10), 82.9%	(14-11), 81.6%	(11-10), 97.7%
Black Students	-0.007	-0.001	0.060	0.020	0.016	-0.002
	(2-2), 73.2%	(1-1)	(1-1)	(13-10), 91.0%	(16-12), 89.1%	(11-10), 97.0%
Hispanic Students	-0.013	-0.071*	0.003	-0.032	-0.033	-0.007
	(2-2), 76.1%	(1-1)	(1-1)	(13-10), 86.8%	(16-12), 85.8%	(11-10), 94.1%
Native American Students	Individual results for elementary, middle and high school students do not yet exist.			-0.147	-0.147	-0.042
				(7-7), 95.6%	(7-7), 95.6%	(7-7), 59.8%
MATH TESTS						
White Students	-0.120	-0.037	-0.116*	-0.057*	-0.058*	-0.002
	(1-1)	(1-1)	(1-1)	(12-10), 96.1%	(14-11), 95.3%	(11-10), 98.9%
Black Students	0.032	0.070	0.007	0.026	0.030*	-0.000
	(2-2), 79.4%	(1-1)	(1-1)	(13-10), 91.4%	(16-12), 90.9%	(11-10), 96.3%
Hispanic Students	-0.121	0.068	0.052*	-0.006	-0.002	-0.001
	(2-2), 93.4%	(1-1)	(1-1)	(13-10), 90.4%	(16-12), 90.9%	(11-10), 92.3%
Native American Students	Individual results for elementary, middle and high school students do not yet exist.			-0.013	-0.013	-0.103*
				(5-5), 0%	(5-5), 0%	(7-7), 53.8%

NOTES: Asterisks indicate effect size significantly different from zero at the 5% 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. Individual results for elementary, middle, and high school Native American students do not yet exist.

in elementary schools, for which the effect size is positive and significant in the full sample as shown in Table 2, but becomes insignificant although larger in the subsample of studies that have an urban focus.

We can only speculate as to the 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 under-serving their students more than their non-urban counterparts.

Results by Student Subgroup

Tables 6 shows results for white, black, Hispanic, and Native American students. The general pattern is as follows. For white students, the estimated effects are always negative and in most cases statistically significant. The main exception is high school reading achievement, for which attending a charter school is associated with a positive and significant effect

Table 7. Effect Sizes for Studies of Selected Subsamples of Student Populations and Significance from Meta-Analysis, by Grade Span and Subject Area

STUDENT POPULATION	M	COMBINED E/M	E, M, AND COMBINED E/M	ALL
READING TESTS				
English Language Learners	0.384*	-0.003	0.006	0.054*
	(1-1)	(9-9), 37.2%	(10-10), 54.7%	(7-7), 74.0%
Special Education Students	0.298	-0.000	-0.000	0.009*
	(1-1)	(9-9), 40.7%	(10-10), 46.2%	(7-7), 0.0%
Students Eligible for Federal Meal Assistance	Results for middle school students are not yet available.	0.011	0.011	0.014
		(9-9), 91.0%	(9-9), 91.0%	(7-7), 89.6%
MATH TESTS				
English Language Learners	0.451*	0.008	0.013	0.025*
	(1-1)	(9-9), 50.3%	(10-10), 62.6%	(7-7), 37.3%
Special Education Students	0.441*	-0.002	0.000	0.012*
	(1-1)	(9-9), 63.2%	(10-10), 70.6%	(7-7), 0.0%
Students Eligible for Federal Meal Assistance	Results for middle school students are not yet available.	0.021	0.021	0.006
		(9-9), 92.4%	(9-9), 92.4%	(7-7), 95.4%

NOTES: Asterisks indicate effect size significantly different from zero at the 5% 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. Results for elementary and high school special education students and English Language Learners are not yet available.

size. For black students, the results are mostly insignificant, the lone exception being math achievement when we pool elementary school studies, middle school studies, and combined elementary/middle school studies, for which a small and significant positive effect of 0.03 emerges. For Hispanic students, estimated effects are mostly insignificant, and of varied signs. Two exceptions are a negative effect on reading tests in middle school studies and a positive effect for high school math, both of which are significant. For studies of Native American students, each of which comes from a CREDO study, effects are negative but significant only in the case of math achievement for studies that combine grades across elementary, middle, and high schools. If we had to rank the effects by student racial/ethnic group, it would seem that from the most positive to the most negative results, the ranking would be: black students, followed by Hispanic, Native American, and finally white students.

These results by racial and ethnic group are much less positive than the results in the earlier tables that include all students. The main reason for this difference is that the number of studies available is far smaller than in the analysis of overall effects on students. A single study from San Diego (Betts et al., 2005) provides the only (or one of only two) estimates

by race/ethnicity for charter schools at the elementary, middle school, and high school levels. This single study is useful but is unlikely to be nationally representative. The studies by CREDO account for all of the estimates for Native American students, over two-thirds of the studies for other racial/ethnic groups in models that combine elementary and middle schools, two-thirds or more of the studies that combine elementary and middle schools, and just under two-thirds of the studies that combine all grade spans (the "All grade spans" category). As shown in Table 4, the CREDO results in general are somewhat less positive than results from other reports. Another partial explanation that cannot explain the lower coefficients, but that can explain the lack of significance, is that the subsamples used in the models by student group are smaller and less likely to yield significant results than when one estimates an overall effect for all students using the same data source.

Table 7 shows results for studies of students eligible for federal meal assistance, special education students, and English language learners. The estimated effect size for three combinations of grade spans for students eligible for federal meal assistance are all positive, but small and statistically insignificant. Again, sample size may be an issue. All of the studies on students eligible for federal meal assistance contributing to this table are from reports by CREDO.

There are some signs of positive effects of attending charter schools for special education students in studies that involve grades from elementary, middle, and high school (as shown in the bottom row of the table), but the effect sizes are very small. Studies that combine elementary and middle grades find no effects. One study of middle schools finds a very large and positive effect size (Angrist et al., 2010); but again, this is for one KIPP school in Massachusetts and is not likely to be nationally representative.

The patterns (and set of underlying studies) for English language learners are similar to those for special education students. Studies that combine grades from all three grade spans find positive and significant effects for both reading and math achievement. Studies that combine elementary and middle grades find no effects. Angrist et al. (2010) find extremely large effect sizes in both reading and math for one KIPP school, but this is unlikely to be nationally representative.

HISTOGRAMS AND VOTE-COUNTING ANALYSIS

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

The histograms support the results of overall findings and additionally offer a view of the distribution of effects. In the previous section we demonstrate that, on average, charter schools are serving middle school and elementary school students well, and serving high school students moderately. 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 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 -1 and 1. We create histograms for each grade span separately in order to examine the different effects according to the grade levels of the students studied. We generate both unweighted histograms, in which all studies receive equal weight regardless of whether it is a study of a single school or an entire state, as well as weighted histograms, in which studies with more students contributing to the estimates receive greater weight. (The formal meta-analysis in the preceding section used variable weights, but as shown by the weights in the figures from that section, they are fairly close to equal weight estimates due to the repeated finding that the sampling variance of individual studies was small relative to the true underlying variation in the effect sizes.)

We start by discussing middle schools because it is these pictures that offer the clearest interpretation. In examining these histograms we find that, compared to all other grade spans, the effect sizes are largest and most often positive in studies of middle school students. The separate figures for middle schools, which display results for reading and math, with and without the KIPP schools, show that most of the mass in all of the histograms studying middle school students is in the positive region, whether these estimates are unweighted or weighted, and in both reading and math. This indicates charter schools are generally serving middle school students very well.

Because a considerable proportion (nearly 75%) of the estimates of charter middle school effectiveness are studies of KIPP schools, it is worth investigating KIPP and non-KIPP studies separately. We do this in order to confirm that the positive middle school charter effect for both reading and math is not driven entirely by strong positive effects in the KIPP schools. We note that while we can identify studies that explicitly study only KIPP schools, many of the studies that do not exclusively study KIPP schools may still contain KIPP schools. For example, a study of charter schools in Florida may contain some KIPP schools, but we cannot distinguish the estimates of KIPP schools versus non-KIPP schools from that sample. However, breaking out the results of exclusively KIPP studies is illustrative. We see in Appendix Figure A1 and A2, which show the unweighted histograms for the exclusively KIPP studies, that KIPP studies generally find large and positive charter school effects.[9]

Because KIPP schools currently comprise a small share of charter schools, we focus on the studies that are not exclusively of KIPP schools. Figure 11 shows the histogram for non-KIPP middle school reading results when each estimate is treated equally, while Figure 12 shows the histogram for non-KIPP middle school reading results when the estimates that are derived from more observations receive greater weight. Figures 13 and 14 show similar histograms for math effect sizes. In all cases, more of the effects lie in the positive region than the negative region. When we compare the observation-weighted figures to the unweighted figures, the charter school effect still looks positive, but it is notably less sizable. In both reading and math, the height of the bars representing the very large positive results seen in the unweighted histograms shrink to nearly zero in the weighted histograms, indicating that most of the estimates that find large positive results are studies of relatively few students. While most of the estimates still fall in the positive region, the effects now appear much smaller. The largest percentage of estimates of effect sizes for both reading and math are found in the bin of 0 to 0.05 standard deviations. Over 70% of the weighted estimates for math, and over 90% of the weighted estimates for reading, fall in this range.

At the middle school level the histograms for math and reading look roughly similar, with more studies finding positive results than studies finding negative results in both cases. The unweighted pictures suggest that the upside for math seems to be somewhat larger than for reading, with more studies finding estimates larger than half a standard deviation in math than in reading. Looking at the weighted pictures, however, suggests

9. We do not present the weighted histograms separately because these look nearly identical to the weighted histograms that combine the KIPP and non-KIPP studies. The similarity of the combined middle school and the non-KIPP study histograms is due to the fact that the KIPP studies are generally low weight, studying one school at a time. This causes their influence in the weighted histograms of the combined studies to diminish.

Figure 11. Distribution of Effect Sizes for Middle School Reading, Non-KIPP Studies Only, Treating Each Estimate Equally

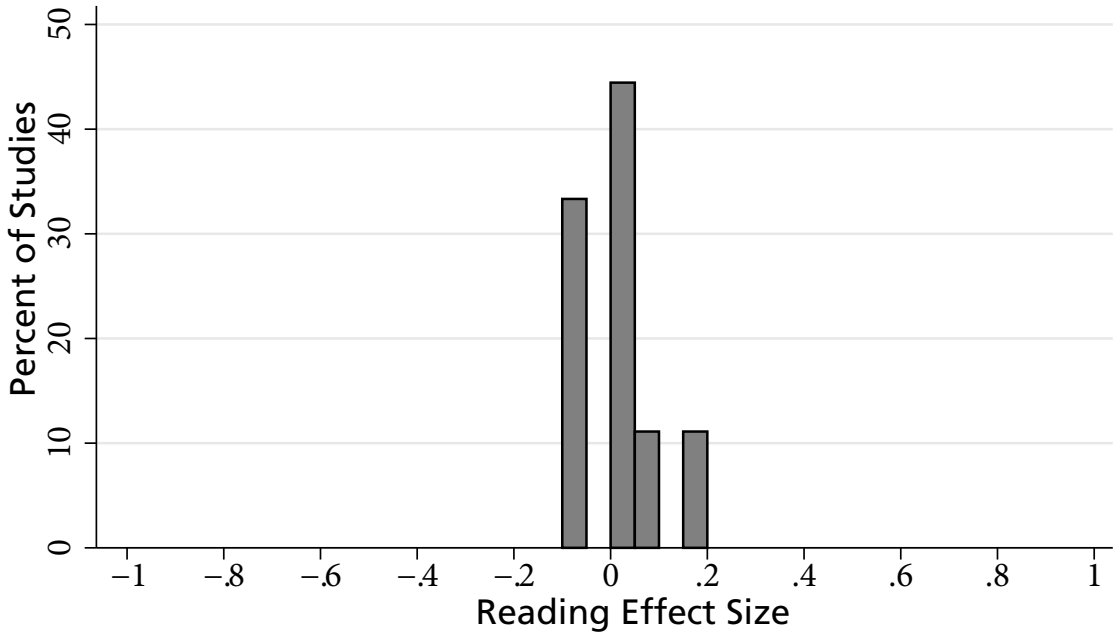


Figure 12. Distribution of Effect Sizes for Middle School Reading, Non-KIPP Studies Only, Weighting Each Estimate by Number of Observations

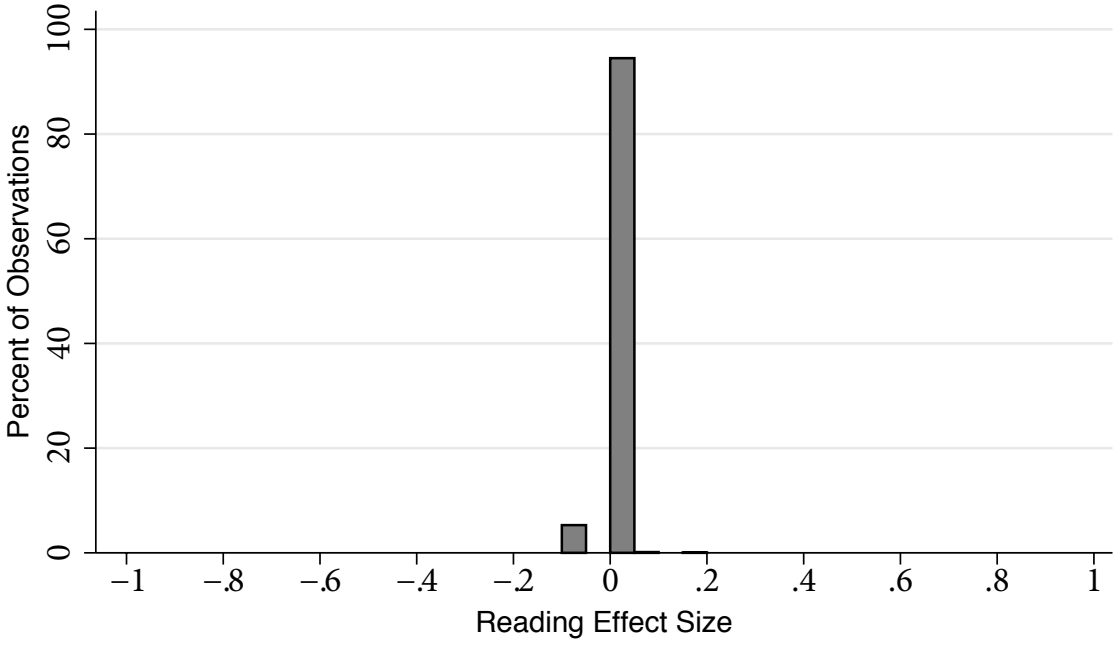


Figure 13. Distribution of Effect Sizes for Middle School Math, Non-KIPP Studies Only, Treating Each Estimate Equally

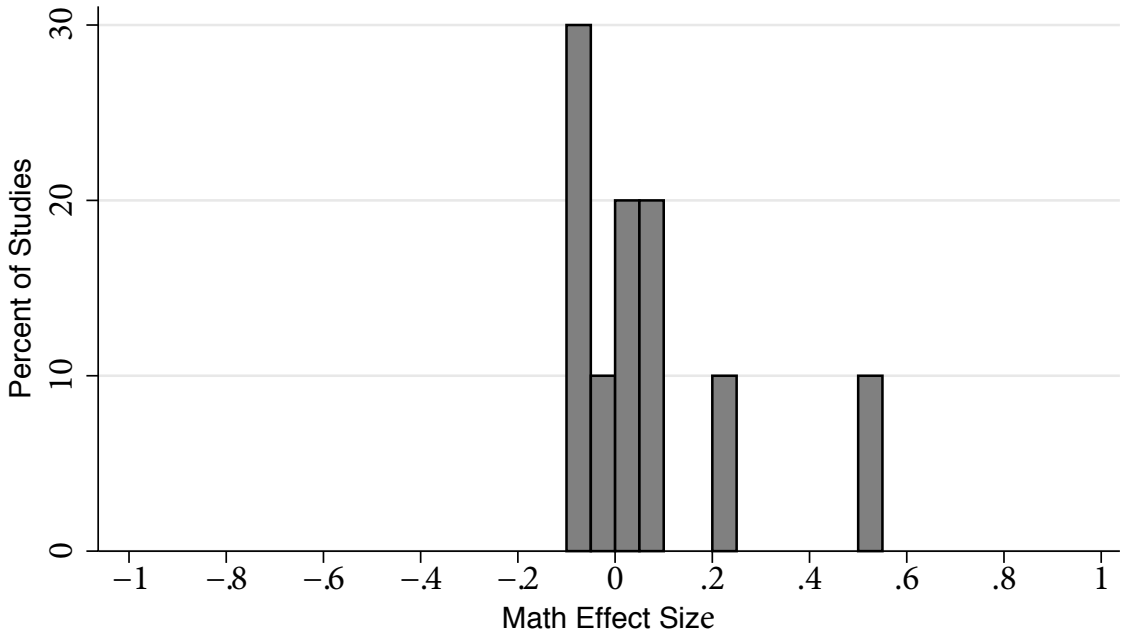
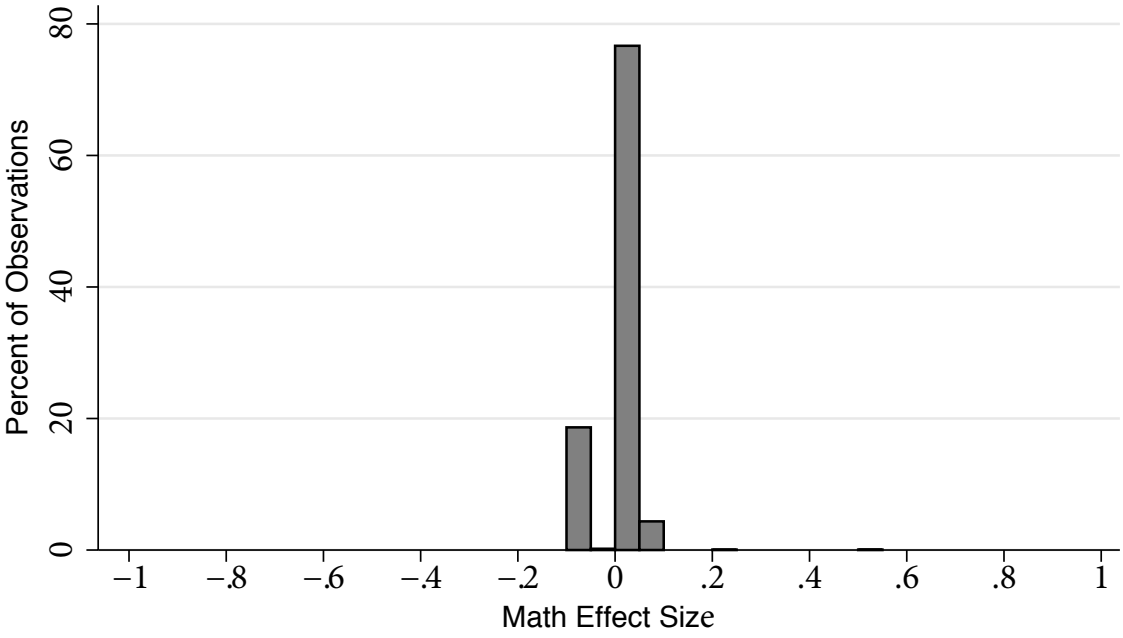


Figure 14. Distribution of Effect Sizes for Middle School Math, Non-KIPP Studies Only, Weighting Each Estimate by Number of Observations



that while the upside for math may be larger than for reading, the downside for math may also be realized more frequently than for reading. In the figures in which the estimates are weighted by number of observations, Figures 12 and 14, the bar representing studies finding negative estimates is much taller in math (Figure 14) than in reading (Figure 12). Nearly 20% of estimates find negative math effects, while in reading fewer than 10% of estimates fall in the negative range. These percentages can also be found in Tables 8 and 9, which report results from the vote-counting analysis.

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

GRADE SPAN	SIGN AND SIGNIFICANCE	(1)	(2)	(3)
		Unweighted Excluding KIPP	Weighted by # of observations Excluding KIPP	Weighted by # of observations Excluding KIPP and Excluding CREDO
E (Elementary)	-/Significant	0	0	0
	-/Insignificant	22	47	91
	+/Insignificant	11	3	6
	+/Significant	67	50	3
M (Middle)	-/Significant	11	5	45
	-/Insignificant	22	0	1
	+/Insignificant	33	6	52
	+/Significant	33	89	2
H (High School)	-/Significant	29	91	3
	-/Insignificant	0	0	0
	+/Insignificant	29	4	45
	+/Significant	43	5	52
Combined E/M	-/Significant	40	49	50
	-/Insignificant	13	2	0
	+/Insignificant	20	3	0
	+/Significant	27	47	50
E, M, and Combined E/M	-/Significant	23	41	42
	-/Insignificant	19	15	15
	+/Insignificant	23	4	2
	+/Significant	35	39	42
Studies of All Grades	-/Significant	24	23	9
	-/Insignificant	12	43	64
	+/Insignificant	29	14	21
	+/Significant	35	20	6

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

GRADE SPAN	SIGN AND SIGNIFICANCE	(1)	(2)	(3)
		Unweighted Excluding KIPP	Weighted by # of observations Excluding KIPP	Weighted by # of observations Excluding KIPP and Excluding CREDO
E (Elementary)	-/Significant	20	43	74
	-/Insignificant	10	42	0
	+/Insignificant	10	0	0
	+/Significant	60	15	26
M (Middle)	-/Significant	10	0	0
	-/Insignificant	30	19	67
	+/Insignificant	10	5	17
	+/Significant	50	77	16
H (High School)	-/Significant	25	74	1
	-/Insignificant	38	26	98
	+/Insignificant	13	0	0
	+/Significant	25	0	1
Combined E/M	-/Significant	47	45	42
	-/Insignificant	0	0	0
	+/Insignificant	13	11	12
	+/Significant	40	44	46
E, M, and Combined E/M	-/Significant	30	47	45
	-/Insignificant	9	3	3
	+/Insignificant	12	10	10
	+/Significant	48	40	41
Studies of All Grades	-/Significant	28	30	0
	-/Insignificant	6	7	10
	+/Insignificant	28	48	70
	+/Significant	39	16	20

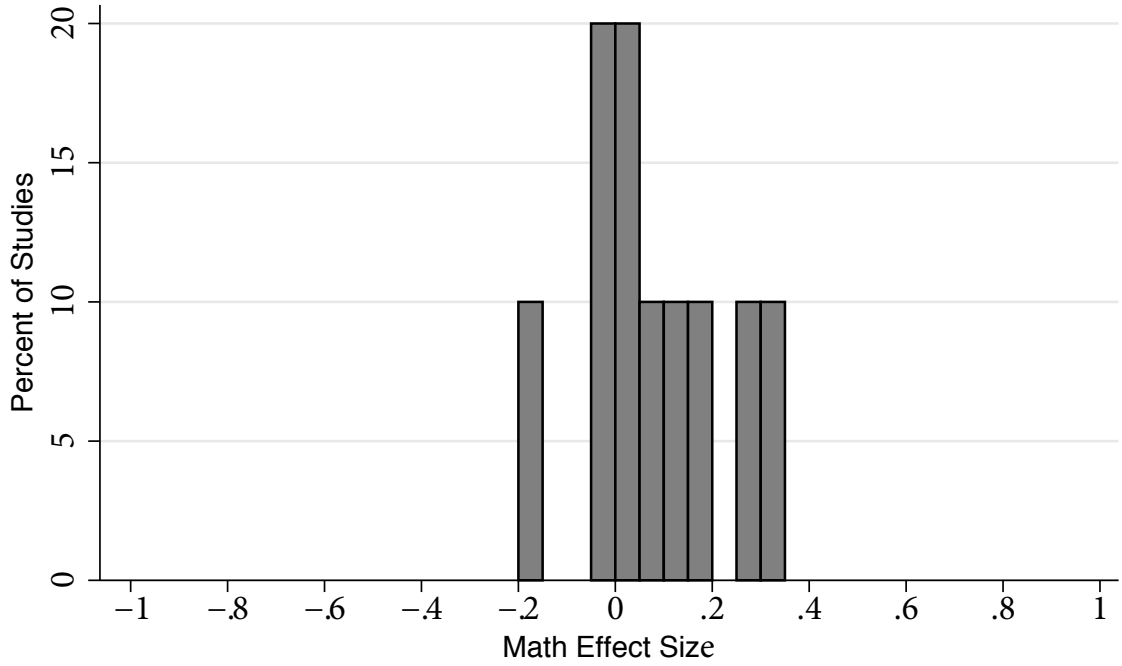
NOTE: Each number indicates the percentage of regression results for the given weighting method and combination of grade spans that fit the stated category of sign and statistical significance. The numbers within each cell may not sum to 100 due to rounding.

Charter schools seem to be doing well at the middle school level, whether they are KIPP schools or not. Middle school charters are outperforming traditional schools, and the KIPP schools especially so.

As in the overall results, the histograms at the elementary school level look generally favorable for charter schools, and more favorable for elementary school reading than for math. However, the pictures are less consistently positive than at the middle school level, and they swing widely on whether we give more weight to the larger studies. Tables 8 and 9 show that when we consider the unweighted effects, 60% of the studies

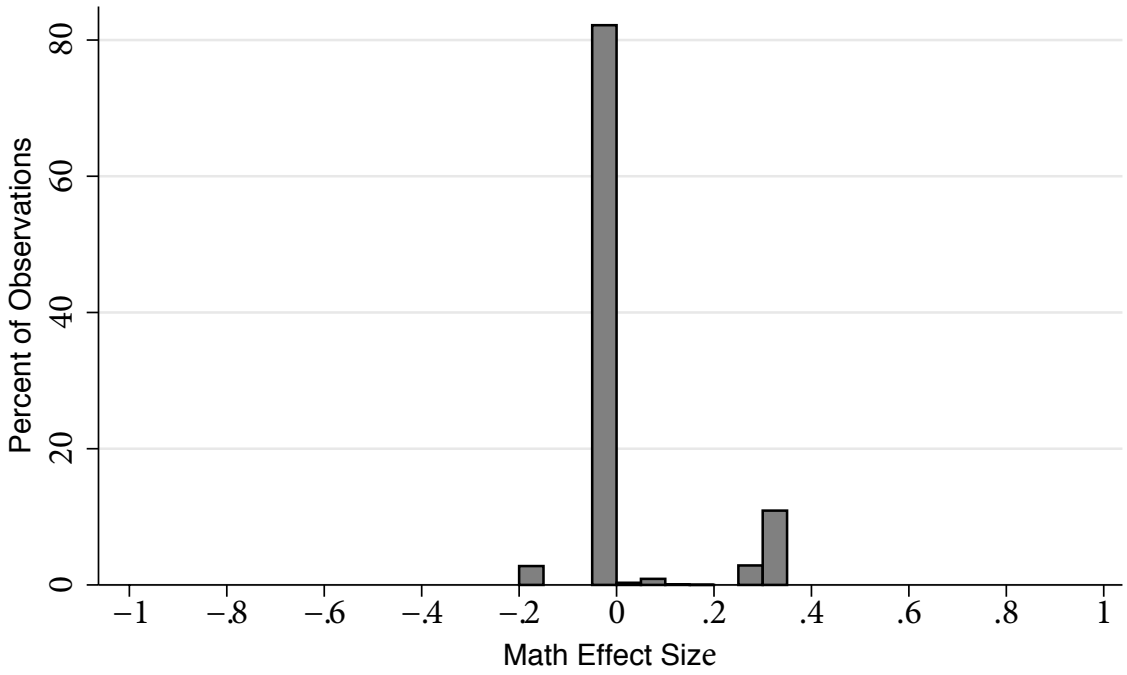
find significantly positive results for elementary math. Another 10% of the studies find positive but not significant results. This implies that only three out of the ten studies of elementary school math are finding negative charter school effects. However, two of these three studies finding negative effects cover, by a substantial amount, the largest numbers of students. The number of observations in these two studies is 1.6 million and 1.7 million, while the average number of observations in the remaining studies is only 22,000. Figures 15 and 16 demonstrate this visually. Comparing the unweighted math results at the elementary school level (Figure 15) with the weighted results (Figure 16), we see that weighting makes the results look a lot less positive. Our formal analysis in the preceding section derives optimal weights, and because this approach suggests that most of the variation across studies is real, those estimates are closer to the unweighted approach than to the extreme approach which weights results by sample size.

Figure 15. Distribution of Effect Sizes for Elementary School Math Studies, Treating Each Estimate Equally



This suggests that observers who tend to give more credence to the larger studies must be cautious about saying that charter schools are outperforming traditional public schools in elementary school math. On the other hand, observers who tend to distrust larger studies (for example, those who doubt that the methods used in large studies fully account for omitted variable biases) might not be dissuaded from seeing charter schools in a positive light.

Figure 16. Distribution of Effect Sizes for Elementary School Math Studies, Weighting Each Estimate by Number of Observations



The story in elementary school reading is similar to the story in elementary school math, but generally more favorable for charter schools. Again, the larger positive effects are found in the smallest studies. However, for elementary school reading one of the two large weight studies finds small positive effects, and the other finds only a very small negative effect. Therefore weighting by the number of observations only causes charters to look somewhat worse, not greatly worse. Furthermore, the downside (the study finding the most negative estimate of the charter school effect) in elementary reading is not worse than -0.1 standard deviation units, and is not significantly different than zero. The upside for math looks to be greater than for reading in elementary schools, as we also found in middle schools.

Overall, elementary charter schools appear to be outperforming traditional schools. However, we must be cautious in making this statement as two large studies find negative results for math.

The pictures in high schools tell us the most mixed story of all the single grade-span studies (elementary, middle, and high). This level has the fewest studies, with only seven estimates for high school reading, and eight estimates for high school math. Moreover,

one of the studies (CREDO's national study that pools all of the data available to them) covers over nine times as many students as the rest of the studies of reading combined. For math, this ratio is 2.75 times, due to the fact that there is an additional large study of math that does not study reading. Therefore, the overall pictures are extremely sensitive to whether we weight the estimates by the number of students studied. The pictures also differ between math and reading much more than at the elementary and middle levels.

The unweighted pictures show that for math, five studies find negative results and three find positive results, and that most of these effects are not large in magnitude. When we weight by sample size, almost all of the mass in the histograms is now found in the negative region. This is due to the fact that the three positive results were found in the three smallest studies. The picture is much more favorable for reading than for math. For reading, five studies find positive results, and only two find negative results. One of these is the large weight CREDO national study. Figure 18 shows that when we do not include the CREDO national study in the picture, in contrast to when we do in Figure 17, the weighted histogram in high school reading shifts toward the right, and the picture looks much more favorable for charter schools.

Figure 17. Distribution of Effect Sizes for High School Reading Studies, Weighting Each Estimate by Number of Observations

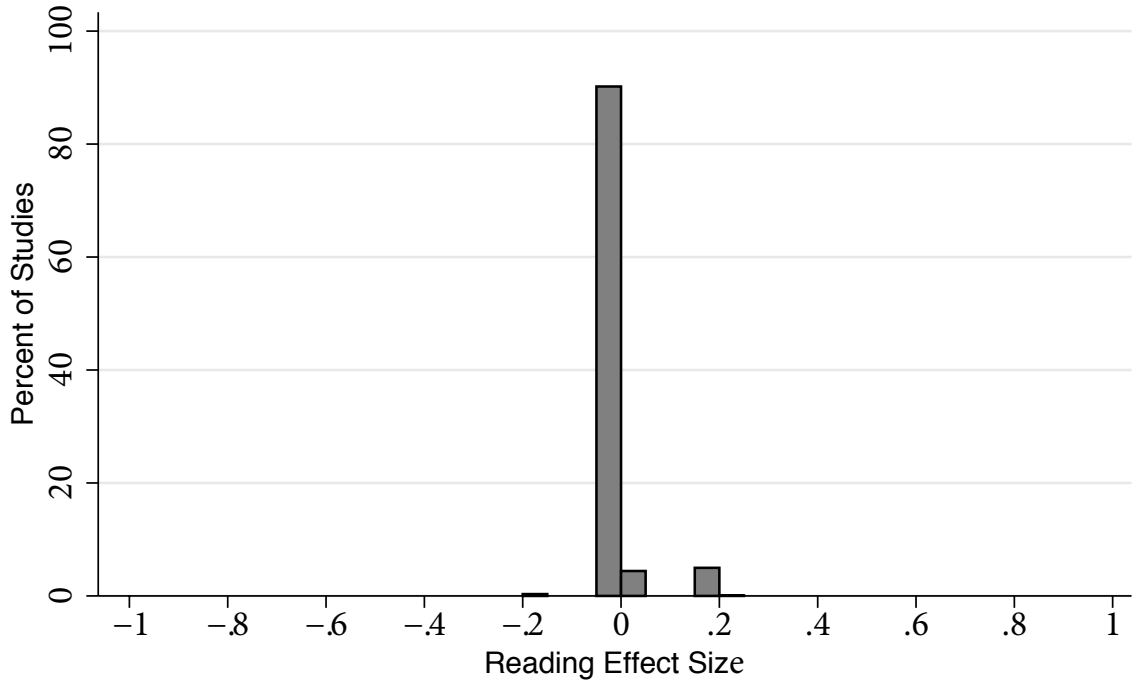
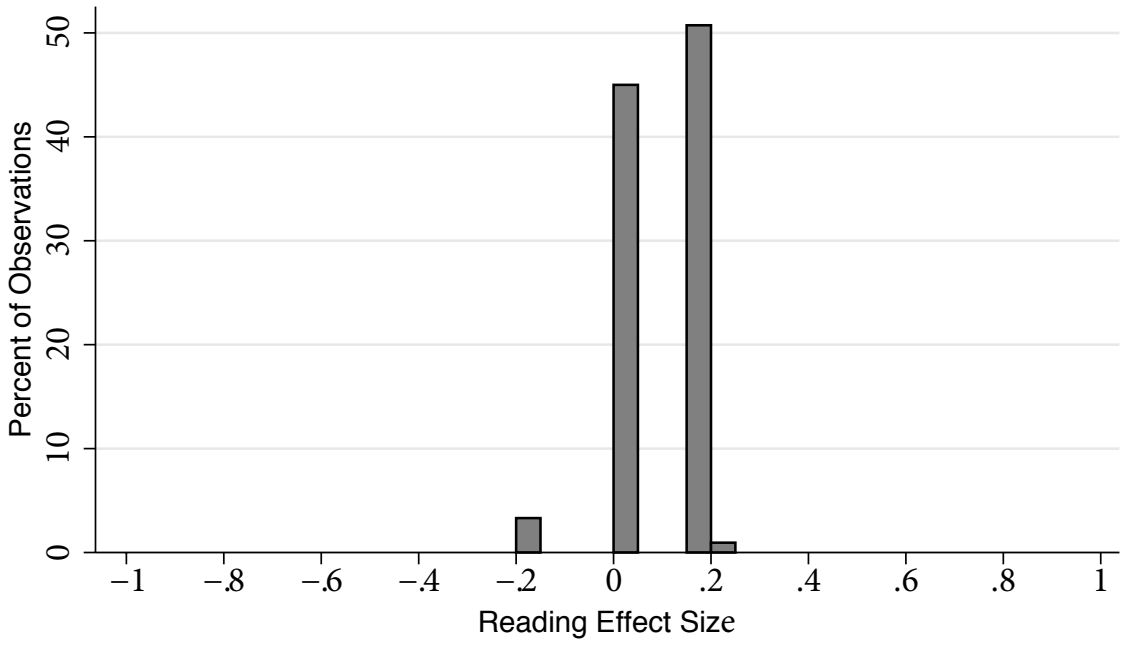


Figure 18. Distribution of Effect Sizes for High School Reading Studies, Weighting Each Estimate by the Number of Observations, Excluding CREDO National Estimate



Now we discuss the elementary/middle combination (EM) estimates. We have fifteen estimates each in both reading and math, from five studies of this type. Two of the studies examine multiple areas. In contrasting these results to the results discussed previously, we find that many more of the EM studies find significant (positive or negative) results than studies at the elementary (E), middle (M), or high school (H) levels. While typically being more significant however, the estimated effects are also much more moderate than studies at the elementary (E), middle (M), or high school (H) levels, whether the effects are positive or negative. For math, there are no studies that find effects in magnitude greater than 20% of a standard deviation, whether positive or negative. For reading, this moderation of results is even more dramatic, with no studies finding effects in magnitude greater than 10% of a standard deviation, whether positive or negative.

Weighting at this level does not change the overall impressions of charter school effectiveness, because most of the estimates come from studies with relatively large samples. There are roughly equal numbers of studies finding negative and positive results. Figures 19 and 20 show that the effects tend to be more moderate in size than those found in purely elementary (E), middle (M), or high school (H) studies, and that again there seems to be more variation in math effects than reading effects, and that effects look generally to be more positive in math than in reading.

Figure 19. Distribution of Effect Sizes for Combined Elementary and Middle School Reading Studies, Weighting Each Estimate by Number of Observations

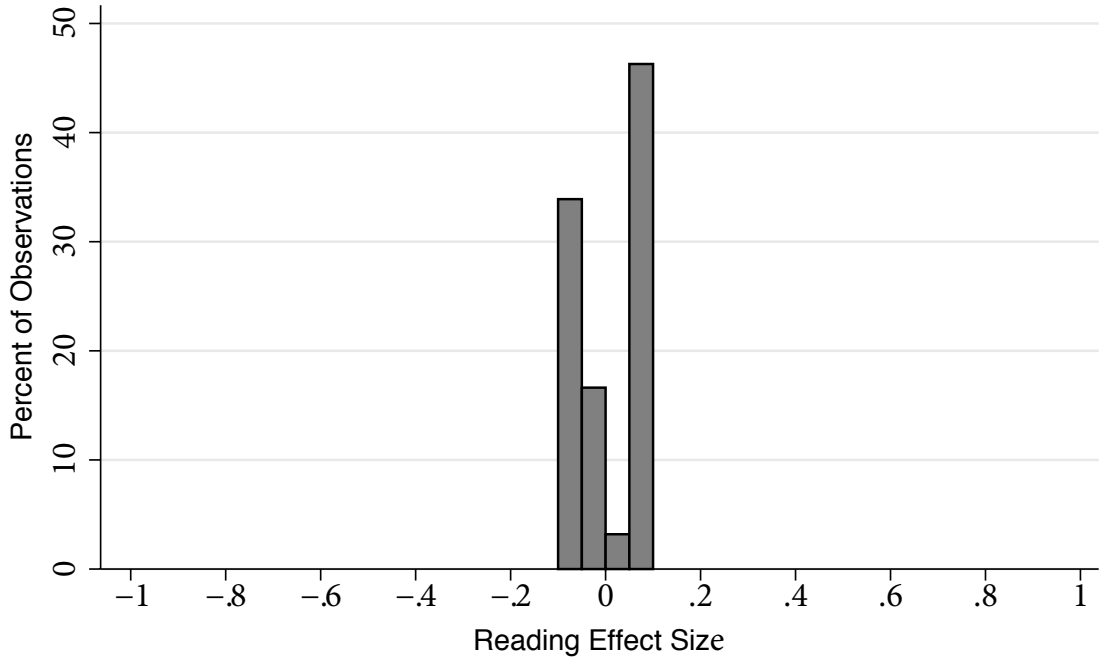
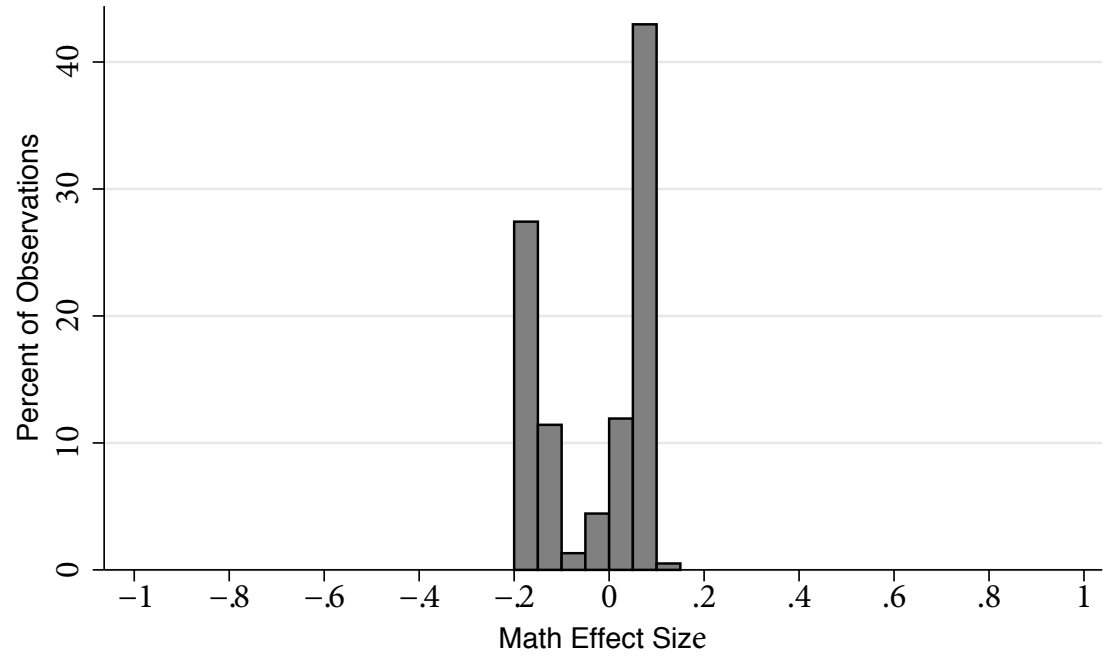


Figure 20. Distribution of Effect Sizes for Combined Elementary and Middle School Math Studies, Weighting Each Estimate by Number of Observations



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. These tend to be smaller studies than the studies finding negative results. Studies of combinations of elementary and middle school students are roughly balanced in 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. These figures are not shown because they essentially replicate the information contained in the earlier figures. Figures 21 and 22 show the histograms of the effect sizes found in these combined grade spans when estimates are weighted by the number of observations. The range of estimates in this case is small effects in reading, and a somewhat larger range of effects for math.

Finally, we discuss those studies studying all grade spans. Figure 23 shows that the distribution of effects in reading is mixed, with many studies finding both negative and positive results. However, the lower bound on these estimates is not large in magnitude. The study finding the most negative estimate for reading finds an effect size of -0.045 standard deviation units. The picture looks more positive for math. Figure 24 shows that many more studies find positive than negative results for math. Similar to the results presented of elementary (E), middle (M), or high school (H) studies, when we weight by number of observations studied, the pictures look slightly worse for charter schools than when we do not weight studies, particularly in reading. Overall, more studies of all grade spans look to be positive rather than negative in math, and more studies look to be negative rather than positive in reading.

Examining all of these results as separate parts of a whole, we conclude that overall charter schools look to be serving students well, at least in elementary and middle schools, and probably better in math than in reading. There appears to generally be more variation in the results for math than reading.

Figure 21. Distribution of Effect Sizes for Elementary, Middle, and Combined Elementary and Middle School Reading, Non-KIPP Studies Only, Weighting Each Estimate by Number of Observations

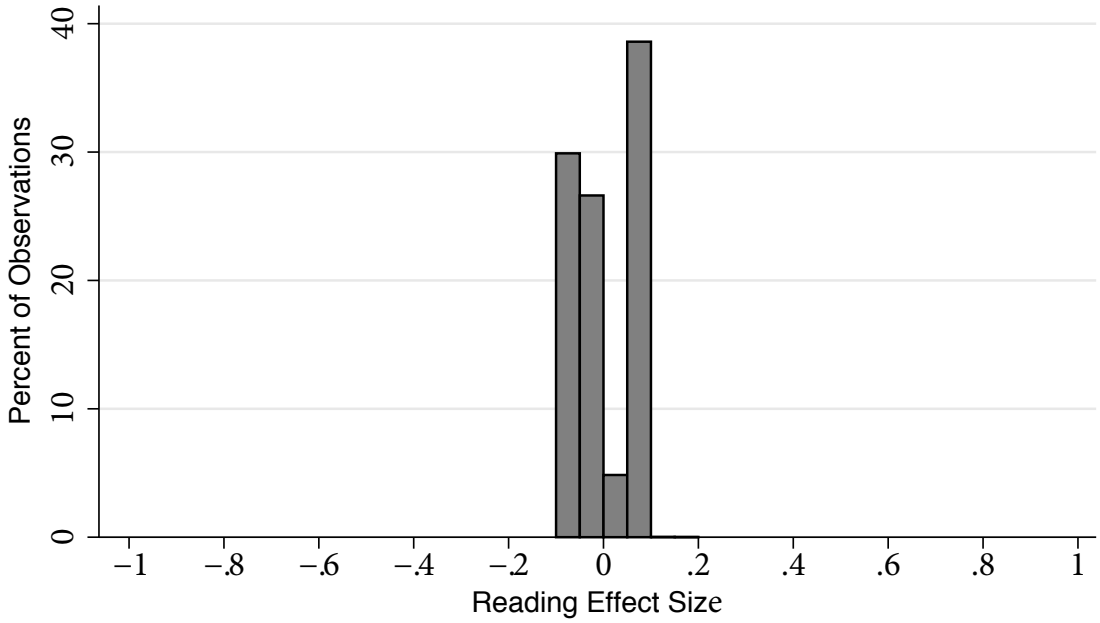


Figure 22. Distribution of Effect Sizes for Elementary, Middle, and Combined Elementary and Middle School Math, Non-KIPP Studies Only, Weighting Each Estimate by Number of Observations

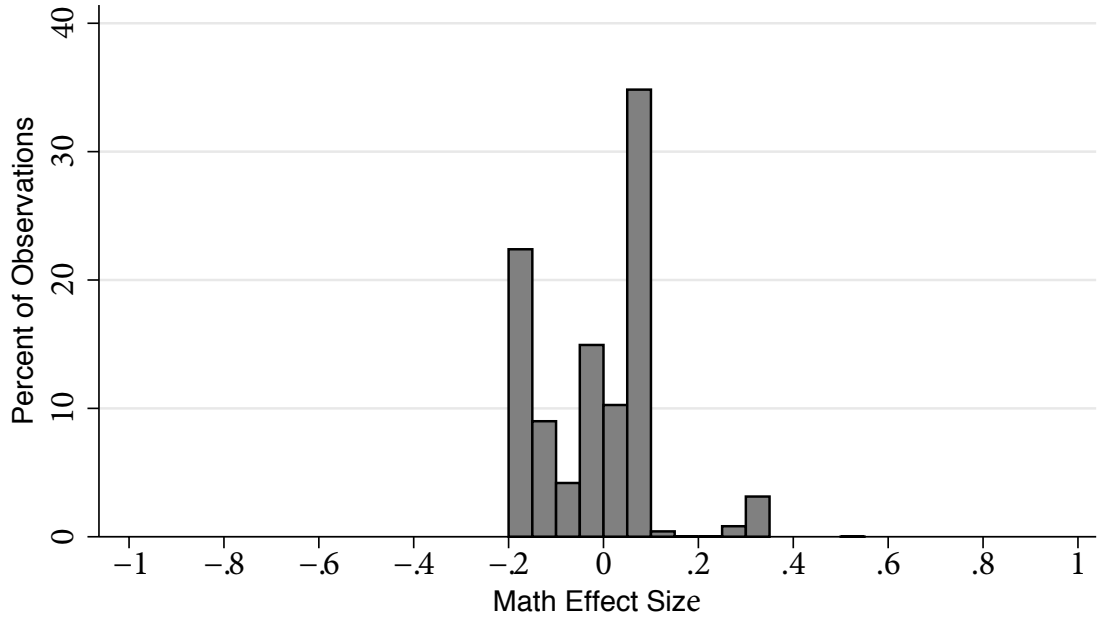


Figure 23. Distribution of Effect Sizes for All Grades Reading Studies, Weighting Each Estimate by Number of Observations

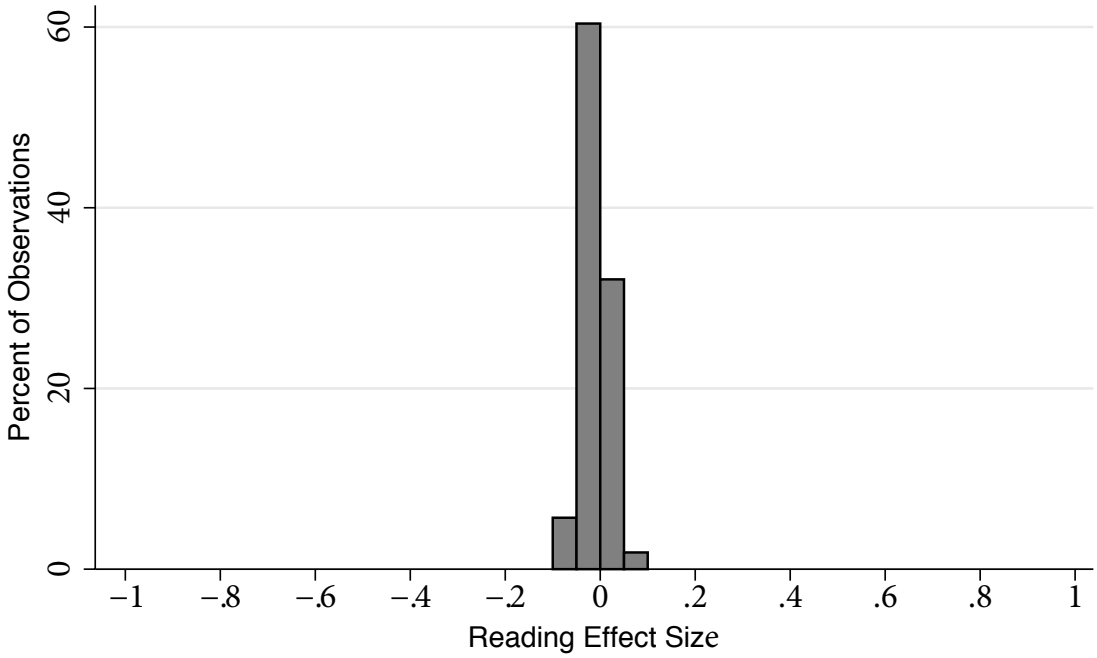
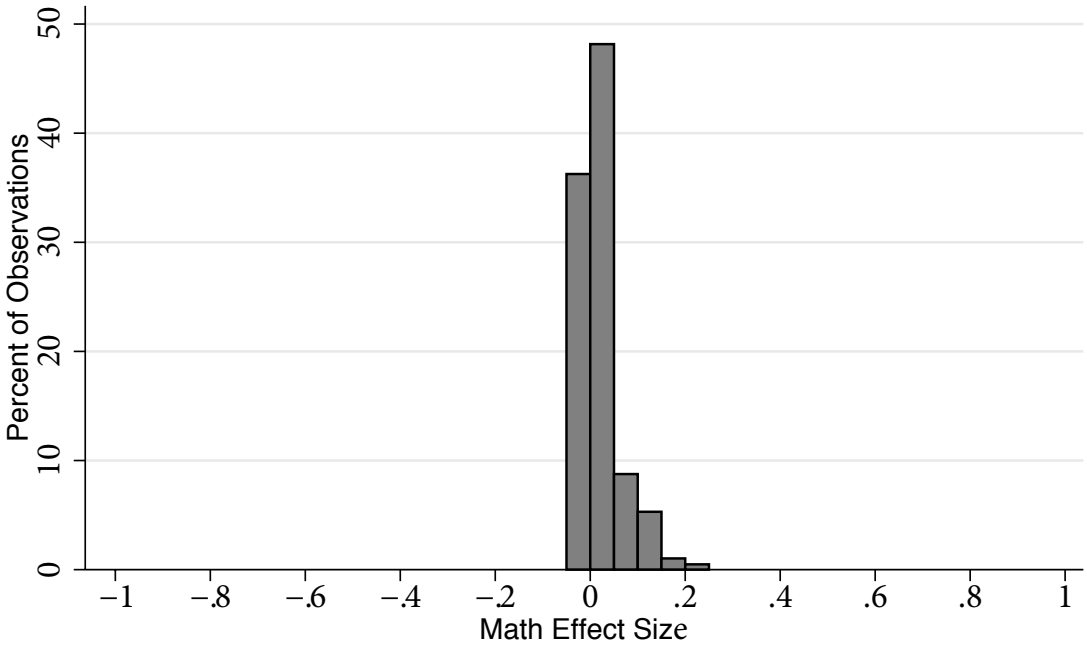


Figure 24. Distribution of Effect Sizes for All Grades Math Studies, Weighting Each Estimate by Number of Observations



DOES METHOD OF ANALYSIS MATTER?

It is worthwhile to check for variations in effect size across the method of study. Lottery-based studies should produce the least biased estimates conditional upon the lotteries producing similar-looking samples of lottery winners and losers, and subject to balanced attrition. But lottery-based studies may not be representative of charter schools that lack lotteries. Among the non-lottery methods, fixed-effect models use the student as his or her own control group, reducing bias, but students who stay in charter schools for the entire period of study do not contribute to the charter school estimate. Finally, matching methods such as propensity score matching or similar methods assume that students select into charter schools based on observable characteristics. To the extent that past test scores and the demographic variables with which researchers match students do not capture unobserved variations across students, the results from such studies could be quite biased statistically.

In the preceding analysis we compiled 91 distinct overall estimates (i.e., not an estimate of a subsample of students) of charter school effectiveness for math, and 87 estimates for reading. (The difference between the number of estimates for reading and math is due to one study offering 4 estimates for math only for Elementary, Middle, High, and All grade spans. All of the other studies contribute estimates for both reading and math.) Table 10 lists the numbers of estimates obtained using each of the four classes of methods—lottery, fixed effects, propensity score matching, or other matching. We can see in the table that more of the estimates use a matching method than use fixed-effects or lottery methods.

Table 10. Number of Math Estimates by Method of Estimation Type

GRADE SPAN	LOTTERY	FIXED-EFFECTS	PROPENSITY SCORE MATCHING	OTHER MATCHING
E (Elementary)	2	4	0	4
M (Middle)	5	3	29	2
H (High School)	2	3	1	2
Combined E/M	1	5	0	9
All Grade Spans	0	10	0	9
Total	10	25	30	26

Twenty-two of these distinct estimates are from one study of 22 KIPP schools. Another six are from one study of two cohorts at 3 KIPP schools. All 28 of these use propensity score matching. One more study is of an individual KIPP school, and it uses a lottery analysis. This means that 29 of 87 estimates we include are estimates for KIPP schools. The remaining 58 estimates come from 21 different reports. Two of these reports, one from RAND and another from CREDO, have substantially greater scope than the others—offering effect sizes for 7 locations and 16 locations respectively. These two large-scope studies apply different methods to obtain their estimates of charter school effectiveness. The RAND report favors the fixed-effect approach, while CREDO applies an approach utilizing a matching method somewhat similar to the synthetic control method. The remaining studies use mostly fixed-effects or lottery methods.

Due to the variation in the method of estimation it is worth discussing whether there may be a pattern in the results found according to methods used to obtain estimates. Table 11 breaks down the vote-count category results according to method used. We can see in this table for both reading and math that lottery studies and studies using propensity score matching find the most consistently positive results. Studies using fixed-effects and other matching methods find more mixed results, with the latter producing negative and significant results the most often. It is not possible to conclude whether the different methods tend to systematically produce different results, because these different methods are used to study different locations, and therefore different charter schools and students. For example, almost all of the studies using propensity score methods are of KIPP schools, and almost all find positive results. We cannot say whether these results are positive due to the propensity score methods (for example, if they fail to sufficiently account for positive selection bias into charter schools) or if they are positive because the KIPP schools studied are in fact outperforming their traditional school counterparts.

It is generally difficult to compare results from lottery and propensity score matching to other estimates, because both of these methods are typically employed in the literature to study an individual school. Fixed effects and other matching methods (in particular the matching of individual students employed by CREDO) are commonly used in the cases of larger samples. By looking closely at the few locations where we do have multiple methods employed to study a geographic sample, we can say more about the results found by fixed-effects methods compared to CREDO's method of matching.

Table 12 shows the sign and significance of the effects found in places where more than one method is employed to obtain an estimate. It is important to note that these differences cannot be solely attributed to the differences in method, because the time periods covered by the studies are not identical. For example, the table suggests that fixed-effects methods

Table 11. Vote-Counting Result Found by Each Method (Any Grade Span), Reading and Math

	LOTTERY	FIXED-EFFECTS	PROPENSITY SCORE MATCHING	OTHER MATCHING
READING TESTS				
Negative and Significant	0	6	1	7
Negative and Not Significant	2	4	5	2
Positive and Not Significant	2	8	7	3
Positive and Significant	6	3	17	14
MATH TESTS				
Negative and Significant	1	5	1	11
Negative and Not Significant	1	5	3	1
Positive and Not Significant	1	7	3	2
Positive and Significant	7	8	23	12

Table 12. Sign and Significance of Effects Obtained in Locations with Multiple Methods Used

		READING		MATH	
LOCATION	GRADE SPANS	FIXED-EFFECTS	MATCHING	FIXED-EFFECTS	MATCHING
Chicago	EM	–*	+	+	+*
Colorado (Denver)	A	+	+*	+*	+*
Florida	A	–	–*	+	–*
Ohio	EM	–*	–	–*	–*
San Diego	A	+*	+	+*	+
Arizona	A/EM	+ (A)	–* (EM)	+ (A)	–* (EM)
California	E/A	– (E)	+* (A)	–* (E)	–* (A)
North Carolina	EM/A	–* (EM)	+* (A)	–* (EM)	–* (A)
Texas	EM/A	–* (EM)	–* (A)	–* (EM)	–*(A)

NOTE: –* indicates negative and significant, – indicates negative and insignificant, + indicates positive and insignificant, +* indicates positive and significant charter school effect. Fixed-effects and matching methods do not study exactly the same time period, though some years overlap. The first five entries study the same grade spans, while in the latter four entries the grade spans studied are not the same. Lottery estimates are available for Chicago, but they are for grade spans elementary and middle only, while the fixed-effects and matching studies are for elementary and middle combined.

find North Carolina charter schools to be underperforming in reading, while matching methods find them to be outperforming. However, the fixed-effects methods were derived from studying students tested between 1996 and 2002, while the matching methods were derived from sample years 2003 to 2007. It is possible that North Carolina charter schools improved in teaching reading between these two periods. Moreover, in some cases, one estimate may be for a grade span that is Elementary and Middle combined while another is for All grade spans combined. Again using the example of North Carolina, the fixed-effects study looked at students in elementary or middle school, while the matching study included students in all grades. The first five entries are consistent in their grade-span selection, though not time periods (though they overlap). We can see here that generally the methods do not contradict each other in finding a negative and significant result using one method and a positive and significant result using another. There are some cases where one method finds a negative result and the other finds a positive result, but in each of these instances (excluding the North Carolina case discussed above) one of the findings is not significant. The magnitudes of the effects found differ widely in some cases between methods of analysis.

One of the studies does offer enough information to compare lottery estimates to non-lottery estimates on the same sample. In a study of Boston's schools, Abdulkadiroglu et al. (2009) include both estimates obtained from lottery analysis as well as estimates obtained from standard regression with demographic controls and baseline test scores (what they call the "observational study"). The estimates they found using these two methods were similar in sign and significance, though somewhat larger in magnitude in math for the lottery study than for the observational study. The observational study tends to have smaller standard errors.

Similarly, Betts, Tang, and Zau (2010) use a variety of non-experimental estimation methods commonly used in the literature to see which, if any, could approximate the lottery-based study by McClure et al. (2005) of the Preuss School at UCSD. Betts et al. find that a variety of methods come close to the lottery-based results as long as one controls for students' gains in achievement, rather than estimating models of levels of achievement.

There is now a third study that attempts to replicate lottery-based results. In a study we read after we had closed our search for new charter school achievement studies, CREDO (2010) applied its method of matching individual charter school students with one or more similar students in traditional public schools to a study of New York City charter schools. This CREDO study finds quite large positive effects of attending a charter school in that city, quite close to the lottery-based findings by Hoxby, Murarka, and Kang (2009). The effect sizes for math were the same in the two studies, at 0.12. For reading, Hoxby,

Murarka, and Kang (2009) report an effect size of 0.09, compared to an effect size of 0.06 reported by CREDO (2010) in its replication study.

It is notable that the CREDO method came close to the experimental estimates of Hoxby, Murarka, and Kang (2009), as the latter find some of the larger effect sizes in the entire literature, while the CREDO studies, as documented earlier, tend to report lower effect sizes than other studies. The lack of discrepancies signal that the effects found in New York City are quite robust and large relative to the charter school effects in most other locations. It also signals that the choice of method may not be as important as generally believed, as long as value-added methods are being used.

Overall, it appears as long as baseline test scores are controlled for (i.e., studies employ a growth-based student-level analysis), the specific method of analysis employed will not severely impact conclusions.

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.[10]

Educational Attainment

A central problem in analyzing years of education, whether a student graduates from high school, or enters college, is that we observe a person's (final) level of education only once. Lottery data are especially useful in this instance because we cannot use student fixed effects to compare a student with himself or herself in such a case.

As far as we are aware, there has been only one lottery-based study of the effects of charter schools on educational attainment, and that study examines only one California charter school. McClure, Strick, Jacob-Almeida, and Reicher (2005) utilize admission lotteries at the Preuss School at UCSD to examine the effect of winning a lottery on student achievement and educational attainment. They did not find big differences in test scores between lottery winners and losers, but they did observe large differences in a variety of measures of educational attainment. First, they studied how many college preparatory courses the students completed, and found large differences emerging as early as grade 10, in favor of lottery winners.

The authors also surveyed lottery losers in the graduating class of 2005 (who had enrolled in traditional public schools in San Diego) when they reached grade 12. The survey found a striking gap in planned college attendance. Among the Preuss school attendees (the lottery winners), 90.3% were set to enroll in a four-year college in fall, and 9.7% were planning to enroll in community college. Only 66.7% of respondents from the group of lottery losers planned to attend a four-year college in the fall, a gap of about 23%.

An issue with this comparison is that just under two-thirds of students in the group that did not win the lottery replied to the survey. By assuming either that none of the non-respondents or alternatively that all of the non-respondents were intending to enroll in

10. For an extended discussion of this "non-achievement" literature, see Betts (2010) upon which this section is largely based.

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. Regardless, then, the lottery winners were more likely to enroll in college than the lottery losers at this school.

The remaining studies of educational attainment do not use lottery data and so potentially suffer from bias caused by omitted variables.

Zimmer et al. (2009) examine the association between educational attainment and charter school attendance in a variety of locations. One of the approaches they take to reduce the self-selection among charter students is to focus on students who attend a charter school in grade 8, then compare educational attainment within this subsample between students who later attend high school charter schools and those who attend traditional public high schools. Because of onerous data requirements, this analysis is limited to Chicago and Florida.

In Chicago, Zimmer et al. (2009) estimate that attending a charter high school is associated with a 7% increase in the probability of graduating from high school and a 10% increase in the probability of attending a community college or four-year college. The corresponding figures for Florida are 12-15% and 8%. The limitations of this method are that we cannot be sure that restricting the analysis to students who attended charter schools in grade 8 removes unobserved variations among students who come to different decisions about whether to attend charter public high schools.

Another perhaps more convincing approach implemented by these same authors uses measures of proximity to charter schools as instrumental variables to take into account students' endogenous choice of whether to attend a charter school. These models produced larger estimates. The probability of graduating from high school is predicted to rise when attending a charter high school by about 15% in Florida and about 32% in Chicago. The estimated changes in probability of attending a two- or four-year college are 18% and 14% in Florida and Chicago respectively.

Quite different results are obtained by Maloney (2005), who studies the probability that students in Texas in grade 10 in spring 2000 obtain high school diplomas or GED diplomas two years later. She reports that students in charter schools overall are less likely to obtain high school diplomas but more likely to obtain GED degrees. But it seems likely that the charter school "effect" in this study is contaminated by the use of instrumental variables such as grade repetition that are highly endogenous.[11]

11. The study uses a method somewhat similar to Two Stage Least Squares, but which differs because not all of the second stage variables are included in the first stage. Numerous instruments are added to the first stage, which models the probability that students attend a charter school. But these instruments, such as whether students are deemed at-risk, whether they were expelled and so on, measured in grade 9, would seem to bear a direct association to the probability of high school graduation, raising questions about the validity of the conclusions.

With the exception of this last paper, the papers all produce large positive estimated effects on educational attainment. But they cover a very limited area—one school in San Diego, and charters in Chicago and Florida.

Evidence on Attendance and Behavior

Imberman (2007) studies two behavioral outcomes: attendance and suspensions from school (combined with more serious disciplinary actions). Using data from an anonymous large urban school district, he finds significant reductions in student disciplinary infractions among those who attend charter high schools.

Imberman also models the percentage attendance rate. The baseline model shows no relation between charter school attendance and attendance rates. However, in models that also control for lagged charter school attendance, a small positive relation between attending a charter school two years ago and attendance in the current period arises.

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. There are several important cases of grade spans in which charter schools are outperforming or performing about as well as traditional public schools. Elementary school math and reading, middle school math and, only if we include the KIPP school estimates, middle school reading all exhibit this pattern of students performing better at charter schools than at traditional public schools. At the high school level, there is no overall significant effect of charter schools, but there is considerable heterogeneity, suggesting that in some locations charter high schools are outperforming, while in others they are underperforming.

One of the most important findings from our meta-analysis is the considerable heterogeneity in effect sizes across studies. We typically found that 90% or more of the variation across studies reflected true variation rather than statistical noise. This realization could have important consequences for how researchers study charter schools and achievement in the future. Note that our finding of heterogeneity is for the most part based on variation in effect sizes across geographic areas (although some also derives from variation across studies of a single geographic area, in some cases at different times). It could well be that the variation across charter schools within a geographic area could be even larger. The CREDO studies of individual states provide some evidence that this is indeed the case, as they show histograms of effect sizes for individual schools.

Our analysis led to some clues as to sources of variation in the effects of charter schools. Charter high schools are not performing as well as charter schools at lower grades, at least in the small number of locations for which data are currently available. Analysis of the subsample of reports that exclusively study urban schools suggests larger effect sizes than for all charter schools in almost all cases. Boston's charter middle and high schools and New York City's charter schools are producing achievement gains far larger than are charter schools in most other areas; we can now be confident that these large gains are not simply a result of the analysis method chosen by researchers studying different areas.

It will always be the case that policymakers will want to have 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 as well. With a few exceptions such as the lottery-based studies of a KIPP school in Lynn, Massachusetts by Angrist et al. (2010), of the Promise Academy of the Harlem Children's Zone by Dobbie and Fryer (2010), and of 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. 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. Not only would this database serve a public purpose, but it also would allow for more nuanced meta-analyses of characteristics of charter schools that are truly making a positive or negative difference for student achievement.

Appendix

Appendix Table A1. 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 TESTS OF COMBINED RESULTS	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 and E/EM/M	E, M, H
Angrist, Dynarski, and Kane	2010	Boston (1 KIPP school)	2006	2009	M	M and E/EM/M	M
Ballou et al.	2006	Idaho	2003	2005	E, M, H, A	E, M, H, A and E/EM/M	E, M, H, A
Betts et al.	2005	San Diego	1998	2002			E, M, H
Betts et al.	2010	San Diego	2001	2006	E, M, H, A	E, M, H, A and E/EM/M	E, M, H, A
Bifulco and Ladd	2005	North Carolina	1996	2002	EM	E/EM/M	EM
Booker et al.	2005	Texas	1995	2002	EM		EM
Buddin and Zimmer	2003	California	1998	2002	E	E and E/EM/M	E
CREDO	2009a	National	2001	2008	E, M, H		E, M, H
CREDO	2009a	Arizona, Arkansas, California, Chicago, Colorado (Denver), DC, Florida, Georgia, Louisiana, Massachusetts, Minnesota, Missouri, New Mexico, North Carolina, Ohio, Texas	varies	varies	EM (9 locations), A (7 locations)	E/EM/M (9 locations), A (7 locations)	EM (9 locations), A (7 locations)
Dobbie and Fryer	2010	NYC (1 school, Promise Academy in Harlem Children's Zone)	2004	2009	E, M	M	E, M
Gleason et al.	2010	National (29 schools)	2004	2008	M		M

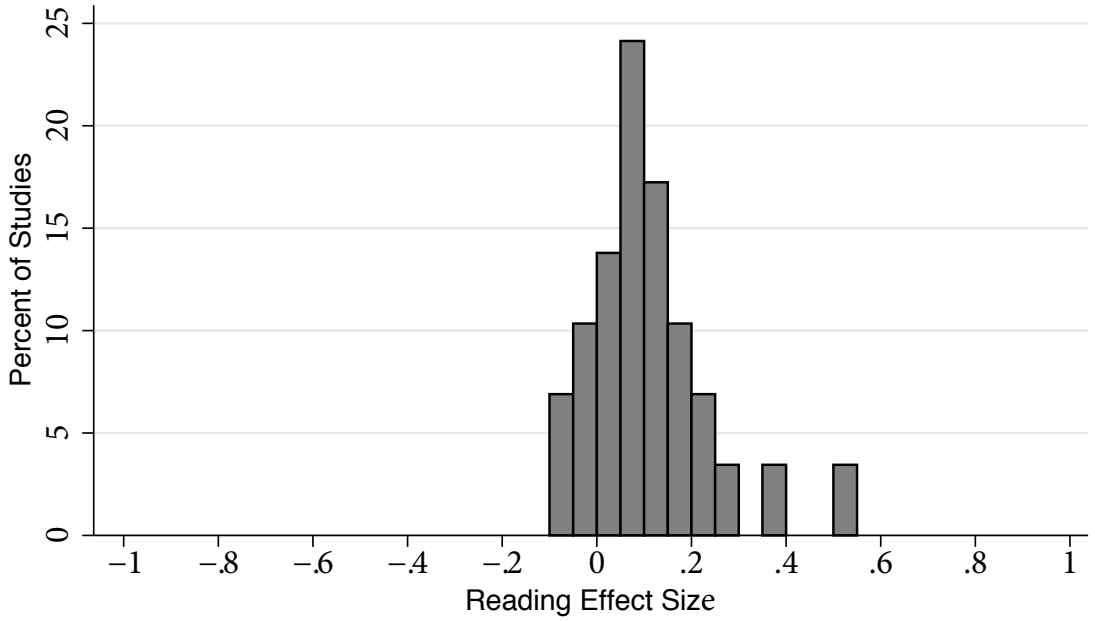
NOTE: E, M, H and A stand for analyses of elementary, middle, high schools, and all grades, respectively, and EM stands for the combinations elementary and middle.

Appendix Table A1. Details on the Studies Used in Any of Our Approaches (continued)

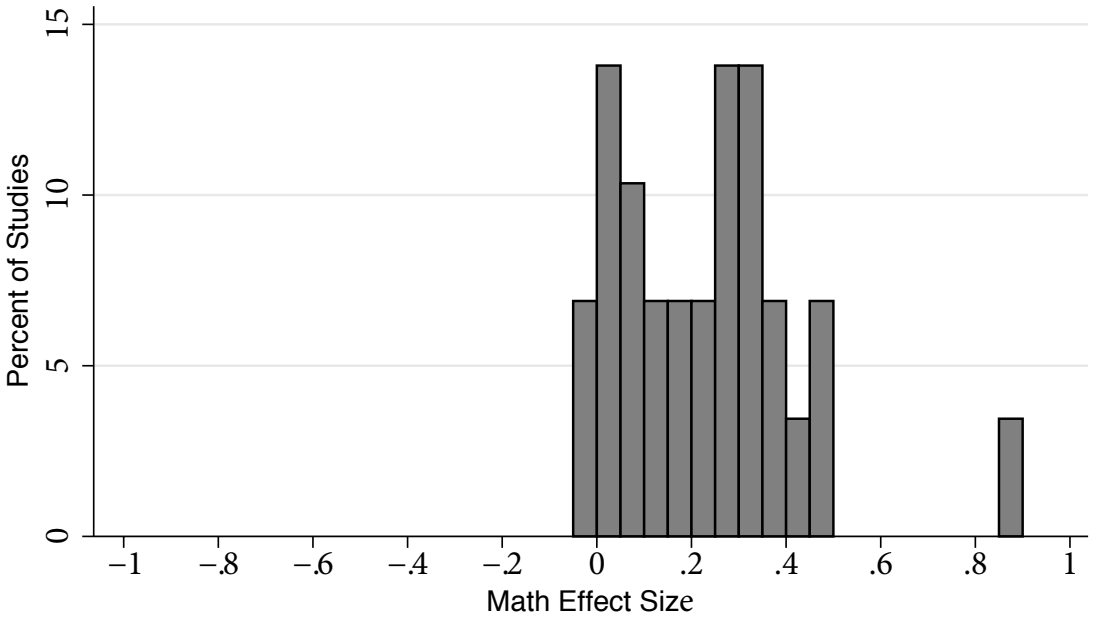
AUTHORS	YEAR PUBLISHED	NAME OF STATE OR CITY	FIRST YEAR OF DATA	FINAL YEAR OF DATA	GRADE SPAN(S) STUDIED	INCLUDED IN TESTS OF COMBINED RESULTS	INCLUDED IN META-ANALYSIS OF EFFECT SIZE, VOTE COUNTING STUDY, AND HISTOGRAMS
Gronberg and Jansen	2005	Texas	2003	2004	M, H	M, H	M, H
Hoxby and Murarka	2007	NYC	2004	2006	E	E	E
Hoxby, Murarka, and Kang	2009	NYC	2000	2008	EM	E/EM/M	EM
Hoxby and Rockoff	2005	Chicago	2001	2004	E, M	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 and E/EM/M	E, M, H, A
Nicotera, Mendiburo, and Berends	2011	Indianapolis	2002	2006	A	A	A
Solmon, Paark and Garcia	2001	Arizona	1997	1999	A	A	A
Sass	2006	Florida	2000	2003	A		A
Tuttle et al.	2010	Anonymous (22 KIPP schools)	varies	varies	M		M
Woodworth et al.	2008	Bay Area (3 KIPP schools)	2003	2005	M	M and E/EM/M using 2004-2005 estimates	M
Zimmer et al.	2009	Chicago, Colorado (Denver), Milwaukee, Ohio, Philadelphia, San Diego, Texas	varies	varies	EM (3 locations), A (4 locations)	E/EM/M (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 the combinations elementary and middle.

Appendix Figure A1. Distribution of Effect Sizes for Middle School Reading, KIPP Studies Only, Treating Each Estimate Equally



Appendix Figure A2. Distribution of Effect Sizes for Middle School Math, KIPP Studies Only, Treating Each Estimate Equally



REFERENCES

Abdulkadiroglu, Atila, Joshua Angrist, Sarah Cohodes, Susan Dynarski, Jon Fullerton, Thomas Kane, and Parag 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.

Ashenfelter, Orley. 1978. "Estimating the Effects of Training Programs on Earnings," *Review of Economics and Statistic*s, 60, pp. 47-57.

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

Betts, Julian R. 2009. "The Competitive Effects of Charter Schools on Traditional Public School." In *Handbook of Research on School Choice*, Mark Berends, Matthew G. Springer, Dale Ballou, and Herbert Walberg (Eds.) New York: Routledge. 195-208.

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., 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 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.

Bifulco, Robert, and Helen F. Ladd. 2006. "The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina." *Education Finance and Policy*, Vol. 1, No. 1, Winter 2006, 50-90. http://www.mitpressjournals.org/toc/edfp/1/1.

Booker, Kevin, Scott M. Gilpatric, Timothy Gronberg, and Dennis Jansen. 2004. "Charter School Performance in Texas." University of Tennessee, Knoxville.

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 Ltd.

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: 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?" National Bureau of Economic Research 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. Downloaded from credo.stanford.edu.

CREDO. 2009b. *Fact vs. Fiction: An Analysis of Dr. Hoxby's Misrepresentation of CREDO's Research*. Stanford, CA: CREDO. Downloaded from credo.stanford.edu.

CREDO. 2009c. *CREDO Finale to Hoxby's Revised Memorandum*. Stanford, CA: CREDO. Downloaded from credo.stanford.edu.

CREDO. 2010. *Charter School Performance in New York City*. Stanford, CA: CREDO. Downloaded from credo.stanford.edu.

Dobbie, Will, & Fryer, Roland, Jr. 2009. *Are high quality schools enough to close the achievement gap? Evidence from a social experiment in Harlem*. National Bureau of Economc Research Working Paper 15473. Cambridge, MA: National Bureau of Economic Research.

Gleason, Philip, Melissa Clark, Christina Clark Tuttle, and Emily Dwoyer. 2010. *The Evaluation of Charter School Impacts: Final Report*. NCEE 2010-4029. Washington, D.C.: 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*. 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*, Vol. 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, Vol. 3A*, O. Ashenfelter and D. Card, eds. Amsterdam: North Holland, pp. 1865-2097.

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

Higgins, J., Thompson, S.G., Deeks, J.J. and D.G. Altman. 2003. "Measuring Inconsistency in Meta-Analyses." *British Medical Journal*, Vol. 327, No. 7414, pp. 557-560.

Hoxby, Caroline M. 2009. *A Statistical Mistake in the CREDO Study of Charter Schools*. Stanford University manuscript, downloaded from 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." http://www.nber.org/~schools/charterschoolseval/.

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.

Hoxby, Caroline M., and Jonah E. Rockoff. 2004. "The Impact of Charter Schools on Student Achievement." http://www.educationnext.org/unabridged/20054/52.pdf.

Imberman, Scott. 2007. "Achievement and Behavior in Charter Schools: Drawing a More Complete Picture." University of Houston. http://www.class.uh.edu/faculty/simberman/imberman2007a_oct.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.

Maloney, Catherine. 2005. *The Effect Of Texas Charter High Schools On Diploma Graduation And General Educational Development (GED) Attainment*. Ph.D. Dissertation, University of North Texas.

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

Nicotera, Anna, Maria Mendiburo, and Mark Berends. 2009. *Charter school effects in an urban school district: An analysis of student achievement gains in Indianapolis*. Paper presented at the National Conference on Charter School Research at Vanderbilt University, Nashville, TN.

Sass, Tim R. 2006. "Charter Schools and Student Achievement in Florida." *Education Finance and Policy*, Vol. 1, No. 1, Winter 2006, 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." Goldwater Institute's Center for Market-Based Education. http://www.goldwaterinstitute.org/pdf/materials/111.pdf.

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, D.C.: Mathematica Policy Research.

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: 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.

The National Charter School Research Project (NCSRP) aims to bring rigor, evidence, and balance to the national charter school debate. For information and research on charter schools, please visit the NCSRP website at www.ncsrp.org. Original research, state-by-state charter school data, and links to charter school research from many sources can be found there.

center on reinventing public education

Center on Reinventing Public Education
University of Washington Bothell

425 Pontius Ave N., Suite 410
Seattle, Washington 98109 Seattle,
Washington 98103-9158
T: 206.685.2214
F: 206.221.7402

www.crpe.org

The Center on Reinventing Public Education at the University of Washington engages in research and analysis aimed at developing focused, effective, and accountable schools and the systems that support them. The Center, established in 1993, seeks to inform community leaders, policymakers, school and school system leaders, and the research community.