



Value-Added and Experimental Studies of the Effect of Charter Schools on Student Achievement



A LITERATURE REVIEW



Julian R. Betts and Y. Emily Tang



Value-Added and Experimental Studies of the Effect of Charter Schools on Student Achievement

A Literature Review

DECEMBER 2008

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

2101 N. 34th Street, Suite 195

Seattle, Washington 98103-9158

www.ncsrp.org



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

Our advisory board guides the selection and methodology of NCSRP research:

- Julian Betts, *University of California, San Diego*
- Susan Bodilly, *RAND Education*
- Anthony Bryk, *Stanford University*
- Lisa Coldwell O'Brien, *Coldwell Communications; New York Charter School Association*
- Abigail Cook, *Public Policy Institute of California*
- Jeffrey Henig, *Columbia University*
- Gisele Huff, *Jaquelin Hume Foundation*
- Christopher Nelson, *Doris & Donald Fisher Fund*
- Michael Nettles, *ETS*
- Greg Richmond, *National Association of Charter School Authorizers*
- Andrew Rotherham, *Education Sector; Progressive Policy Institute*
- Priscilla Wohlstetter, *University of Southern California*

CONTENTS

Abstract	1
Acknowledgments	2
About the Authors	2
Introduction	3
Methods and Practical Issues	5
Method 1: Testing Whether Charter Schools in Any Study Underperform or Outperform Traditional Public Schools	8
Method 2: Patterns of Statistical Significance	10
Method 3: The Distribution of Effect Sizes	16
Conclusion	26
References	28

FIGURES

Table 1. Details on the Studies Used in Any of Our Three Approaches	6
Table 2. Tests for Existence of Positive or Negative Effects of Charter Schools Among All Studies	9
Table 3. Percentage of Reading Results by Level of Statistical Significance and by Method of Weighting Studies	12
Table 4. Percentage of Math Results by Level of Statistical Significance and by Method of Weighting Studies	15
Table 5. Median Effect Sizes	17
Table 6a. Distribution of Effect Sizes, Unweighted	19
Table 6b. Distribution of Effect Sizes, Weighted by Number of Charter Schools	20
Table 6c. Distribution of Effect Sizes, Weighted by (Number of Charter Schools) * (Number of Years)	20
Figure 1a. Distribution of Effect Sizes for All Math Studies, Treating Each Estimate Equally	22
Figure 1b. Distribution of Effect Sizes for All Math Studies, Weighting Each Estimate by Number of Charter Schools	22
Figure 2a. Distribution of Effect Sizes for All Reading Studies, Treating Each Estimate Equally	23
Figure 2b. Distribution of Effect Sizes for All Reading Studies, Weighting Each Estimate by Number of Charter Schools	23
Figure 3. Distribution of Effect Sizes for Elementary Reading Studies, Weighting Each Estimate by Number of Charter Schools	24
Figure 4. Distribution of Effect Sizes for High School Math Studies, Weighting Each Estimate by Number of Charter Schools	25

ABSTRACT

We assess the literature that uses either experimental (lottery) methods or student-level growth-based methods to infer the causal impact of attending a charter school on student performance. We focus on the evidence on math and reading scores. We test that none of the effects is positive and, conversely, that none of the effects is negative. We find compelling evidence that charter schools underperform traditional public schools in some locations, grades, and subjects, and outperform them in other locations, grades, and subjects. We examine the distribution of effect sizes suggested by the existing work, using various methods to weight one study relative to another. We find a fairly wide range of estimated effects. Two cases in which charter schools appear quite frequently to outperform traditional public schools are elementary school reading and middle school math, although the effect sizes are small in the latter case. Conversely, charter schools often significantly underperform in high school reading and math. The 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 that we needed to incorporate their papers into our literature review. Dale Ballou, Richard Buddin, Caroline Hoxby and Jonah Rockoff, Chris Reicher and Larry McClure, Tim Gronberg, and Scott Imberman all provided key data that we needed for one or more of our analyses. We are also indebted to Laura Hamilton and Bryan Hassel for comments that substantially improved this paper.

ABOUT THE AUTHORS

Julian Betts is Professor of Economics and Chair of the Department of Economics at the University of California, San Diego. He also serves as a Research Associate of the National Bureau of Economic Research and an Adjunct Fellow at the Public Policy Institute of California (PPIC). He has written extensively on the link between student outcomes and measures of school spending, including class size, teachers' salaries, and teachers' level of education. More recently, he has examined the role that standards and school choice play in student achievement. His work includes a theoretical analysis of the impact of educational standards published in the *American Economic Review* (1998), the co-edited book *Getting Choice Right: Ensuring Equity and Efficiency in Education Policy* (Brookings Institution Press, 2005), and the co-authored book *From Blueprint to Reality: San Diego's Education Reforms* (PPIC, 2005). Current research includes a national study of magnet schools and a local study of career and technical education, both for the U.S. Department of Education. He is also a member of the National Charter School Research Project's Charter School Achievement Consensus Panel.

Y. Emily Tang is a Faculty Fellow in Economics at the University of California, San Diego. She obtained her Ph.D. in Economics from UCSD in 2007. Her studies have focused on social issues, including education, crime, and racial tensions. She is a co-author of *Does School Choice Work: Effects on Student Integration and Achievement* (Public Policy Institute of California, 2006) and is also co-author of a forthcoming book chapter illustrating how the choice of empirical method can lead to sharply varying estimates of the effects of charter schools on achievement.

INTRODUCTION

The volume of research on charter schools and achievement has mushroomed over the last half decade. However, the majority of studies has used relatively unsophisticated “snapshots” of student achievement at a single point in time, or has instead looked at changes in test scores in a given grade over time without accounting for the fact that a school enrolls different students in that grade in different school years. Such methods can be quite misleading because charter schools do not attract “typical” students, and the demographic background of schools’ populations can fluctuate from year to year. A number of studies, both national and statewide, suggest that charter schools disproportionately attract minority students and students from less affluent backgrounds. Without taking these differences into account, academic studies may be prone to understating the benefits of attending charter schools.

The Charter School Achievement Consensus Panel (2006) documents these patterns, and argues that these “snapshot” approaches are unlikely to produce unbiased estimates of the causal effect of attending a charter school on a student’s achievement. The same paper argues that there are two sets of approaches that provide more accurate results. The first is to compare those who win and lose lotteries to attend a given charter school. Only three papers have used this approach to date, and the total number of charter schools studied in these three papers is just under fifty. The next best approach, argues the Consensus Panel, is to use one of several variations of value-added models. These models follow individual students over time and examine improvement in test scores over time. This approach is helpful because it takes into account a student’s past academic history. Although these approaches were selected as the two best, both are susceptible to various potential weaknesses.¹

1. 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. A secondary potential problem 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. The two primary weaknesses of the most rigorous form of the value-added model, the fixed-effect model, stem from the fact that identification comes from students who switch between charter and traditional public schools. In elementary schools, many students start in charter schools and do not switch, so that it is hard to extrapolate fixed-effect results to such students. A more general issue is that fixed-effect models can control for unobserved heterogeneity among students only to the extent that the heterogeneity is fixed over time. But students who switch between the two types of schools may have done so due to unobserved factors that evolve over time. Given the small number of lottery studies, we have not attempted to compare results between these models and the growth models. However, the patterns of significance and signs of estimated effects are roughly similar between these two approaches: both lean toward findings of zero or positive charter school effects.

This paper provides an up-to-date review of charter school studies that utilize either of these approaches—randomization based on lotteries, or taking into account a student’s past achievement through value-added modeling. It may seem controversial to some readers that we would focus on a subset that constitutes roughly one-third of the literature. However, Betts, Tang, and Zau (2007), using data from San Diego, show that non-value-added models produce quite different results than the more sophisticated value-added models, and that the changes in estimated effects of charter schools are consistent with the idea that the weaker approaches fail to take into account the relatively disadvantaged backgrounds of students who attend charter schools. Further, Betts, Tang, and Zau 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.

METHODS AND PRACTICAL ISSUES

We use several approaches to summarize the results. First, we test whether we can reject two hypotheses. The first is that the effects of charter schools are never positive, and the second is that the effects are never negative. In a second approach, traditional vote-counting methods show the number of studies that yield positive and significant, insignificant (either negative or positive), or negative and significant results. This method is transparent and easy to understand. Researchers have 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 we will show, in the charter school literature far more studies produce significant results than one would expect if small samples were biasing researchers toward concluding “no significant effects.” Our third approach is more direct: we study the range or “distribution” of effects regardless of statistical significance.

A recurring issue in this literature is that the existing papers are strongly clustered geographically. The clustering partly reflects the sharp variations in availability of high-quality student-level achievement data across states. In general, such “longitudinal” data from Texas, North Carolina, Florida, and individual districts in California have received the most attention. It could be misleading to use many studies of the same schools in the same years and to claim that those studies are independent of each other. On the other hand, we often see two or more studies examine charter schools in two different time spans that do not overlap or only very partially overlap. Such studies really do complement each other by providing new information. For this reason, we tend to include studies of the same area that use different time periods, while focusing on just one study in the case of two or more studies of virtually the same schools in the same time periods. For one case in which a multi-district study included all of the schools in a second smaller single-district study, with overlap between time periods studied, we chose the broader study. In a few cases we could not include a study because it did not report math and reading test scores separately, or because either the papers or the authors did not provide the information needed to convert results into effect sizes.

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

Table 1. Details on the Studies Used in Any of Our Three 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 VOTE-COUNTING STUDY	INCLUDED IN ANALYSIS OF EFFECT SIZES
Ballou et al.	2006	Idaho	2003	2005	E, M, H	E, M, H	E, M, H	E, M, H
Betts et al.	2005	San Diego	1998	2002	E, M, H	E, M, H	E, M, H	E, M, H
Betts et al.	2007	San Diego	2001	2006	E, M, H	E, M, H	E, M, H	E, M, H
Bifulco and Ladd	2006	North Carolina	1996	2002	EM	EM	EM	EM
Booker et al.	2004	Texas	1995	2002	EM	EM	EM	E, EM
Gronberg and Jansen	2005	Texas	2003	2004	M, H	M, H	M, H	M, H
Hoxby and Murarka	2007	NYC	2004	2006	E	E	E	E
Hoxby and Rockoff	2005	Chicago	2001	2004	E, M	E, M	E, M	E, M
Imberman	2007	Anonymous	1995	2005	A	A	A	A
McClure et al.	2005	San Diego	2003	2004	H	H	H	H
Miron et al.	2007	Delaware	2000	2005	E, M, H	E, M, H	E, M, H	E, M, H
Sass	2006	Florida	2000	2003	A	A	A	A
Solmon et al.	2001	Arizona	1997	1999	A	A	A	A
Zimmer et al.	2003	California	1998	2002	E, MH	E, MH	E, MH	E, M, H

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

We quickly realized that a review of the charter school literature can ask three quite distinct questions:

1. In the typical study, what is the effect of attending a charter school?
2. In the typical charter school studied, what is the effect of attending a charter school? (Here, by “typical,” we mean “when we average across the charter schools studied,” not “for a charter school with the typical curriculum and typical student demographics.”)

3. In the typical charter school and year studied, what is the effect of attending a charter school?

The first question, which we think many researchers at least subconsciously ask, averages results across studies. However, because studies vary widely in the number of charter schools in their samples, and the number of years studied, some papers contribute more to our general understanding of how charter schools are contributing to student achievement. For example, one study that examines 5 charter schools over two years inherently tells us less than another study of equal rigor that follows 200 charter schools over five years. Yet in answering the first question above, we weight these two hypothetical studies equally. The second and third questions are much more relevant to policymakers because they ask, in slightly different ways, how charter schools are faring in an average sense. To answer these questions we weight studies either by the number of charter schools included in each study, or the number of charter schools times the number of years of data. We will find that weighting in favor of studies that contain above-average numbers of charter schools sometimes yields different results than a more naïve unweighted review.²

In many cases, we needed to request additional data from authors in order to include their work in our literature review. The most common data requests we made were for standard deviation of test scores (within grade), so that we could translate results from diverse testing systems into effect sizes, and for the number of charter schools included in the analysis. Comparisons across papers would be far simpler if authors routinely included these statistics.³

2. One could also imagine weighting by the number of students in each charter school. Individual studies in fact do this, because charter schools with relatively more students within a study will contribute more to the estimate of the charter school coefficient than will charter schools with relatively fewer students. In theory, one could also weight across studies by the number of charter school students per study. But the number of students per charter school is not typically provided in existing studies, and further, it would take large variations in charter school size across studies for this reweighting to make much of a difference.

3. To make clear that we do not intend to browbeat our many colleagues, we ruefully admit that we had to re-analyze some of our own data to obtain the number of charter schools in specific regressions. We thank the many authors who speedily sent us data we requested. Another factor that sometimes arose was that the achievement metric was averaged across subject areas. For this reason, we could not use the results in Hanushek, Kain, Rivkin, and Branch (2007). Fortunately, we were able to use results from two other studies from Texas.

Method 1: 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: even 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.⁴ 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.

For Fisher’s method to be valid, the studies that are used in the test must be statistically independent. For this reason, 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 2 shows the probability that charter school effects are negative/zero or positive/zero for various combinations of studies. The top row 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 positive effects of charter schools is miniscule, below 0.0001. The same applies to the probability that there are no studies showing true negative effects. The 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.

4. See, for instance, chapter 3 of Hedges and Olkin (1985).

Table 2. Tests for Existence of Positive or Negative Effects of Charter Schools
Among All Studies

STUDIES THAT INCLUDE CHARTER SCHOOLS FROM THE GRADE SPANS:	NUMBER OF STUDIES	PROBABILITY OF NO POSITIVE EFFECTS	PROBABILITY OF NO NEGATIVE EFFECTS
READING			
All studies	14	<0.0001	<0.0001
Elementary	5	0.0098	0.8982
Elementary and Combined Elementary/Middle	7	<0.0001	0.0011
Middle	5	0.0253	0.0202
Middle and Combined Middle/High	5	0.0023	0.6724
High School	5	<0.0001	<0.0001
MATH			
All studies	17	<0.0001	<0.0001
Elementary	6	<0.0001	0.0028
Elementary and Combined Elementary/Middle	8	<0.0001	<0.0001
Middle	6	0.0020	0.5798
Middle and Combined Middle/High	6	0.0311	0.0466
High School	5	0.0767	<0.0001

NOTES: The columns showing probabilities show the p-value, or probability, that there are either no positive effects or no negative effects. The row that shows "Middle and Combined Middle/High" school studies has the same number of studies as the row for "Middle" school studies. The reason for this is that when we added the one combined middle/high school study (Zimmer et al. 2003), we removed Betts et al. (2005) to minimize the chance of non-independent tests.

The next row of table 2 strongly suggests that some elementary charter schools outperform in reading, and that none underperform. (More precisely, the probability of no negative effects is 89.8%. Conversely, the probability that none of the studies find a positive effect is only 1.0%.) For math, there is strong evidence that elementary charter schools both underperform and outperform, depending on the time and location.

As shown in the next line of table 2, when we add in two studies that include elementary *and* middle school students, there is strong evidence of both negative and positive charter school effects in both math and reading. The North Carolina study by Bifulco and Ladd (2006) drives the evidence of negative effects in reading.

When we examine studies that focus on middle schools alone, there is ample evidence for both positive and negative effects on reading. For math, we find evidence of positive effects, but there is a very high (58%) probability that none of the studies show any negative effects.

When we add to the middle school studies a California study by Zimmer et al. (2003) that combines middle and high school grades, and drop the overlapping results from Betts et al. (2005) from San Diego, we find a quite high probability of no negative reading effects, but conversely, the quite high probability of no negative math effects falls from about 58% to 5%. It is not clear which of these estimates is preferable: the former study includes data from Los Angeles and San Diego at the secondary level, but it combines middle and high school students, unlike the other studies which focus on middle school students alone.

When we examine high school studies by themselves, we find evidence of both overperformance and underperformance, except that there is some evidence for no positive effects of charter high schools on math.

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 two cases in which charter schools do not seem to underperform in any of the studies: elementary school reading and middle school math.

Method 2: Patterns of Statistical Significance

Armed with the knowledge that the literature in most cases provides statistically meaningful evidence of both positive and negative effects from attending a charter school, we next study the sign and level of statistical significance of the charter school variable in different studies.

We looked in each paper for the main estimate, and found that often papers provide separate estimates of effects of charter schools for elementary, middle, and high school students. Such papers contribute three estimates to our analysis. In other cases a paper will study just one grade span, and in a few cases the paper studies only overall effects of charter schools at “all” grades, that is, elementary, middle, and high schools. In this last case, we apply this estimate to our summary of what “all studies” show, but do not use them for our separate estimates by grade span.

For each study, we categorized the estimated charter school effects as either positive or

negative, and as either statistically significant or insignificant. By “significant,” we mean that there is less than a 5% chance that the estimated effect could have been as far away from zero if the true effect really were zero. Thus, we report each study as falling into one of four categories. It is useful to remember that, roughly speaking, if charter schools produce the same achievement gains as traditional public schools, then we should find about 2.5% of results are “negative and statistically significant,” and 2.5% are “positive and statistically significant.” We almost always found that far more results than 2.5% fell into each of these two categories.⁵ This is what the test of the last section detected: there is strong evidence in most grade spans and subject areas of locations in which charter schools outperform traditional schools, and of locations in which charter schools underperform.

Table 3 shows results for models of reading achievement. We aggregate the studies in three ways. The first method treats each result equally, and so provides an answer to the question: “In the typical study, what is the effect of attending a charter school?” The second and third columns instead give increased weight to the studies that include larger numbers of charter schools, and to those that have more observations of charter schools by year, respectively. In some cases, the way in which we aggregate the studies matters a great deal.

5. One interpretation of this finding is that, as we have stated, charters tend both to outperform and underperform traditional schools. A second possibility is publication bias, which would arise if “insignificant” results have a harder time making it into print. We don’t think this second interpretation drives the results, as “zero” results are just as relevant to policymakers as positive or negative results. Second, we have included many studies that, although vetted by conference attendees and others in the field, have yet to be formally published.

Table 3. Percentage of Reading Results by Level of Statistical Significance and by Method of Weighting Studies

	SIGN AND SIGNIFICANCE	(1) Unweighted	(2) Weighted by # of schools	(3) Weighted by (# of schools) * (# of years)
All studies	-/Significant	18.2	22.3	23.5
	-/Insignificant	18.2	22.7	20.9
	+/Insignificant	27.3	32.7	18.6
	+/Significant	36.4	22.3	36.9
Elementary	-/Significant	0.0	0.0	0.0
	-/Insignificant	33.3	41.7	52.7
	+/Insignificant	16.7	14.2	22.4
	+/Significant	50.0	44.1	24.9
Elementary And Combined Elementary/ Middle	-/Significant	12.5	19.3	19.8
	-/Insignificant	25.0	12.2	9.8
	+/Insignificant	12.5	4.2	4.2
	+/Significant	50.0	64.4	66.3
Middle	-/Significant	20.0	7.1	14.8
	-/Insignificant	20.0	1.6	2.6
	+/Insignificant	40.0	88.5	75.6
	+/Significant	20.0	2.7	7.1
Middle and Combined Middle/High	-/Significant	16.7	4.1	11.0
	-/Insignificant	16.7	1.7	2.4
	+/Insignificant	33.3	85.4	63.8
	+/Significant	33.3	8.8	22.8
High	-/Significant	20.0	85.6	52.0
	-/Insignificant	0.0	0.0	0.0
	+/Insignificant	40.0	2.7	8.0
	+/Significant	40.0	11.7	40.0

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 first panel of table 3 shows that when we combine all studies, over one-third of them find positive and significant charter school effects. Although this evidence in favor of positive effects weakens when we weight studies by the number of charter schools included per study (because some studies included relatively small numbers of charter schools), when we instead weight by the number of charter schools times the number of years for which data were available, the results are virtually identical to the unweighted results.

The second panel of table 3 shows results for elementary schools. No studies find negative and statistically significant reading effects, and two-thirds find positive effects, most of

which are statistically significant. The evidence of positive effects weakens somewhat when we give more weight to studies with more charter schools and school years, but the overall picture of outperformance by charter schools remains.

When we add in studies that combine elementary and middle schools, we now find that some studies do produce statistically significant evidence of negative effects, but, depending on the weighting method used, the evidence suggests statistically significant positive effects for anywhere from one-half to two-thirds of the sample.

Middle school results that weight all studies equally suggest a disproportionate number of significant results, both negative and positive. In the estimates that weight by the number of charter schools or by the number of charter schools times the number of years of data, over 80% of results are positive (but only about 3%-7% of all results are positive and statistically significant, depending on which weight we use).

When we added to the middle school studies estimates that combined both middle school and high school grades, disproportionate shares of studies were either positive and significant or positive but not significant. The share of positive and significant results roughly triples when we add the combination schools, at least when we weight the results by the number of charter schools or by the number of charter schools times the number of years of data.

At the high school level, a U-shaped pattern emerges, with the majority of studies showing significant and positive or negative results. This pattern becomes especially strong when studies are weighted by the number of charter schools or by the number of charter schools times the number of years of data. For example, in the latter case, 52% of studies find negative and significant effects and 40% find positive and significant effects.

Table 4 shows the corresponding results for models of math achievement. As with the studies of reading achievement, far more than 5% of studies appear in the two “statistically significant” categories. There is ample evidence that at least some charter schools outperform and others underperform. As revealed by the Fisher test in the last section, middle school math results stand out in that literally none suggest a negative and significant effect of attending a charter school.⁶ The U-shaped pattern we found for reading in charter high schools does not appear in the math results, but a similar pattern emerges in several cases for math: the “all studies,” elementary, and elementary

6. One possible exception is the aforementioned California study that combined middle and high school students, which did seem to obtain a negative effect. Again, it is unclear whether the results in that one study reflect the districts studied, or the combination of middle and high school students.

and combined elementary/middle school studies. So again, we see strong signs that in some locations charter schools underperform, and in other locations they outperform. This appears most strikingly in the elementary and combined elementary/middle school math results. For instance, when we weight by the number of charter schools times the number of years of data, 29% of studies suggest negative and significant effects, and 70% suggest positive and significant effects. Notably, the negative effects derives mainly from the North Carolina study and from two studies of California districts up through 2002, while the 70% positive and significant result derives from a slightly more diverse set of areas: Texas, Delaware, New York City, and San Diego (in a study covering 2002 through 2006).

In the middle and higher grades, the effect of weighting studies tends to lower the percentage of studies with positive and significant estimated effects. This pattern is most striking at the high school level. If we simply combine all studies, without regard to the number of charter schools or years of data covered, about 17% show positive and significant effects, and 17% show negative and significant effects. But when we weight studies instead by the number of charter schools and years of data covered, the percentage of negative and significant results mushrooms to 69% and the percentage of positive significant results declines to just under 4%. More than any other result, this finding highlights the notion that merely averaging across studies, some of which have a relatively tiny sample of charter schools, can produce misleading estimates of what is happening at the typical charter school in the combined studies.

Together, the findings that math effects in charter middle schools are never negative and significant, but very often negative and significant at charter high schools, present a puzzle, and, at the high school level, a source of concern.

An overall assessment of these patterns is that for most of the 36 calculations shown in tables 3 and 4, the majority of estimated effects of charter schools are positive. This imbalance is sometimes mild, but in other cases, such as for reading scores in elementary schools, and for math scores in middle schools, the literature strongly suggests that charter schools are outperforming traditional public schools. There are important exceptions. Charter high schools appear to underperform significantly in math. In several cases we find considerable evidence of both positive and negative effects of charter schools.

Table 4. Percentage of Math Results by Level of Statistical Significance and by Method of Weighting Studies

	SIGN AND SIGNIFICANCE	(1)	(2)	(3)
		Unweighted	Weighted by # of schools	Weighted by (# of schools) * (# of years)
All studies	-/Significant	24.0	25.4	27.9
	-/Insignificant	20.0	20.9	18.5
	+/Insignificant	24.0	31.9	16.9
	+/Significant	32.0	21.8	36.7
Elementary	-/Significant	28.6	35.2	46.4
	-/Insignificant	0.0	0.0	0.0
	+/Insignificant	28.6	12.4	7.7
	+/Significant	42.9	52.4	45.9
Elementary and Combined Elementary/Middle	-/Significant	33.3	29.9	28.8
	-/Insignificant	0.0	0.0	0.0
	+/Insignificant	22.2	4.1	1.6
	+/Significant	44.4	66.1	69.6
Middle	-/Significant	0.0	0.0	0.0
	-/Insignificant	50.0	77.6	45.2
	+/Insignificant	16.7	13.3	34.4
	+/Significant	33.3	9.2	20.4
Middle and Combined Middle/High	-/Significant	14.3	5.6	15.1
	-/Insignificant	42.9	80.4	42.5
	+/Insignificant	14.3	7.6	25.8
	+/Significant	28.6	6.4	16.6
High	-/Significant	16.7	87.6	68.7
	-/Insignificant	33.3	5.7	21.5
	+/Insignificant	33.3	5.1	6.0
	+/Significant	16.7	1.6	3.8

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.

Of course, we must be careful not to extrapolate wildly. Table 1 shows the relatively narrow geographic coverage of the value-added and lottery-based studies included in our review. At the high school level, our somewhat pessimistic conclusions for math derive from Texas, Idaho, Delaware, and a small number of large urban districts in California. Researchers obviously have huge and important frontiers yet to study.

Method 3: The Distribution of Effect Sizes

After confirming the statistical significance of both positive and negative differences in performance between charter schools and traditional public schools, it is natural to ask about the magnitudes of these differences. Our first approach toward understanding how large or small these effects typically are is simply to look at the median estimated effect size across studies. By effect size, we mean the estimated effect of attending a charter school for one year on test scores, measured as the proportion of one standard deviation of test scores in the given grade and year. A second approach shows the sizes of the effect at the 25th and 75th percentiles, and presents histograms of the effect size distributions.

Table 5 presents the median effect sizes by grade span and subject. The first column shows the unweighted medians. In this column, each estimate is given an equal weight regardless of how many schools or years are covered in the study. By giving greater weight to studies that examine more charter schools, the second column accounts for the fact that some studies cover only a few schools in a district while others cover an entire state. The median values in this column are calculated after weighting the estimates by the number of charter schools that the study covers. The third column additionally considers that some studies include only one year of gains in test scores across students, while others analyze performance over a much longer period, and weights by the number of schools times the number of years covered in the study. In this column, studies covering a large number of schools over a long period of time are given the most weight. Below each entry in this table is the number of observations, or the number of weighted observations.

Table 5. Median Effect Sizes

		(1) Unweighted	(2) Weighted by # of schools	(3) Weighted by (# of schools) * (# of years)
All studies	Math	0.0305 (19)	0.00519 (1,277)	0.00519 (6,044)
	Reading	0.0197 (16)	0.0175 (1,243)	0.0220 (5,976)
Elementary	Math	0.0863 (6)	0.0807 (300)	0.0807 (1,854)
	Reading	0.039 (5)	0.086 (288)	0.086 (1,830)
Elementary and Combined Elementary/ Middle	Math	0.0807 (7)	0.0807 (367)	0.0807 (2,256)
	Reading	0.0363 (6)	0.086 (355)	0.086 (2,232)
Middle	Math	.00519 (5)	.00519 (226)	.00519 (1,879)
	Reading	-.00460 (4)	.0220 (213)	.0220 (1,853)
Middle and Combined Middle/High	Math	0.00519 (5)	0.00519 (232)	0.00519 (1,927)
	Reading	0.00659 (4)	0.0220 (219)	0.0220 (1,901)
High	Math	-0.0206 (4)	-0.215 (190)	-0.0155 (369)
	Reading	0.0592 (3)	-0.163 (181)	0.0592 (351)

NOTE: Number of studies, number of represented schools, or number of represented schools times years in parentheses.

The first pattern apparent in table 5 is that, regardless of the weighting scheme, in most cases the median effect is positive. This supports the previous vote-counting analysis that there are far more positive findings than negative findings. The only exception is at the high school level, in which the median effect is negative for math scores regardless of weighting, and negative for one of the cases of reading. Again, this closely matches the patterns of significance discussed earlier.

While generally positive, the median effect sizes tend to be small. In all cases but two, the absolute value of the median effect size is less than 0.10, or less than 1/10 of a standard

deviation of a test score. This is true for all of the unweighted median effects, as well as the weighted effects when the number of schools times the number of years serves as weight. Again, the only exception is at the high school level, and only under one of the weighting schemes. When we weight by the number of schools only, the Gronberg and Jansen (2008) study, which finds large and negative effects of charter schools, dominates the sample. Weighting by number of schools times number of years downweights the influence of this study as it studies only one year. With the exception of this subset of studies, the differences using these two types of weighting schemes are relatively small. Only at the high school level do the two methods offer qualitatively different results.

Looking across grade spans, it appears that the only consistently sizable positive effects are found at the elementary school level. Regardless of weighting, the median effect size in elementary math is around 8% of a standard deviation of a test score. In elementary school reading, while the charter school effect is also consistently positive and around 8% of a standard deviation when using weights, it is only around half that size when looking at the unweighted median.

Some simple comparisons provide some perspective on whether an effect size of 0.08 is large or small. A student with median test scores—ranking 50th out of 100 students—would be predicted to move up to about the 47th rank out of 100 students after one year at a charter school. This is not a large change, but over several years of such gains it could be quite meaningful. 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.

The effect at the middle school level is fairly small (between 1% and 2% of a standard deviation) regardless of whether we are looking at math or reading and whether we are looking at unweighted or weighted medians. Charter schools do not seem to have sizable effects in either direction at the middle school level.

At the high school level, the findings are more mixed. In high school math, all weightings give a negative overall effect, but these differ in size enormously, from around 20% of a standard distribution to only 2% of a standard deviation. In reading, the differing weighting schemes actually have different signs on their median estimates, which in turn reflects the different signs on the charter school variable across studies. It seems that we are not able to confidently conclude much about the performance of charter schools in high school reading.

Looking at medians gives us a good idea of what the “typical” study finds. We are also

interested in looking at the range of estimates. We can do this by looking at different points in each distribution of estimates. Tables 6a-c additionally present the estimates at the 25th and 75th percentiles as well as the mean estimates. Again, we present unweighted estimates and also employ two different weightings. The entries in the first columns of this set of tables repeat the median values shown in table 5.

Table 6a. Distribution of Effect Sizes, Unweighted

		(1)	(2)	(3)	(4)
		Median	Mean	25th Percentile	75th Percentile
All	Math	0.0305	0.0199	-0.0257	0.0905
	Reading	0.0197	0.0180	-0.0165	0.0694
Elementary	Math	0.0863	0.110	0.0377	0.116
	Reading	0.039	0.0509	0.0336	0.086
Elementary and Elementary/ Middle	Math	0.0807	0.0718	0.00815	0.116
	Reading	0.0363	0.0266	-0.00591	0.086
Middle	Math	0.00519	-0.00359	-0.0553	0.0328
	Reading	0.00460	0.00173	-0.0473	0.0508
Middle and Middle/High	Math	0.00519	-0.00794	-0.0553	0.0110
	Reading	0.00659	0.00732	-0.0361	0.0508
High	Math	-0.0206	-0.0414	-0.120	0.0375
	Reading	0.0592	0.0356	-0.163	0.211

Table 6b. Distribution of Effect Sizes, Weighted by # of Charter Schools

		(1)	(2)	(3)	(4)
		Median	Mean	25th Percentile	75th Percentile
All	Math	0.00519	-0.0122	-0.0190	0.0305
	Reading	0.0175	-0.00460	-0.00591	0.0220
Elementary	Math	0.0807	0.0777	0.0807	0.0807
	Reading	0.086	0.0598	0.039	0.086
Elementary and Elementary/ Middle	Math	0.0807	0.0343	0.00815	0.0807
	Reading	0.086	0.0306	-0.00591	0.086
Middle	Math	0.00519	0.00551	0.00519	0.00519
	Reading	0.0220	0.0156	0.0220	0.0220
Middle and Middle/High	Math	0.00519	0.00321	0.00519	0.00519
	Reading	0.0220	0.0176	0.0220	0.0220
High	Math	-0.215	-0.177	-0.215	-0.215
	Reading	-0.163	-0.129	-0.163	-0.163

Table 6c. Distribution of Effect Sizes, Weighted by (# of Charter Schools) * (# of Years)

		(1)	(2)	(3)	(4)
		Median	Mean	25th Percentile	75th Percentile
All	Math	0.00519	0.00948	0.00519	0.0479
	Reading	0.0220	0.0153	-0.00591	0.039
Elementary	Math	0.0807	0.0661	0.0377	0.0801
	Reading	0.086	0.0602	0.0337	0.086
Elementary and Elementary/ Middle	Math	0.0807	0.0258	0.00815	0.0807
	Reading	0.086	0.0323	-0.00591	0.086
Middle	Math	0.00519	0.00819	0.00519	0.00519
	Reading	0.0220	0.0164	0.0220	0.0220
Middle and Middle/High	Math	0.00519	0.00582	0.00519	0.00519
	Reading	0.0220	0.0182	0.0220	0.0220
High	Math	-0.0155	-0.0964	-0.215	-0.0155
	Reading	0.0592	-0.0338	-0.163	0.0592

There is very little qualitative difference between the two weighted tables at the elementary and middle school levels. In comparing the unweighted to the weighted ranges, however, we see that weighting leads to significant shrinking in the range of effects. The unweighted ranges tend to be much wider than the weighted ranges, indicating that both the relatively large negative and large positive effects tend to come from smaller-weight studies. Put differently, the studies that find large effects in either direction study a relatively small number of schools over a relatively short period of time. In some cases, when using either weighting scheme, the 25th and 75th percentile estimates are identical, especially in the middle and high school samples. This indicates that the median effect size in those cases come from a large-weight study, one that studies a lot of schools over a long time period. It also reflects the relatively small number of studies in those grade spans.

With the exception of the case of high schools, the bulk of these studies appear to find small positive effects. Of the 36 estimates of the 75th percentile effects in tables 6a-c, only 3 are negative. Conversely, at the 25th percentile only 18 of 36 estimates are negative, and only 4 of 36 median effects are negative. At the 25th percentile, the lower extreme that we present, the most negative effect is 5% of a standard deviation in middle school math, and this is replaced by a tiny positive effect when weights are applied. In high school, the picture is a bit more mixed, as discussed earlier.

Figures 1 and 2 show the histograms for effect sizes in all of the studies across grade spans, for math and reading respectively. In comparing the unweighted to the weighted cases, we can see that in both cases of math and reading the large positive effects tend to diminish in importance when we weight by the number of schools studied. When we give more weight to studies that cover more schools, the fraction of observations in the large positive bins falls, and the fraction in the small positive bin rises. This is another way of seeing the typical shrinking of the distribution when weights are applied, as discussed previously and shown in table 6.

Figure 1a. Distribution of Effect Sizes for All Math Studies, Treating Each Estimate Equally

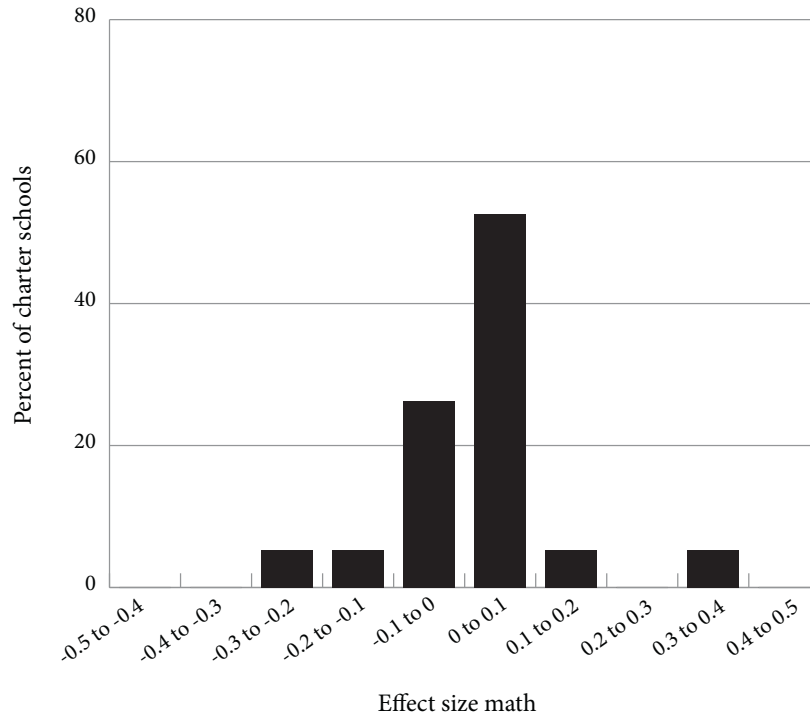


Figure 1b. Distribution of Effect Sizes for All Math Studies, Weighting Each Estimate by Number of Charter Schools

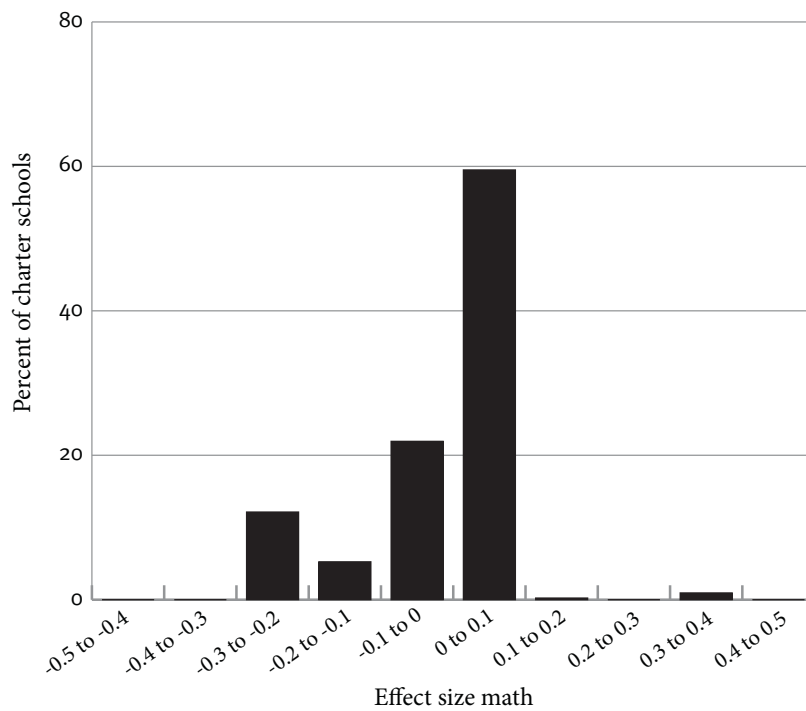


Figure 2a. Distribution of Effect Sizes for All Reading Studies, Treating Each Estimate Equally

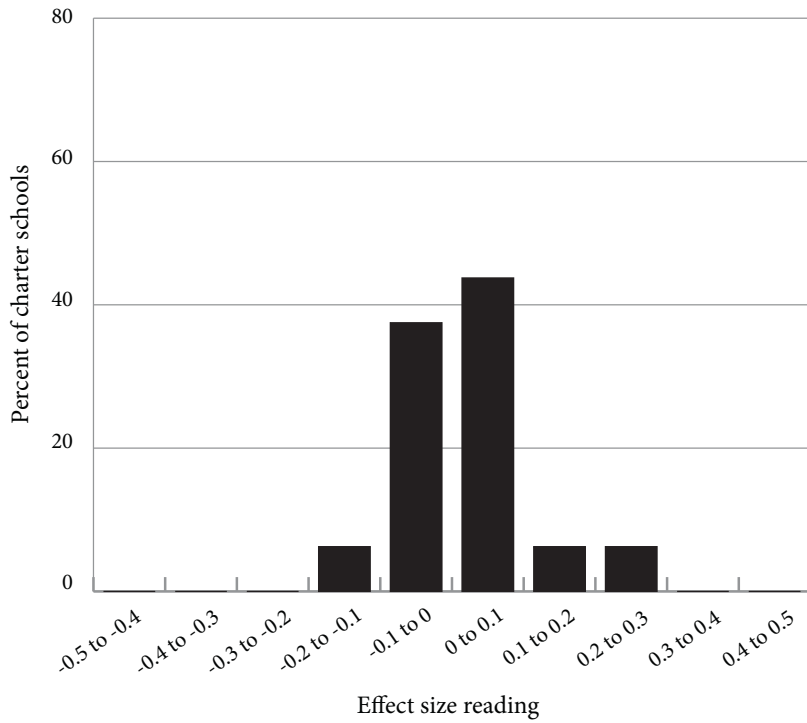
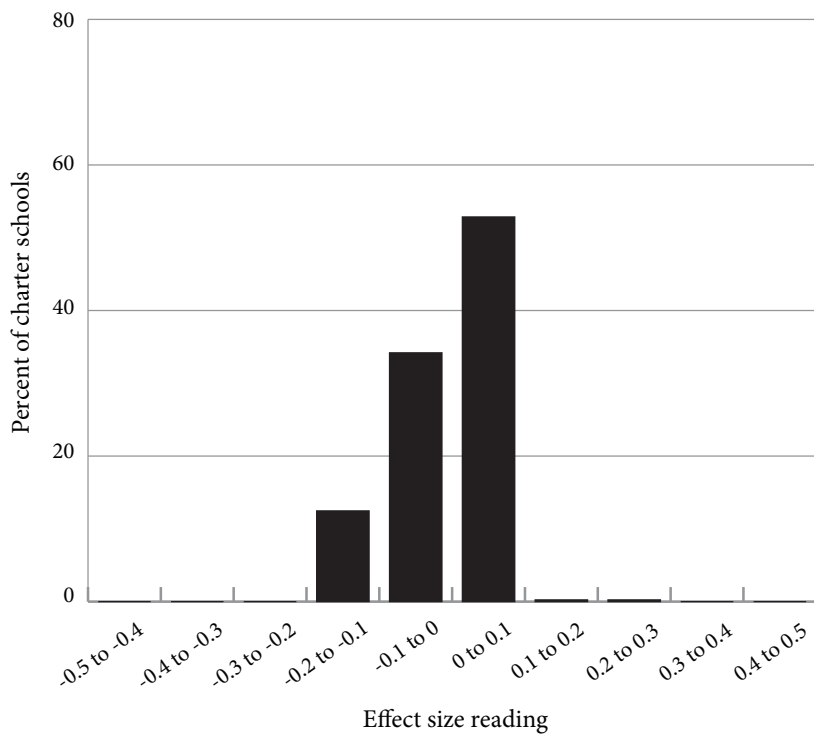
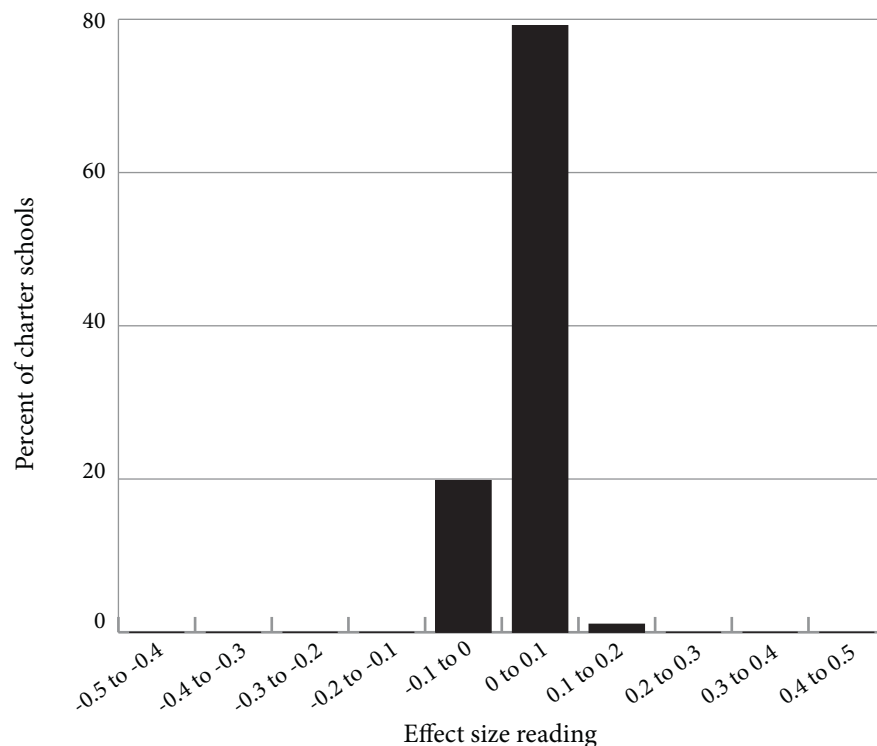


Figure 2b. Distribution of Effect Sizes for All Reading Studies, Weighting Each Estimate by Number of Charter Schools



We do not present the histograms for each grade span/subject combination due to space limitations. At the elementary school level, the general story told by the unweighted and weighted histograms in figures 1 and 2 hold—in both reading and math, weighting adds to the importance of the small positive effects. Figure 3 presents the weighted histogram for elementary school reading, re-affirming previous tests indicating that there is unlikely to be a negative charter school effect.

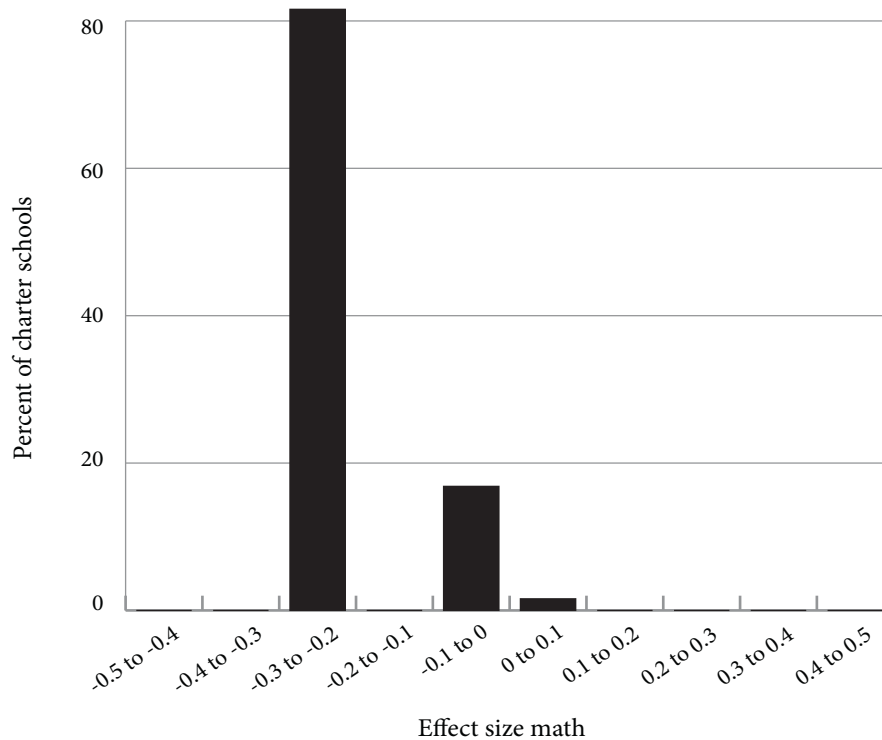
Figure 3. Distribution of Effect Sizes for Elementary Reading Studies, Weighting Each Estimate by Number of Charter Schools



There are no studies that find sizable effects in either direction at the middle school level. In the unweighted cases for both reading and math there are about the same number of studies on either side of zero, echoing the preceding vote-counting analysis. However, weighting distributes more importance to the small positive bin, in both reading and in math, meaning the negative effects in middle school in both subjects are found in low-weight studies. While the bulk after weighting ends up on the positive side of zero, most of these positive effects are quite small, as shown in table 5 and discussed above. We have already discussed the mixed picture at the high school level. For math, while there are studies finding large negative results, as well as small negative and small positive results,

weighting the studies using either method results in a virtual disappearance of the small positive effect. In high school math, charter schools are probably not outperforming traditional schools overall, echoing the result found using Fisher's inverse Chi-squared test. The school-weighted histogram for high school math in figure 4 demonstrates this clearly, with the bulk of the distribution on the negative side of zero.

Figure 4. Distribution of Effect Sizes for High School Math Studies, Weighting Each Estimate by Number of Charter Schools



In summary, the main lesson learned from looking at the overall distribution of the estimated effect sizes found by researchers is this: while studies have found both large negative and large positive effects, most of the effects that are relatively large in magnitude are found in studies of relatively small scope. The larger-weight studies typically find small positive effects.

CONCLUSION

Our review of the literature on the effect of charter schools on math and reading achievement appears to establish several facts.

First, broadly speaking, there is ample evidence that charter schools outperform traditional public schools in some areas, and, to a slightly lesser degree, that charter schools underperform in other areas. There are far more significantly positive and significantly negative charter school effects in the value-added and lottery literature than we would expect if charter schools were never different from traditional public schools.

The mission of charter schools is to offer innovative curricula and teaching methods. The finding of considerable heterogeneity among charter schools probably reflects this spirit of innovation and experimentation. Of course, this begs the question of whether over time the weaker charter schools will shrink or close, while the better charter schools expand. Hanushek, Kain, Rivkin, and Branch (2007) provide fascinating evidence from Texas that parents are more likely to pull their children out of charter schools that boost students' test scores by less than average. This is just one state, but the finding suggests that in the long run, heterogeneity in quality could lead to uniformly higher school quality in the charter sector. This statement is speculative, because for the weaker schools to shrink in size it is necessary that students who leave are not immediately replaced by new students. Further, we also need studies that check whether charter schools that close down, either voluntarily or because the chartering authority refuses to renew the charter agreement, are those with lower gains in student achievement.

Second, in a few cases the evidence to date is quite one-sided. For instance, elementary charter schools seem to be producing better gains in reading achievement than traditional public schools. Similarly, charter middle schools often outperform in math. Conversely, charter high schools seem to lag behind traditional public schools, especially in math. We urgently need more rigorous studies at the high school level.

Third, the overall evidence suggests that charter schools more typically outperform than underperform their traditional public school counterparts.

Fourth, median effect sizes are positive and roughly 8% of a standard deviation in elementary school for both math and reading. To put this into perspective, a student

with median test scores—ranking 50th out of 100 students—would be predicted to move up to about the 47th rank out of 100 students after one year at a charter school. This is not a large change, but over several years of such gains it could be meaningful. The median effect size at the middle school level is positive and varies between 1% and 2% of a standard deviation. These effects are relatively small, on the same order as some recent estimates of the effects of class size reduction. At the high school level, all methods suggest a negative median effect, but these differ in size enormously, from around 20% of a standard distribution in weighted estimates to only 2% of a standard deviation in unweighted estimates.

Fifth, the way in which we pose the question about achievement truly matters. Asking “What does the typical study show?” in some cases produces quite different answers than asking “For the typical charter school studied, what is the estimated effect on achievement?” We think the latter question holds far more relevance for policymakers, who must make decisions about whether to maintain charter school legislation, to increase or decrease the number of charter schools, or to eliminate charter schools altogether. When we give more weight to studies that include a greater number of charter schools, we tend to find less evidence of variation in the effects of charter schools.

Finally, even though we have learned a great deal about charter schools and achievement over the last decade, we need to exercise some caution in reading this literature. Researchers have conducted rigorous value-added or lottery-based studies of charter schools in only a very few states and major cities to date, and even here, the quality of the data and hence the quality of the analyses vary. In many cases, a few new soundly designed studies in a given grade span and subject area could lead to meaningful revisions to our understanding.

REFERENCES

- 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., 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. 2007. Madness in the Method? A Critical Analysis of Popular Methods of Estimating the Effect of Charter Schools on Student Achievement. University of California, San Diego.
- Bifulco, Robert, and Helen F. Ladd. 2006. The Impacts of Charter Schools on Student Achievement: Evidence from North Carolina. *Education Finance and Policy* 1 (1): 50-90. <http://www.mitpressjournals.org/toc/edfp/1/1>.
- Booker, Kevin, Scott M. Gilpatric, Timothy Gronberg, and Dennis Jansen. 2004. Charter School Performance in Texas. University of Tennessee, Knoxville.
- 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. <http://www.nber.org/papers/w12828>.
- 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* 91:823-848.
- Hedges, Larry V., and Ingram Olkin. 1985. *Statistical Methods for Meta-Analysis*. London, United Kingdom: Academic Press.
- Hoxby, Caroline M., and Sonali Murarka. 2007. New York City's Charter Schools Overall Report. <http://www.nber.org/~schools/charterschoolseval/>.
- Hoxby, Caroline M., and Jonah E. Rockoff. 2005. The Impact of Charter Schools on Student Achievement. Harvard University. http://media.hoover.org/documents/ednext20054unabridged_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. The Center for Research on Educational Equity, Assessment and Teaching Excellence. University of California, San Diego. http://create.ucsd.edu/Research_Evaluation/PreussReportDecember2005.pdf.
- Miron, Gary, Anne Cullen, Brooks Applegate, and Patricia Farrell. 2007. Evaluation of the Delaware Charter School Reform: Year One Report. The Evaluation Center, Western Michigan University. http://www.doe.k12.de.us/files/pdf/sbe_decseval.pdf.
- Sass, Tim R. 2006. Charter Schools and Student Achievement in Florida. *Education Finance and Policy* 1 (1): 91-122. <http://www.mitpressjournals.org/toc/edfp/1/1>.
- Solmon, Lewis C., Kern Paark, and David Garcia. 2001. Does Charter School Attendance Improve Test Scores? The Arizona Results. Phoenix, AZ: Goldwater Institute's Center for Market-Based Education. <http://www.goldwaterinstitute.org/pdf/materials/111.pdf>.
- 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/>.

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
University of Washington Bothell
2101 N. 34th Street, Suite 195
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.