

 Open access • Journal Article • DOI:10.2139/SSRN.3045966

You Can't Always Get What You Want: Using "Broken Lotteries" to Check the Validity of Charter School Evaluations using Matching Designs — [Source link](#)

Leesa M. Foreman, Kaitlin P. Anderson, Gary W. Ritter, Patrick J. Wolf

Institutions: Arkansas Department of Education, Lehigh University, Saint Louis University

Published on: 01 Jul 2017 - Social Science Research Network

Topics: Charter, School choice and Internal validity

Related papers:

- [Using "Broken" Lotteries to Check the Validity of Charter School Evaluations Using Matching Designs.](#)
- [Can Broad Inferences Be Drawn from Lottery Analyses of School Choice Programs? An Exploration of Appropriate Sensitivity Analyses](#)
- [Where You Come from or Where You Go? Distinguishing between School Quality and the Effectiveness of Teacher Preparation Program Graduates](#)
- [Evaluating the provision of school performance information for school choice](#)
- [Improving Selection of the Comparison Group from State Longitudinal Data Systems Using Two-Stage Matching](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/you-can-t-always-get-what-you-want-using-broken-lotteries-to-2nzlyqu9wy>

University of Arkansas, Fayetteville

ScholarWorks@UARK

Education Reform Faculty and Graduate
Students Publications

Education Reform

7-1-2017

You Can't Always Get What You Want: Using "Broken Lotteries" to Check the Validity of Charter School Evaluations using Matching Designs

Leesa M. Foreman

University of Arkansas, Fayetteville, lforema@email.uark.edu

Kaitlin P. Anderson

Michigan State University

Gary W. Ritter

University of Arkansas, Fayetteville, garyr@uark.edu

Patrick J. Wolf

University of Arkansas, Fayetteville, pwolf@uark.edu

Follow this and additional works at: <https://scholarworks.uark.edu/edrepub>



Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Education Policy Commons](#), and the [Other Education Commons](#)

Citation

Foreman, L. M., Anderson, K. P., Ritter, G. W., & Wolf, P. J. (2017). You Can't Always Get What You Want: Using "Broken Lotteries" to Check the Validity of Charter School Evaluations using Matching Designs. *Education Reform Faculty and Graduate Students Publications*. Retrieved from <https://scholarworks.uark.edu/edrepub/5>

This Article is brought to you for free and open access by the Education Reform at ScholarWorks@UARK. It has been accepted for inclusion in Education Reform Faculty and Graduate Students Publications by an authorized administrator of ScholarWorks@UARK. For more information, please contact scholar@uark.edu.



UNIVERSITY OF
ARKANSAS

College of Education & Health Professions
Education Reform

WORKING PAPER SERIES

You Can't Always Get What You Want: Using "Broken Lotteries" to Check the Validity of Charter School Evaluations using Matching Designs

Leesa M. Foreman, Kaitlin P. Anderson, Gary W. Ritter, Patrick J. Wolf

July 2017

EDRE Working Paper 2017-15

The University of Arkansas, Department of Education Reform (EDRE) working paper series is intended to widely disseminate and make easily accessible the results of EDRE faculty and students' latest findings. The Working Papers in this series have not undergone peer review or been edited by the University of Arkansas. The working papers are widely available, to encourage discussion and input from the research community before publication in a formal, peer reviewed journal. Unless otherwise indicated, working papers can be cited without permission of the author so long as the source is clearly referred to as an EDRE working paper.

You Can't Always Get What You Want: Using "Broken Lotteries" to Check the Validity of Charter School Evaluations using Matching Designs

Abstract

We consider situations in which public charter school lotteries are neither universally conducted nor consistently documented. Such lotteries produce "broken" Randomized Control Trials, but provide opportunities to assess the internal validity of quasi-experimental research designs. Here, we present the results of a statewide charter school evaluation using a broad-based student matching evaluation design, and run two additional analyses using the charter application waitlists as robustness checks. Our additional models, which address concerns of self-selection by using only charter applicants as matched comparison students, yield similar effect estimates and thus provide support for the use of matching designs in charter school evaluations.

Keywords: school choice; charter schools; research methods; student matching

You Can't Always Get What You Want: Using "Broken Lotteries" to Check the Validity of Charter School Evaluations using Matching Designs

1. Introduction – Using Charter School Lotteries to Support Matching Designs

In the field of school choice broadly and charter schools in particular, the existence of oversubscribed schools or programs that perform lotteries as a means of student selection provide a great opportunity for studying school effectiveness. School admission lotteries can provide the foundation for randomized experiments because winners are chosen at random from a group of applicants, each of whom were motivated to seek out a charter school. Because the randomized selection process reduces the likelihood of systematic observable or unobservable differences between those who receive the charter school assignment and those who do not, we are able to estimate charter school effects by comparing the outcomes for both groups of students.

Such experimental approaches have been lauded for their strong internal validity, as they generate unbiased causal estimates of programmatic effects (e.g. Mosteller & Boruch, 2002; Tufto 2006). However, such experiments have also been properly criticized for their limited external validity, due to the potential lack of generalizability of those causal estimates to the entire sample of program participants (Rossi et al., 2004; Gill et al., 2007; Betts & Hill, 2010). With experimental charter evaluations, for example, we are likely to obtain unbiased estimates of the impact of higher-quality charter schools on student achievement, since such schools are more likely to be over-subscribed than are lower-quality charters. However, there is likely something different about these schools, which leads them to be oversubscribed – higher achieving students, better teachers, more involved parents, different curriculum, etc. Thus, while evaluations of oversubscribed schools provide the best opportunity for randomized experiments, such studies may tell us very little about charter schools that are having trouble filling all available spaces.

In addition to the questions regarding the external validity of lottery-based charter studies, there are also potential complications with the administration of student lotteries in oversubscribed schools that undermine rigorous random assignment evaluations.

1.1 When Lotteries do not Allow Experimental Evaluations

Two specific issues connected to the implementation of the admission lotteries limit their usefulness for evaluation. First, in oversubscribed charter schools that are popular with parents and students, there may be very few “lottery seats” available in grade levels other than the entry grade. In such schools, students generally continue in the school from one year to the next, and the only open seats are found in the school’s “initial” grade. For example, in an oversubscribed K-12 charter school, the vast majority of “lottery” seats will be in kindergarten. Moreover, in schools that allow for sibling preference, many of these seats can be filled by siblings rather than by randomly selected applicants. Oversubscribed schools that serve the youngest students are particularly difficult to study because very little achievement data are gathered in the earliest school years, and there is generally no baseline information for students entering Kindergarten. Thus, perhaps the best opportunity for careful “lottery” studies exists in charter middle schools, where there is a “pipeline” grade each year in which all seats in the initial grade are open for lottery admission. For this reason, many of the best known charter school lottery studies are conducted at middle schools where the set of entering students have previously attended traditional public schools and have pre-treatment data available.

Moreover, the best opportunity to implement a randomized study with the greatest number of lottery seats available occurs in the initial year of charter school operation. However, with charter schools and most new programs, the initial years of operation involve start-up challenges

that may cause results that are not typical or representative of future results. Indeed, most relevant studies have found that the performance of new charter schools improves over time (CREDO, 2009). Therefore, achievement effects at newly established and/or oversubscribed charter schools are not likely to be representative of the effects of charters across a district or a state.

Second, careful Randomized Control Trials (RCTs) require excellent data management. That is, charter school personnel need to track carefully initial applicants, initial admits, waitlist admits, and those who declined offers of enrollment. Many charters may not have the types of data management systems required to organize such information. Additionally, some states do not require charter schools to report lottery and waitlist results, or maintain them over time. In the state we evaluate, we find charter school operators report (1) that they held a lottery, (2) the students who eventually matriculated to the school (which includes those who won the initial lottery, those who lost the initial lottery but were later admitted off the waitlist, and excludes lottery winners who failed to enroll), and (3) the students remaining on the waitlist after enrollment was finalized. Thus, their lists represent the outcome after post-lottery student sorting and not the actual randomization of the students.

1.2 Motivation for the Current Study

Due to shortcomings of both experimental and quasi-experimental designs, in this study we combine both strategies to examine the effectiveness of charter schools across a mid-sized U.S. state. Indeed, our initial plan was to use charter school enrollment and waitlist information from the 2012-13 school year to conduct a lottery-based evaluation for the State Department of Education. Toward that end, we were provided with academic and demographic data for all students in the state's public schools (charter and traditional) for the academic years from 2011-

2012 to 2013-2014. However, we were unable to conduct a full-scale lottery-based evaluation because fewer than half of the 17 open-enrollment charter schools that we could evaluate over this time period reported a waitlist that resulted from a student admission lottery. Moreover, because schools did not engage in detailed record-keeping, we only had access to the enrolled students and 2012-13 waitlisted students at each of the oversubscribed charter schools. We did not have access to a list of lottery winners; thus, we had no information on students who were offered admission through the lottery but subsequently declined or on students who initially lost the lottery but then gained admission off the waitlist. Additionally, we were not informed if students were awarded automatic admission outside of the lottery for sibling preference or previous mid-year transfers. We inferred that lottery winners were those students who were first time enrollees in charter schools with waitlists in their grade levels in 2012-13. Finally, we also had no information on lottery losers who did not attend a public school in the state in 2012-13 because out of state, private school, or home school applicants were not reported to us.

Essentially, the waitlist data allowed us to identify some of the lottery losers and to infer (albeit imperfectly) who the lottery winners were in the relatively small number of grade levels in oversubscribed schools with meaningful admission lotteries. Clearly, we had imperfect lotteries, or a “broken RCT”¹ (Barnard et al., 2003; 1998), that would not allow for an actual experimental evaluation. Nevertheless, we had waitlist information for several charter schools in the state that had large waitlists even in grades that are mostly filled with continuing students. Thus, we had experiments in which few people get the actual treatment, but the control population is large.

¹ Barnard et al. (1998) refer to randomized experiments that suffer from missing data and noncompliance as broken randomized experiments.

Thus, our primary design would necessarily be quasi-experimental, but we could also exploit the broken lotteries as robustness checks.

1.3 Quasi-Experimental Designs – Alternatives to Lotteries

Because all charter schools are not oversubscribed, oversubscription does not always allow for experimental studies, and true experimental studies will not provide information about the universe of charter schools, we must employ other designs to examine the effectiveness of charter schools and provide important information to policymakers. Quasi-experimental design (QED) studies, such as those involving student matching and longitudinal analysis, can be conducted in schools with or without viable student admission lotteries. Admittedly, such designs are weaker than experiments in internal validity due to the possibility that specific types of students are selected to charter schools. Nevertheless, for all of the reasons described above, our primary analytic strategy in this evaluation is quasi-experimental.

Our primary quasi-experimental design employs a propensity score matching method to create a comparison group of non-charter students. These matched students are drawn from traditional public schools (TPS) within the same district they would have otherwise been assigned had they not attended the charter school, with control variables for pretreatment achievement measures.² Of course, this TPS-Matching analysis is subject to reasonable criticism due to concerns related to selection bias (Rossi et al., 2004; Cook & Campbell, 1979); therefore, we supplement our matching design with robustness checks from our lottery waitlist to address such concerns. We use the waitlist to conduct robustness checks in two different ways - one quasi-experimental, the other observational.

² This method is recommended by Bifulco (2012).

Specifically, because both enrolled charter students and reported waitlist students were all applicants to charter schools, it is reasonable to use the very large waitlist as a population from which to draw a second set of matched comparison students in a quasi-experimental design. We refer to this model as the Waitlist-Matching analysis. While this strategy clearly does not replicate an actual lottery, the fact that the entire sample in such a design did indeed actively apply for charter school entry addresses issues of selection bias in a way that straight matching designs cannot. As an additional robustness check, we conduct an observational regression analysis that includes all charter applicants (those enrolled and those on the waitlist), which we refer to as the Applicant (descriptive) model. This design does not limit the sample to charter school students and their particular student “matches”, but uses information from all applicants on the waitlist.

Specifically, then, in this study we address two overarching questions, one practical and one methodological:

- (1) Based on quasi-experimental evaluation strategies, do we find that charter schools in this state are effective at improving student achievement?
- (2) How confident should we be in our findings? Do our robustness checks, using waitlist students as the comparison population yield similar results as a “matching study” comparing charter students to similar students in traditional public schools?

This paper is organized in the following way. We first review the relevant studies of charter schools that have used both randomized control trials and matching methods . We then present details on the matching method and regression model in the Methods section. Results of our analyses are reported in the Findings section. We conclude with a discussion of the limitations, policy implications, and opportunities for future research.

2. Review of Literature - Charter School Studies

School choice as a school improvement strategy, through a variety of methods, has been seriously contemplated since the 1960s. Providing choice to families and students who otherwise are residentially assigned to traditional public schools could, in theory, create competition that spurs innovation in traditional public schools (Friedman, 1962; Bohte, 2004). One prominent form of school choice is public charter schooling, first developed in Minnesota in 1991 (NAPCS, 2016). Charter schools are distinctive public schools allowed the freedom to be more innovative while being held accountable for advancing student achievement (ibid). Charter schools are held accountable in primarily two ways: by their charter authorizer and the threat of closure, and by the families they serve. They are generally open to all children who wish to apply and do not charge tuition (NAPCS, 2016). These schools provide parents with an alternative public school option to the traditional public schools in their neighborhoods. Currently, there are more than 6,700 charter schools in 42 states and the District of Columbia (NAPCS, 2016).

To frame our findings and to provide context, we examined the relevant studies of charter schools that have used both randomized control trials and matching methods. The literature review included here is by no means exhaustive. For a more complete review of the literature on charter school research, see Betts and Tang (2014). Betts and Tang find that students in charter schools generally perform better in math than those in traditional public schools in most grades, with middle schools producing the largest gains. They find reading achievement effects of charters to be positive but not statistically significant, though this appears to be driven by a few studies with large negative effect sizes. They note that impacts in the charter sector vary considerably, in particular across geographic areas and the age of the charters, generally described as the “maturity” effect.

2.1 Charter RCT Studies

Mathematica Policy Research (MPR) conducted an evaluation of 36 charter middle schools in 15 states (Clark et al., 2011). MPR limited its evaluation to charter schools that were over-subscribed and used random lotteries to determine which students were and were not admitted. The MPR Randomized Controlled Trial (RCT) study found that, overall, the charter middle schools in the sample produced average achievement gains that were similar to those of the control group. Urban charters tended to have statistically significant positive effects on student achievement while rural charters tended to have statistically significant negative effects. Lower-income students realized more positive achievement gains from charters while higher-income students experienced more negative achievement effects.

Fortson et al. (2012) conducted an RCT which examined a sample of middle school students from 15 charter schools in six states over a two-year period. Students randomly selected to attend charter schools through the lottery had nearly identical average math and reading test scores (baseline) as did students in the control group. From a restricted subsample of treatment or control students with outcome, baseline, and pre-baseline test scores, they found charter school students had insignificant differences in average math test scores as students in the control group, but statistically significant lower reading test scores.

Hoxby et al. (2009) conducted a random assignment evaluation of New York City charter schools that included 93% of NYC charter students in grades 3-12 over an eight-year period. Ninety-four percent of the charter students included in the study were admitted through lottery. They found significantly positive effects for students who attended a charter school for all grades, kindergarten through eight, for both math and English scores. The magnitude of the effects were

large; the authors reported the math effect to be about 86 percent of the black-white achievement gap in that subject and the effect in English represented about 66 percent of that achievement gap. Students who attended fewer years improved by commensurately smaller amounts. Hoxby et al. also found that students who attended a charter high school were more likely to receive a Regents diploma and to score higher on the Regents exams than their control group counterparts.

In an RCT study done by Abdulkadiroglu et al. (2009), which included students in Boston's middle and high schools in grades 6-12 over a seven-year period, the authors found large positive effects for charter school students in reading and math at both the middle and high school levels. They found the largest effects in math at the middle school level.

These studies of oversubscribed charters generally report positive results. This is consistent with the theory that parents and students will seek out the highest quality options. In fact, this is the reason that we should exercise caution in drawing inferences on the effectiveness of all charter schools from the randomized trials alone. We should expect that the results found in these oversubscribed schools are not representative of the results found in the universe of charter schools.

2.2 Charter Matching Studies

Many studies have been conducted on charters that are not necessarily oversubscribed. The Center for Research on Educational Outcomes (CREDO) at Stanford University has performed three national evaluations of charter school performance. In all three studies, CREDO used a "Virtual Twin" Matching (VTM) method, which matches each charter school student with a composite of multiple traditional public school students that, collectively, reflect the charter student's

observable characteristics assumed to affect achievement. This type of study serves, to some extent, as a model for the student matching method we use in our analyses.

The first of these studies examined charter school populations in 15 states and the District of Columbia with data available from 2003-04 through 2007-08 (CREDO, 2009). That evaluation concluded that 17 percent of charter schools in the sample generated achievement outcomes that were significantly higher than the outcomes for the comparison students, 37 percent of charters delivered achievement outcomes that were significantly worse, and for 46 percent of the charters there was no statistical difference between the outcomes of their students and the virtual twins. Charter performance was somewhat more positive for low-income students and at charters that had been open longer.

The second study evaluated the same states as in the 2009 report as well as 10 new states and the city of New York with data released since the 2009 report (CREDO, 2013). This follow-up study reported more favorable findings for public charter schools, concluding that charter schooling generated a very small but statistically significant benefit in reading that amounted to seven extra days of achievement growth per year. This study reported no difference between charters and traditional public schools in math. Low-income, English Language Learner (ELL), and special needs students appeared to benefit the most from charter schools in terms of achievement gains.

CREDO released a third national study focused on 41 urban areas with substantial concentrations of public charter schools (CREDO, 2015). It concluded that urban charters deliver achievement benefits in both reading and math that are statistically significant and substantially meaningful, amounting to 28 additional days of learning growth in reading and 40 days in math.

Summary

The results from these national studies of charter school effects on student test scores are remarkably similar. They all suggest that the average effect of charter schooling in general on student achievement is modest. The national studies with longitudinal components indicate that average charter effects tend to be somewhat negative in the earlier years of charter schooling but is now somewhat positive. The studies all provide evidence that urban charters, especially with disadvantaged populations of students, demonstrate much larger and more consistently positive achievement results than do rural charters. .

2.3 Contributions of this Paper

We present the results of a charter school evaluation of a medium-sized U.S. state using a student matching methodology similar to the one used in the CREDO studies except that we match students 1:1 instead of 1:many to avoid the potential problem of bias raised in critiques of the CREDO approach.³ We discuss the extent to which our primary findings are consistent with the general patterns of charter school effects from the national studies reviewed above. Importantly, because several of the charter schools in our sample were oversubscribed and we have information on the waitlist population, we conduct sensitivity and robustness checks of our primary results using the waitlist samples, which address concerns of student selection bias.

Bifulco (2012) indicates that longitudinal student matching studies that include baseline test scores and geography as matching variables can produce effect estimates that are less than

³ The most salient criticism of the “Virtual Twin Matching” is that the use of a larger number of students in the comparison group in an estimation of achievement gains over time biases the charter effects estimates towards zero because the estimations are more precise on the comparison group side of the analysis than on the charter group side (Hoxby, 2009).

four percent different from those obtained through experimental analysis. We use Bifulco’s approach for our primary analysis (the TPS-Matching analysis) and for a robustness check in which we match charter students with students on charter waitlists (referred to as the Waitlist-Matching analysis throughout this paper). Matching within the population of students who sought entrance to charters essentially controls for parent motivation to seek charter schooling and thereby addresses the main threat to the internal validity of QED charter studies.

Furthermore, as we cannot verify the conduct and validity of the charter school lotteries performed in schools that we study, but we can confirm which students were eligible applicants and which students were left on the waitlist, we classify the experimental analysis of the charter effects as a “broken RCT” which could contain selection bias. Conducting a straightforward lottery analysis here, unfortunately, would not generate meaningful information on overall program impacts, because it would be based on a small subsample of students in “lottery grades” in a subsample of non-representative schools. However, because students on the waitlists did indeed apply for charter school admission but attended a traditional public school when they were not selected, we can use these students in the analysis to check the robustness of our general matching design. This student matching approach, within the pool of charter applicants, can address the issue of selection bias and increase internal validity. We now turn to a description of the data used for the analyses.

3. Data

The research team was provided de-identified student level data for a mid-sized U.S state, for years 2011-12 to 2013-14. Each ID was paired with information for each school year including the

student's school attended, race/ethnicity, gender, free and reduced lunch status, English language learner status, special education status, and test scores for math and literacy.

The student test scores came from the state criterion referenced exams in both math and literacy. In this state, during the study period, assessments were taken by students in 3rd through 8th grade. We use the prior year test score as the baseline measure and compare outcomes of student achievement in math and literacy for 2012-13 and 2013-14. Scores are standardized within grade and school year.

This state includes both open enrollment charter schools and district-conversion charter schools. Conversion charters are formerly traditional public schools which later converted to charter, but remain under the purview of the local district. These conversion schools, while technically called charters in this state, are often the only public school available to a community serving particular grade levels. More importantly, only students living within district boundaries can attend these schools. Open enrollment charter schools, on the other hand, are created from scratch to be their own school district and are open to all students who apply, regardless of residence. These open enrollment charters are the types of schools that most researchers and observers are referring to when they study and debate the pros and cons of charter schools (Greene et al., 2005; Hanushek et al., 2006; Booker et al., 2007; Forman, 2010). As a result, our analyses here focus solely on the open enrollment charter schools in the state. Throughout this paper, "charter schools" refers exclusively to these open enrollment charters.

Charter students represent about 1.6% to 1.8% of all K-12 students statewide depending on the year (Table 1 below). Charter students' share of total enrollment has increased over the two years covered by this study. Seventeen charter schools, feeder schools (the local TPS that students generally would have attended based on residence), for a total of approximately 5,500-7,000

charter and TPS students, in each of the two years, were included in each year of analysis of the TPS-Matching evaluation.⁴ Ten charter schools and approximately 2,500-3,200 students, respectively, were included in each year of analysis of the Waitlist-Matching robustness check.

The demographics of charter students differ in some observable ways from the state as a whole. Charter schools serve a smaller proportion of low-income students but a larger proportion of minority students.⁵ For the primary TPS-Matching analysis, the sample of minority students is proportionate to open enrollment charters statewide but has a lower proportion of low income students.⁶ Charter schools included in the Waitlist-Matching robustness check sample have greater proportions of minority students than the population of charters and students statewide.⁷ This is likely due to the fact that the pool of potential matches is much smaller and the oversubscribed schools tend to be in large metropolitan areas that serve more diverse student populations.

⁴ While there were 20 open enrollment charters in the state at this time, two schools with less than 15 matches for analysis were excluded. In addition, two schools that were part of the same system and shared the same campus were combined as one for analysis. Thus 17 start-up charter schools were included.

⁵ In further analyses, propensity score probabilities indicate that charter students were generally less likely to be designated free or reduced lunch or black, and more likely to have lower test scores. They were also somewhat more likely to be Asian, Hispanic, Native American, or female, however, likelihoods vary slightly by geographic region.

⁶ The TPS-Matching Analysis includes charter students in a region of the state which has a much smaller fraction of minority students and free and reduced lunch students. This region was not included in the Waitlist-Matching Analysis.

⁷ The matched group is slightly more low income and minority than the set of students in these charter schools.

Table 1. Student Demographics: Charter Students Compared to State, 2012-13 and 2013-14 ⁸

2012-13	State (All Students)	Open Enrollment Charter Students	Charter Students in TPS-Matching Analysis	Charter Students in Waitlist-Matching Analysis
Enrollment	471,867	7,150	3,148	2,148
Charter as % of Total		1.5%	0.7%	0.5%
FRL %	61%	57%	50%	52%
Minority %	36%	61%	60%	68%
<i>N of Schools</i>	930	17	17	10
2013-14				
Enrollment	474,995	8,393	4,012	2,669
Charter as % of Total		1.8%	0.8%	0.6%
FRL %	61%	57%	54%	56%
Minority %	37%	63%	59%	71%
<i>N of Schools</i>	929	17	17	10

4. Methods

4.1 Quasi-Experimental Design

Our research questions are: 1) how well do charter students perform in math and literacy achievement relative to their traditional public school peers; and 2) do the robustness checks using charter waitlist students as comparison students generate similar results as did the broad-based charter-TPS student-matching design?

Both the primary TPS-Matching analysis and the Waitlist-Matching robustness check match each individual student attending a charter with a student who is not in a charter school. The goal is to create two sets of students that are as similar as possible on key characteristics: one in charter schools and the other in traditional public schools. Any differences will not be based on

⁸ Charter students in the matching analyses only include those students in tested grades with prior year test scores.

observable student characteristics (such as race, income, gender, or prior test scores) as matched twins are intentionally selected to be nearly identical on these characteristics. Students were re-matched the second year to account for movement between the groups.

Students are excluded from the analysis if they did not have test scores from both the baseline year and the outcome year. Thus, the results should be interpreted as the impacts for the matched student population, which may not generalize to the broader student population. We report single year estimates. In the TPS-Matching analysis 69-90% of students in tested grades were actually matched and for the Waitlist-Matching analysis 53-67% were actually matched.⁹

For the primary TPS-Matching analysis, the sample available for matching includes all charter and traditional public school students from the districts where the students would have attended had they not been admitted to a charter school. However, we were unable to match many charter students to comparison students one-to-one. Charter students whom we were not able to match were not included in the analysis.¹⁰

Despite the similarity in observable characteristics between the charter students and the comparison students, these students could still differ on one fundamental unobservable attribute – the motivation, due to positive or negative traits or past experiences, to seek out a seat in a charter school. Our Waitlist-Matching analysis, which matches charter students only to their peers on charter waitlists, overcomes this shortcoming.

⁹ Matched students in the TPS-Matching analysis account for 36-42% of all open enrollment charter students and 16-19% of open enrollment charter students are matched in the Waitlist-Matching analysis. The main reason for this sample limitation is the matching requirements. Each student in the study must have test scores from both the baseline test year and the outcome year. Nevertheless, as Table 1 shows, the various samples are quite similar on all observable measures.

¹⁰ The matched student population includes more minority students and free and reduced lunch students than the broader charter student population.

For the robustness check using the Waitlist-Matching analysis, the sample consists of all charter students in the geographic area of the oversubscribed schools and all waitlist students.¹¹ Most oversubscribed schools fall within a central urban metropolitan area, and one is located in a rural area. All charter schools in these areas are included in the analysis as treatment schools, whether or not they allotted any open seats in the previous year using a lottery. Of the 17 charters in the primary TPS-Matching analysis, only 10 charter schools are located in geographic regions with oversubscribed schools with waitlists; thus, only these schools were included in the Waitlist-Matching analysis. Therefore, matched comparison students are similar in that they applied to a charter school in that geographic area, but comparison students did not necessarily apply to the same charter school that their treatment group-match is attending.

For greater comparability of results, we examine whether we can replicate results with these two different matching methods. It is important to note that the overall sample in the primary TPS-Matching design is larger than the robustness check waitlist sample. To have an apples to apples comparison of the methods, we restrict the sample of the TPS-Matching analysis to include the same 10 schools that were used in the Waitlist-Matching analysis. Thus, when we compare results of each method, we are sure to analyze the same sub-sample of students. As we may be concerned that by restricting the full TPS-Matching sample we may be excluding very different types of schools, we examine the baseline characteristics below of the schools included (see Table 2).

¹¹ It is possible that traditional public school students matched in the primary matching analysis (TPS-Matching) may have been re-matched in the secondary analysis (Waitlist-Matching) as they were enrolled in traditional public schools. The samples may not be mutually exclusive. However, it is also possible that charter students not matched in the primary analysis were matched in the secondary analysis as waitlist students may have included those outside the feeder districts.

4.2 Matching Method

Students who have received the “treatment” of being in the charter school are matched to comparison students on observable characteristics from the previous school year using a multi-step matching procedure. To account for movement between the groups students were re-matched the second year, therefore, students were matched based on 2011-12 test scores when evaluating 2012-13 and students were matched based on 2012-13 when evaluating 2013-14. This identical process is used for both the primary TPS-Matching analysis and the robustness check using the Waitlist-Matching analysis:

Matching Process (Conducted Separately for Math and Literacy)

1. Students are first matched with a student in the same grade in both the outcome year and baseline or matching year (always the year before).
2. For the math and literacy analyses, separately, all students are matched based on *previous year scores on the same subject test*, rounded to the nearest 0.01 z-score unit. The other subject test score is used as part of the propensity score in step 4 (below), as having a matched test score in the same subject is more relevant for controlling for prior performance. Therefore, the math analysis matches first on math examination scores and later factors in literacy scores, while the literacy analysis matches first on literacy examination scores and later factors in math scores.
3. A propensity score is then created using free and reduced lunch (FRL) status, race/ethnicity (African-American, Asian-American or Pacific Islander, Hispanic-American, Native American, White, or “Two or more races”), gender, and the “other” test score (literacy for the math analysis and math for the literacy analysis).

4. Finally, all matches are based on guaranteeing exact matches from step 1 and 2, and the closest available propensity score match from step 3.¹²

4.3 Analytic Samples

For greater clarity, we define the analytic samples used for these analyses. The primary **TPS-Matching Analysis** includes charter school students matched to traditional public school students in districts where the charter students would have attended had they not been admitted to a charter school. Seventeen charter schools are included in this analysis. As a robustness check, the **Waitlist-Matching Analysis** includes charter schools in the geographic area of the oversubscribed schools (with reported waitlists in 2012-13) matched to TPS applicants that were not admitted to charters. Ten charter schools are included in this analysis. We refer to this as the Restricted TPS-Matching sample. Finally, as an additional robustness check, in the **Applicant (Descriptive) Model**, we include all charter admitted students and all non-admitted waitlist students for 2012-13 without matching, thereby using the same 10 schools used in the Waitlist-Matching analysis and all of the waitlist information. The main results we present focus on the overall samples used in the TPS-Matching. We then compare these results to the results from the robustness check that compares the same schools in the Restricted TPS-Matching and the Waitlist-Matching, and the robustness check of the Applicant (descriptive) model. All sets of results are presented in the Findings section that follows.

We conducted baseline equivalency analyses to test how similar the matched groups were to each other (Table 2). P-values below 0.05 indicate statistically significant differences that might

¹² For some of the analysis, we had to replace the original matched groups with broader matches in order to have a large enough sample of students. Broader matches were accomplished by relaxing the degree of similarity of the baseline test score for the two students.

raise concerns about the comparability of our samples in terms of observable characteristics. For the primary TPS-Matching analysis in math, one of six comparisons between the matched groups revealed a statistically significant difference (the percent FRL was higher in the TPS group). For literacy, one of seven differences was statistically significant (the percent FRL was higher in the TPS group). For the robustness check which includes the Restricted TPS-Matching sample in math, one of six differences was statistically significant (baseline scores in literacy were lower in the charter group). For literacy, two of seven comparisons revealed statistically significant differences (percent FRL was higher in the TPS group but percent minority was higher in the charter group). For the robustness check that uses the Waitlist-Matching sample in math, one of seven comparisons revealed a statistically significant difference (baseline literacy lower for the charter group) and in literacy none of the five differences were statistically significant.

Table 2. Baseline Equivalency for All Samples in Two Matching Designs for Math and Literacy Exams, 2011-13

<i>TPS-Matching</i>	<i>Math</i>						<i>Literacy</i>					
N of Schools = 17	2011-12			2012-13			2011-12			2012-13		
	Charter	TPS	Difference	Charter	TPS	Difference	Charter	TPS	Difference	Charter	TPS	Difference
Number of Observations	2822	2822	-	3493	3493	-	2775	2775	-	3360	3360	-
Tested Grades	4-8		-	4-8		-	4-8		-	4-8		-
Prior Year Math Z-Score	-0.14	-0.14	-	-0.05	-0.05	-	-0.11	-0.09	(0.02)	0.00	-0.03	0.03
Prior Year Literacy Z-Score	-0.03	-0.05	0.02	-0.04	-0.01	(0.03)	0.02	0.03	(0.01)	0.03	0.03	-
% FRL	0.50	0.54	(0.04) ***	0.61	0.65	(0.04)	0.48	0.56	(0.08) ***	0.78	0.79	(0.01)
% Minority	0.57	0.55	0.02	0.55	0.55	-	0.57	0.57	-	0.54	0.54	-
% Female	0.51	0.51	-	0.50	0.49	0.01	0.51	0.50	0.01	0.51	0.50	0.01
<i>Restricted TPS-Matching</i>												
N of Schools = 10	Charter	TPS	Difference	Charter	TPS	Difference	Charter	TPS	Difference	Charter	TPS	Difference
Number of Observations	2266	2266	-	2378	2378	-	2235	2235	-	2281	2281	-
Tested Grades	4-8		-	4-8		-	4-8		-	4-8		-
Prior Year Math Z-Score	-0.17	-0.17	-	-0.07	-0.07	-	-0.13	-0.12	(0.01)	-0.02	-0.05	0.03
Prior Year Literacy Z-Score	-0.05	-0.06	0.02	-0.06	-0.01	(0.05) *	0.01	0.01	-	0.03	0.03	-
% FRL	0.56	0.55	0.01	0.56	0.55	0.01	0.54	0.58	(0.04) ***	0.54	0.55	(0.01)
% Minority	0.65	0.64	0.01	0.66	0.66	-	0.65	0.63	0.02 *	0.66	0.65	0.01
% Female	0.51	0.51	-	0.51	0.50	0.01	0.52	0.51	0.01	0.51	0.51	-
<i>Waitlist-Matching</i>												
N of Schools = 10	Charter	Waitlist	Difference	Charter	Waitlist	Difference	Charter	Waitlist	Difference	Charter	Waitlist	Difference
Number of Observations	1257	1257	-	1428	1428	-	1422	1422	-	1569	1569	-
Tested Grades	4-8		-	4-8		-	4-8		-	4-8		-
Prior Year Math Z-Score	-0.16	-0.16	-	-0.09	-0.09	(0.00)	-0.11	-0.11	-	-0.02	-0.04	0.02
Prior Year Literacy Z-Score	-0.02	0.00	(0.02)	-0.07	0.01	(0.08) ***	0.05	0.06	(0.01)	0.07	0.07	-
% FRL	0.53	0.53	-	0.57	0.56	0.01	0.52	0.51	0.01	0.54	0.54	-
% Minority	0.64	0.64	-	0.67	0.66	0.01	0.62	0.61	0.01	0.65	0.65	-
% Female	0.49	0.55	(0.06) ***	0.50	0.52	(0.02)	0.52	0.52	-	0.52	0.51	0.01

Note: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

These summary statistics show that we were able to match our samples on most characteristics, though in some of the samples charter students were somewhat more economically disadvantaged and tended to perform slightly lower in literacy at baseline (these differences in baseline literacy scores were only evident in analyses of math scores).. Recall, as well, that characteristics such as FRL-eligibility and race were included in the propensity score, so we would not necessarily expect 1:1 matches on each characteristic, particularly if they were not strong predictors of charter school attendance. Importantly, our regressions, which control for these characteristics, account for the small differences observed in these matching variables so they will not introduce bias into the analysis. By comparing estimates based on the same populations of schools but produced by different matching methods, we can determine if our conclusions are sensitive to these shortcomings in our samples or strategies.

4.4 Regression Model

Restricting the samples to our sets of matches, we then conduct Ordinary Least Squares (OLS) regression analysis to assess how much of the academic growth for students can be attributed to attending charter schools. We use heteroskedastic-robust standard errors (Angrist & Pischke, 2009; White, 1980). The same regression model was used for all analyses.

OLS Regression Model.

$$y_{it} = \beta_0 + \beta_1 \text{charter}_{it} + \beta_2 \theta_{it-1} + \beta_3 X_{it-1} + \beta_4 \delta_{it} + \varepsilon_{it} \quad (1)$$

Where,

- y_{it} represents a given outcome of interest (math or literacy achievement) for student i in time t ,
- charter_{it} is an indicator for charter school treatment,
- θ_{it-1} represents controls for prior year test scores (both math and literacy achievement),
- X_{it-1} represents controls for student baseline characteristics (gender, race, FRL status),

δ_{it} represents control for switched schools, and
 ε_{it} represents the error term.

Our regression analyses statistically control for any minor differences in demographic characteristics.

4.5 Robustness Checks

To take advantage of all of the information provided by the waitlist, we conduct robustness checks in two different ways - one quasi-experimental, the other observational. First, we match charter students in the geographic area of the oversubscribed schools with all waitlist students for the Waitlist-Matching analysis. Second, we use the full sub-sample of charter and waitlist students *without matching*, and run an OLS regression where we control for prior achievement, student characteristics, and schools for which students applied. As this model includes all students who applied to these charter schools (those attending and those on the waitlist), we refer to this as the “Applicant” (descriptive) model.

Results of the primary TPS-Matching model(s), and the Waitlist-Matching and Applicant robustness check models are reported together in the following section.

5. Findings

Our objective in this study was to answer the following questions:

- (1) Are charters more effective than traditional public schools in this state?
- (2) Should we believe these results? Does our strategy of using waitlist students as the comparison population yield similar results as does a more broad-based “matching study” comparing charter students to similar students in TPS schools?

5.1 Overall and Subgroup Results

The academic impacts represented in Table 3 indicate that, for all 17 open enrollment charter schools in this analysis, students in public charter schools, on average, demonstrated positive and statistically significant impacts of 0.09 standard deviations on math scores and 0.07 standard deviations on literacy scores, in 2012-13. In 2013-14, we find no clear effect of attending a charter school on math scores, while there was a negative and statistically significant impact of -0.03 standard deviations on literacy scores.

Table 3. Primary Analysis: Charter and Subgroup Effect by Assessment, 2012-14

	2012-13	2013-14
<i>N of Schools=</i>	17	17
MATH		
Charter Effect	0.09 ***	-0.02
<i>Treatment n=</i>	2822	3493
N=	5644	6986
Subgroup Effect		
Low-Income	0.15 ***	-0.02
High-Income	0.05 **	-0.01
LITERACY		
Charter Effect	0.07 ***	-0.03 **
<i>Treatment n=</i>	2775	3360
N=	5550	6712
Subgroup Effect		
Low-Income	0.13 ***	0.06 **
High-Income	0.04 *	-0.08 ***
Control for prior reading/math tests	x	x
Control for student characteristics	x	x
Control for school applied		
Control for school switchers	x	x

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: High-income ($FRL < \text{State Average } 61\%$); Low-income ($FRL \geq \text{State Average } 61\%$)

We examine a subgroup of charter schools by level of poverty, defined by the State average, for both math and literacy outcomes. Low-income schools are defined those with percentages of free and reduced lunch students that are greater than or equal to the state average of 61%; high-income schools are the remaining schools. These groups were based on the charter school's overall enrollment, not necessarily the students that were actually matched. This subgroup is particularly relevant considering that open enrollment charter schools are designed as

a method of public school choice for students who may not be able to afford private schools or other options.

In math, we find positive and significant effects for both low and high income populations with a greater impact for low-income students in 2012-13. The effect on math scores of low-income students is 0.15 standard deviations and the effect on high-income students is about 0.05 standard deviations. For 2013-14, the effects are small and insignificant. In literacy, we find similar results for both years with positive and significant results that are larger for low-income populations. In 2012-13, the effect on literacy scores of low-income students is 0.13 standard deviations and in 2013-14 it is 0.06 standard deviations.

5.2 Model Comparisons

To assess the extent to which we should have confidence that our quasi-experimental matching design using all TPS students as the comparison group is internally valid (in the face of self-selection into charter schools), we can compare the results from this method to one in which we only match to charter school applicants. Thus, we restrict the primary TPS-Matching analysis sample to the same schools included in the Waitlist-Matching and Applicant analyses for comparison. If the results of the Restricted TPS-Matching sample provide similar estimates to those of the Waitlist-Matching, this convergence of the results of the primary TPS-Matching analysis enhances our confidence that our matching strategy is not significantly threatened by self-selection.

In the Restricted TPS-Matching analysis, in Table 4, we see the same results for 2012-13 as those found in the primary full TPS-Matching sample, with statistically significant impacts of

0.09 standard deviations in math and 0.07 standard deviations in literacy. For 2013-14, we find statistically significant impacts on both math and literacy scores of 0.03 standard deviations.

Table 4. Robustness Checks: Charter Effect by Assessment, Model Comparisons, 2012-14

	2012-13			2013-14		
<i>N of Schools=</i>	10	10	10	10	10	10
Model	TPS-Matching (Restricted)	Waitlist-Matching	Applicant	TPS-Matching (Restricted)	Waitlist-Matching	Applicant
MATH						
Charter Effect	0.09 ***	0.05 **	0.11 ***	0.03 *	0.02	0.02
<i>Treatment n=</i>	2266	1257	1838	2378	1428	2279
<i>N=</i>	4532	2514	3527	4756	2856	3748
LITERACY						
Charter Effect	0.07 ***	0.04 **	0.06 ***	0.03 **	0.01	0.04 **
<i>Treatment n=</i>	2235	1422	1838	2281	1569	2279
<i>N=</i>	4470	2844	3527	4562	3138	3748
Control for prior reading/math tests	x	x	x	x	x	x
Control for student characteristics	x	x	x	x	x	x
Control for school applied			x			x
Control for school switchers	x	x		x	x	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Preliminary statistical tests suggest that when the same ten schools are used in the samples, the results from various models (TPS-Matching (Restricted), Waitlist-Matching, and Applicant) provide similar results. In only two out of 12 comparisons, were the estimated effects statistically different. This gives us greater confidence in support of the findings of the larger TPS-Matching model results.

Comparing the results of the Restricted TPS-Matching analysis to the findings from the robustness check using the Waitlist-Matching analysis, we find similar results with statistically significant impacts for math (0.05 s.d.) and literacy (0.04 s.d.) in 2012-13. For 2013-14, we find no significant effects for math or literacy. This gives us greater confidence in the results of the primary full TPS-Matching analysis.

Examining the robustness check using the Applicant (descriptive) model analysis, we see that this analysis also appears to support the findings of both matching analyses. For 2012-13, we see positive and statistically significant impacts in math of 0.11 standard deviations, and 0.06 standard deviations in literacy. For 2013-14, we find no clear effects in math, but positive and significant effects in literacy of 0.04 standard deviations. The magnitudes and signs of these effects are quite similar to those found in the traditional matching method we used for the study of the full state. Thus, we believe that it is possible to use information from “imperfect” lotteries to test the robustness of broader analyses using standard quasi-experimental designs.

In fact, we do find that both the robustness checks, the Waitlist-Matching and the Applicant (descriptive) analyses, provided similar estimates to the primary TPS-Matching analysis results in most instances.

6. Discussion

This study sought to estimate the academic impacts of charter schools in a medium-sized state for the 2012-13 to 2013-14 school years. Initially, we hoped to assess the effectiveness of charter schools in the state using a lottery-based, experimental design. However, due to all of the circumstances and limitations described throughout this paper, we were not able to draw firm conclusions about oversubscribed open enrollment charter schools through a Randomized Control Trial (RCT) analysis. Given the data available, the quasi-experimental model seems to be a stronger form of analysis available to us. Using a careful student matching method, charter students in each school were matched with similar students in their feeder districts in each of these years in the primary TPS-Matching analysis.

Reasonable conclusions that can be drawn from this study are that charters across the state, and particularly in vibrant charter markets with waitlists (where at least some charters are oversubscribed), have a modest positive effect on student test scores in math and literacy, however, this finding is not consistent over both years of the analysis. The school year 2012-13 appeared to be the strongest year for charter school performance, compared to 2013-14.¹³ Effects for both math and literacy appear to be positive and somewhat larger for low-income populations. Perhaps more interesting is the opportunity to check these results against a method that addresses self-selection.

While this analysis does provide some information on charter school effectiveness in a single state, the more important contribution may be our attempt to make use of imperfect lotteries. As we have noted at length, even in the cases of oversubscribed charter schools, clean lottery studies of the impacts of full schools will be difficult to conduct in most environments. It may be possible to salvage “broken” lotteries using a combination of quasi-experimental methods to provide robust estimates using data from oversubscribed schools. What we find here is encouraging. Because traditional matching models do a good job of ensuring baseline equivalence on observables, the big concern is that charter and traditional students differ on unobservable characteristics due to student self-selection into charters. With robustness checks that employ charter waitlists – even from imperfect lotteries – to address concerns of self-selection, researchers can test the validity of the conclusions drawn from traditional matching models, which can be conducted in far more contexts than lotteries can.

¹³ Further analyses, not included here, indicate that the positive results in 2012-13 are primarily driven by particular schools ([newer schools](#), [schools with waitlists](#), [schools in the urban metropolitan area](#), and [schools serving less well-off students](#)).

In the case of this analysis, the robustness checks generated similar results as did the main models and thus increased our confidence in the validity of our findings. .This is important as it may be necessary to employ quasi-experimental designs in evaluations of most charters as clean RCTs are hard to find.. We recommend using multiple methods to assess how strong the effects of self-selection may be.

Researchers should continue to analyze the academic impacts of public charter schools. One of the most celebrated aspects of charter schools is that they are held accountable for their outcomes. This current evaluation seeks to add to that accountability and provide a means of checking the robustness of results found in the primary TPS-Matching analysis. In doing so, we have increased confidence in our traditional matching strategies as our waitlist analyses have generated similar findings. While academic impacts do not encompass the entire mission of a charter school, or any school, these results help provide information on the performance of public charter schools.

References

- Abdulkadiroglu, A., Angrist, J., Cohodes, S., Dynarski, S., Fullerton, J., Kane, T., & Pathak, P. (2009). *Informing the debate: Comparing Boston's charter, pilot and traditional schools*. Boston, MA: The Boston Foundation.
- Angrist, J.D., and Pischke, J.S. (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Barnard, J., Frangakis, C.E., Hill, J.L., & Rubin, D.B. (2003). Principal stratification approach to broken randomized experiments: A case study of school choice vouchers in New York City. *Journal of the American Statistical Association*, 98(462), 299-314.
- Barnard, J., Du, J., Hill, J.L., & Rubin, D.B. (1998). A broader template for analyzing broken randomized experiments. *Sociological Methods & Research*, 27(2), 285-317.
- Betts, J.R. & Hill, P.T. (Eds.) (2010). *Taking Measure of Charter Schools: Better Assessments, Better Policymaking, Better Schools*. Lanham, MD: Rowman & Littlefield Publishers.
- Betts, J.R., & Tang, Y.E. (2014). A meta-analysis of the literature on the effect of charter schools on student achievement. Center on Reinventing Public Education, Working paper.
- Bifulco, R. (2012). Can nonexperimental estimates replicate estimates based on random assignment in evaluations of school choice? A within-study comparison. *Journal of Policy Analysis and Management*, 31(3), 729-751.
- Bohte, J. (2004). Examining the impact of charter schools on performance in traditional public schools. *Policy Studies Journal*, 32(4), 501-520.
- Booker, K., Gilpatric, S.M., Gronberg, T., & Jansen, D. (2007). The impact of charter school attendance on student performance. *Journal of Public Economics*, 91(5), 849-876.

- Clark, M. A., Gleason, P., Clark-Tuttle, C., & Silverberg, M.K. (2011). "Do Charter Schools Improve Student Achievement? Evidence from a National Randomized Study." Mathematica Policy Research, Working Paper. Web. http://www.mathematica-mpr.com/~media/publications/PDFs/education/charterschools_WP.pdf.
- Cook, T.D. & Campbell, D.T. (1979). "Validity" in *Quasi-Experimentation: Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin.
- CREDO. (2009). "Multiple Choice: Charter School Performance in 16 States." Center for Research on Education Outcomes (CREDO), Stanford University. Web. http://credo.stanford.edu/reports/MULTIPLE_CHOICE_CREDO.pdf.
- CREDO. (2013). "National Charter School Study: 2013." Center for Research on Education Outcomes (CREDO), Stanford University. Web. <https://credo.stanford.edu/documents/NCSS%202013%20Final%20Draft.pdf>.
- CREDO. (2015). "Urban Charter School Study Report on 41 Regions: 2015." Center for Research on Education Outcomes (CREDO), Stanford University. Web. <http://urbancharters.stanford.edu/download/Urban%20Charter%20School%20Study%20Report%20on%2041%20Regions.pdf>.
- Forman Jr, J. (2010). Do charter schools threaten public education? Emerging evidence from fifteen years of a quasi-market for schooling. University of Illinois Law Review, Vol. 2007, May 2007; Georgetown Public Law Research Paper No. 921101. Available at SSRN: <https://ssrn.com/abstract=921101>
- Fortson, K., Verbitsky-Savitz, N., Kopa, E., & Gleason, P. (2012). Using an Experimental Evaluation of Charter Schools to Test Whether Nonexperimental Comparison Group Methods Can Replicate Experimental Impact Estimates. Institute of Education Sciences,

- National Center for Education Evaluation and Regional Assistance, U.S. Dept. of Education, April 2012.
- Friedman, M. (1962). *Capitalism and Freedom*. Chicago: University of Chicago Press.
- Gill, B., Timpane, P.M., Ross, K.E., Brewer, D.J. & Booker, K. (2007). *Rhetoric Versus Reality: What We Know and What We Need to Know About Vouchers and Charter Schools*. Santa Monica, CA: RAND Corporation.
- Greene, J.P., Forster, G., & Winters, M.A. (2003). Apples to Apples: An Evaluation of Charter Schools Serving General Student Populations. Education Working Paper No. 1. Center for Civic Innovation.
- Hanushek, E.A., Kain, J.F., Rivkin, S.G., & Branch, G.F. (2007). Charter school quality and parental decision making with school choice. *Journal of Public Economics*, 91(5), 823-848.
- Hoxby, C.M. (2009). "A Statistical Mistake in the CREDO Study of Charter Schools." Web. http://users.nber.org/~schools/charterschoolseval/memo_on_the_credos_study.pdf.
- Hoxby, C., Murarka, S., & Kang, J. (2009). How New York City's Charter Schools Affect Achievement, August 2009 Report. Second report in series. Cambridge, MA: New York City Charter Schools Evaluation Project, September 2009.
- Mosteller, F. & Boruch, F. (Eds.) (2002). *Evidence Matters: Randomized Trials in Education Research*. Washington, D.C.: The Brookings Institution.
- National Alliance for Public Charter Schools (NAPCS). (2016). About Charter Schools. Available at: <http://www.publiccharters.org/get-the-facts/public-charter-schools/>.
- Rossi, P.H., Lipsey, M.W. & Freeman, H.E. (2004). "Measuring and Monitoring Program Outcomes" in *Evaluation: A Systematic Approach, Seventh Edition*. Thousand Oaks, CA: Sage Publications.

Tufte, E. (2006). *Beautiful Evidence*. Cheshire, CT: Graphics Press.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*: 48, 817–838.