Choices for Studying Choice: Assessing Charter School Effectiveness Using Two Quasi-Experimental Methods

by

Devora H. Davis
Research Manager
Center for Research on Education Outcomes (CREDO) at Stanford University

and

Margaret E. Raymond
Director
CREDO at Stanford University

Abstract

In the current fiscally constrained environment, there is a consensus that traditional public school reform has not worked and funding for it cannot continue unabated. With the spotlight thus shining ever brighter on charter schools as a school reform effort, reaching actionable conclusions about charters has become essential. However, media coverage of charter schools is often polarized rather than productive, and data can be found to support both sides of the debate. This is largely because the literature on charter school effectiveness, until recently, was dominated by single-location studies that each used a slightly different research approach. The CREDO study, "Multiple Choice: Charter School Performance in 16 States," was among the first studies to attempt a broader picture of overall charter school effectiveness, and it used a novel quasi-experimental method called virtual control records (VCRs).

As a new analytic method, questions arose about the strengths and weaknesses of the VCR methodology vis-a-vis the most commonly used quasi-experimental method in education research, student fixed effects (FE). Using data from longitudinal student data from the United States, this study documents the results of a first look at such a head-to-head comparison of the two methods. Fixed effects follow a student who has changed from one sector to another – from traditional public school to a charter school or vise verse – thereby creating a time-variant student-as-own-control condition. The VCR method creates concurrent controls by melding the performance of traditional public school students who exactly match a charter school student, including baseline achievement.

The study compares samples that result in each type of model, compared to the full population of charter school students. The comparison is important, since the market for charter is highly

dynamic; the profile shows important variation in the proportion of charter school students that are included in a fixed effect design. These differences aside, the critical feature of the paper is a head-to-head test of designs for students that meet the criteria for a student fixed effect model using each technique. In addition, the study examines the "all-in results that are possible given the superior sampling fraction that VCR affords.

We find that estimates of charter school effectiveness are highly similar between the two methods when the sample is restricted to the same set of charter students. Whether the samples are restricted in this manner or allowed to vary between the VCR and FE models, results showed variation in charter school effectiveness across student subgroups and across locations regardless of the methodology being used.

Other recent multi-location studies of charter schools have also uncovered variation in charter school quality. There are two possible conclusions to this now-common finding. Detractors decry charter schools because they are not a silver-bullet answer to all education's problems. But this knee-jerk reaction ignores a more nuanced point: some charter schools are effective, and they are often effective for students whom the traditional public school sector has historically failed to serve well, such as low-income students, minority students and English language learners.

Recognizing this fact, the conclusion reached about the widespread variation in quality is that the conditions that bring about a high-quality charter school sector need to be investigated. This discussion focuses on the accountability side of the flexibility-for-accountability bargain made with charter schools, including the parameters of strong charter school authorizing and the frameworks that can be used to revoke or deny renewal of under-performing charters. If used properly, these policies and procedures increase the odds that only high-quality charter schools will be allowed to open and to continue operating.

Introduction

Media attention toward charter schools tends to either demonize or canonize their practices, and data is regularly marshaled to strengthen the case. In the midst of this often highly charged and confusing environment, policymakers are tasked with important decisions ranging from the fate of individual charter schools to the overall effectiveness of charter schools as a tool for improving student achievement. To responsibly execute their duties, policymakers must have dependable – and understandable – information about charter school performance.

In attempting to provide this information, the research community has delivered an array of studies of charter schools, employing a variety of analytic techniques. When different techniques yield differing results, and the effort is not made to explain the sources of difference, the research community inadvertently contributes to the confusion. Policymakers, who view research from a practical rather than an academic standpoint, need guidance from the research community about the strengths and limitations of the various analytic techniques.

The research on charter schools has not kept up with the increasing importance of charter schooling as a policy alternative. Until recently, the research literature on charter schools was constrained to studies within single states. Accordingly, the insights likewise were constrained to within-state or even within-city variation, in which all schools would be covered by the same policies and changes to them. This led to two flawed inferences: either policy thinkers took each study result to be indicative of the entire charter school population or they rejected the results of studies conducted outside of their own state.

Policymakers, educators and researchers alike need a better understanding of the benefits and limitations of the different study designs used to measure charter school performance. This is a two-way responsibility. On the one hand, decision-makers need to be sensitive to the analytic

methods used when assessing the relative strengths of results from different studies. On the other hand, the research community needs to be clearer in their discussions about the suitability and limitations of the techniques they choose to illuminate various policy questions.

The purpose of this paper is to begin such a discussion of two quasi-experimental methods that have been recently used to measure charter school effectiveness on a large scale. These two techniques are student fixed effects (FE) and virtual control records (VCR), and they are compared here using an unusually large multi-state student-level dataset. Student fixed effects is the most commonly used quasi-experimental method in education research. The virtual control record methodology, in contrast, is a newly developed technique that was designed to address some of the limitations of student fixed effects.

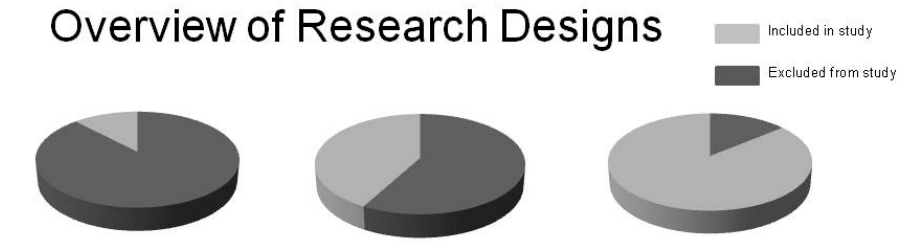
Two issues are central to the debate about analytic methods: selection bias and generalizability. Minimizing the former often comes at the expense of the latter, as in the case of lottery studies, which draw on the experience of a limited number of charter schools that randomly select students for admission. Quasi-experimental methods such as student fixed effects include greater proportions of students than lottery studies but introduce issues of selection bias to varying degrees. What remains an open question is whether generalizability can be improved in the quasi-experimental realm while maintaining low levels of selection bias. The research presented here compares the advantages and disadvantages of two quasi-experimental approaches, student fixed effects and virtual control records, with a focus on the issues of selection bias and generalizability.

To address both issues, two different comparisons were conducted. The first, discussed in Section 2, creates a head-to-head test of FE and VCR designs. The test involves the control conditions of the two methods, restricting the analysis to only the charter students who switch

between traditional public and charter school sectors and the control specifications used in the two approaches. This computational comparison will effectively test whether the students acting as their own controls in the FE model produce equivalent results to those of the virtual controls in the VCR model, i.e., whether selection bias is comparable between the two methods. Section 3 compares the results for FE and VCR models that are not restricted a priori. The biggest difference between the two methods is that fixed effects models are limited to students who are observed in both a traditional public school and a charter school, i.e., sector switchers. The VCR method allows controls to be found for both sector switchers and charter-only students. The policy insights gained from each model will be the main focus of the second comparison.

Methods Descriptions

In recent years, the methods debate in academia has focused on the relative merits of experimental and quasi-experimental research designs. In charter school research, these are most often lottery studies and student fixed effects. Figure 1 summarizes these two methods and their strengths and limitations along with the new quasi-experimental methodology examined in this paper, virtual control records.



Lottery / Random Assignment	Student Fixed Effects (FE)	Virtual Control Record (VCR)
Unit = Schools	Unit = Students	Unit = Students
Narrow subset of older schools included	Resulting sample biased to middle schools	Majority of students included
Random assignment minimizes selection bias – no guarantee on single draw	Students act as own controls, selection bias minimized	Students matched on fixed characteristics and prior achievement
Uncertain ability to generalize results	Do they accurately reflect all students?	Strong ability to generalize to full population

Figure 1. Overview of Research Designs.

Lottery studies

The ideal study would take the population of traditional public school (TPS) students and randomly place some in charter schools while keeping the rest in the TPS sector. Such an experimental design would eliminate selection bias from the estimates of charter school effectiveness. Although there are no cases of true random assignment between sectors, the rules of entry into charters mimic many aspects. In most states, if a charter school is oversubscribed (i.e., more children want to attend than there are spaces available), the charter school must select students at random by holding a lottery for entry. Thus, on average, the losers of the lottery would be expected to be a good comparison group for the winners on both observable and unobservable characteristics such as motivation, thereby giving lottery studies a high degree of internal validity. This approach was employed in recent analyses of charter school effectiveness in Boston and New York City and, in one study, in multiple locations across 15 states.

However, charter school lotteries have an unfortunate limitation: The conditions that give rise to lotteries, namely over-subscription, are neither universal nor randomly distributed across schools. Research suggests that oversubscribed schools are older and academically better than the average charter school (Hoxby & Murarka, 2007; Tuttle, Gleason & Clark, 2012; Abdulkadiroglu et al., 2009). There are also practical limitations. It only takes one student beyond a school's capacity to trigger a lottery, and many schools have subscription rates less than 125 percent of capacity . In these cases, random selection does not produce equal numbers of students to serve as controls, and therefore can lead to problems in the estimation phase. Consequently, only a fraction of oversubscribed schools can fully support a lottery study. Thus, the results that arise from lottery studies are difficult to generalize to the full population of charter schools.

Student fixed effects

The alternative to lottery studies is to rely on quasi-experimental approaches, and the most commonly used approach in the charter school literature is student fixed effects . This method uses a panel dataset and is relatively more inclusive of schools than lottery studies. The design includes only the students who have at least one growth period in each sector: TPS and charter. Each student's TPS experience essentially serves as his or her own control condition for the charter school experience, which minimizes many potential sources of selection bias. One notable exception is the occurrence of an event that coincides with (or immediately precedes) a student's transfer between sectors and also influences the student's learning gains. In this case, estimates of the charter school effect will be biased (Hoxby & Murarka, 2008; Hanushek et al., 2007).

Like lottery studies, there are concerns about the generalizability of student fixed effects studies. First, mobility between the TPS and charter sectors is not extremely high, as will be shown below. Furthermore, as the charter sector expands, the number of students who never attend a TPS school is likely to grow. To the extent that these so-called "always charter" students look different from their sector-switching peers, either in demography or in terms of early performance, generalizing fixed effects results to the entire charter student population may become untenable .

Virtual control records

In an attempt to produce generalizable results on national charter school effectiveness, a new quasi-experimental method was devised. The virtual control record design draws on principles used in propensity score matching and a recently developed synthetic matching technique by Abadie, Diamond & Hainmueller (2007). At its heart is a matching methodology that creates a set of synthetic control records that closely mirrors the matched charter school students on known demographic attributes, grade in school, eligibility or participation in special support programs (subsidized lunch program eligibility, English language learner status and special education status) and a baseline test result.[1] The baseline test result reflects both earlier schooling and family and neighborhood background, including motivation, home-based learning and other inputs.

For each charter school, the virtual control record (VCR) design first identifies all the traditional public schools that had students transfer to that charter school (i.e., "feeder schools"). The students still enrolled in the feeder schools constitute the pool of potential matches. These TPS students are required to meet further criteria to be matched with a particular charter student:

[1] To be included in a synthetic control record, the baseline test result had to be no more than 0.10 of a standard deviation higher or lower than the matching charter student's baseline test score.

1) They must attend a traditional public school for all observed time periods; 2) They must have equivalent values on all the match variables; and 3) They must have all the relevant outcome data. Matches are developed separately for reading and math. For example, a third-grade charter student will be matched to a third-grade TPS student from the feeder pool who has a similar baseline math score and the same gender, race/ethnicity, lunch program eligibility, English learner status, special education status and subsequent grade-level promotion/retention pattern. For all the TPS students who meet all the match criteria, scores from each subsequent test period are averaged across all the matched records, and a single record is produced – the VCR. The result is a composite of the expected value of experience the charter school student would have had if enrolled in the traditional public schools that compete with the charter school.[2]

The advantage of the VCR design is that it permits the inclusion of a far greater proportion of charter school students than either lottery studies or student fixed effects. Only charter students who do not have the requisite minimum of two consecutive test scores will be excluded from the sample pool. Widespread use of unique statewide student identifiers and the federal requirement that all students be tested in at least grades three through eight ensures that relatively few students in elementary or middles schools will fall into this category.

The VCR method faces many of the same challenges for lower elementary and high school students as the other methods, owing to a lack of consistent test data. A concern that has been raised about the VCR design is that it could possibly underestimate the charter treatment effect for students who are exclusively enrolled in charter elementary schools. This is because the approach matches students after they have an initial test score, thereby capturing student achievement midway through the elementary charter school treatment. The fundamental

2 This composite score has less noise than a nearest-neighbor matching scheme with the added benefit of survivability of the controls over time; if a student leaves the TPS sector in subsequent time periods, the VCR can be recalculated without that student.

question is whether performance in earlier (largely unobserved) grades at charter schools is stronger than at traditional public schools. If this is true, the VCR method would lead to an understatement of true charter school performance. The purest way to test this conjecture is to conduct a randomized trial of students in kindergarten and follow them until their first observed score. Absent that approach, we took advantage of a more limited opportunity to look at performance in earlier grades using the data amassed for this study. In Georgia student testing begins in first grade, and California tests start in second grade. In both states, the first observed achievement scores for students exclusively enrolled in charters were higher on average than those for students whose earliest score was recorded while enrolled in TPS. Whether high early achievement in the charter school sector relative to traditional public schools is due to selection bias or early charter school performance – or a combination of both – remains unclear. Two points bear noting. First, the pattern of charter student achievement in the earliest scores for both these states mirrors the patterns for growth we see in the later grades for these same students, so the same general signal is obtained even when the first score is delayed. Second, there is greater variation across states in the performance of charters relative to TPS than we see across the student groups in Georgia and California, leading to the suggestion that the VCR method does an adequate job of representing the impacts of each sector.

In the sections that follow, the FE and VCR methods are compared to determine the relative strengths and limitations of each with respect to generalizability and selection bias. Two different comparisons will be made. The first comparison, in Section 2, focuses on a comparison of the control conditions in the FE and VCR models, which necessitated using the same charter students in both models. To be included, therefore, charter students had to meet multiple criteria: 1) Be included in a regular FE model; 2) Obtain a growth score in TPS before attending a

charter; and 3) Receive a VCR match. In Section 3, the second comparison of the FE and VCR models dispenses with multiple criteria for student inclusion. Instead, the FE model includes all sector switchers regardless of whether they attended a charter or TPS first or whether they have a matching virtual control. The VCR model includes all charter students with a matching virtual control regardless of whether they were included in the FE model. Throughout the rest of the paper, the models used for the head-to-head methods test discussed in Section 2 are called restricted models, while those used in the comparison discussed in Section 3 are called unrestricted models.

Data Description

The laboratory for comparing the two study designs is a large multi-state student-level dataset with records that are longitudinally linked across multiple years. In addition to math and reading test scores, information on student characteristics is included: grade level for each year, race/ethnicity, gender, subsidized meal eligibility (a proxy for economic disadvantage), English language learner status and participation in special education programs.

For this analysis, fourteen states and two large urban districts were followed for four years to measure the effectiveness of charter schools on student achievement growth. [3] Information on the years and grade levels included in the analysis can be found in the Appendix. For both the FE and VCR methods, observations were pooled across states. Consolidating experience across states is not without its challenges. As noted by the National Consensus Panel on Charter School

[3] On a practical note, this study required an approach that met the multiple and conflicting interpretations across states of the Family Education Records Privacy Act (FERPA). The only realistic avenue to conduct a study of this scope was to negotiate agreements with state education agencies for permission to use administrative datasets with student-level records. In spite of changes to the implementation regulations in late 2008, the law remains unclear about the permissibility of providing independent researchers access to student-level data. Several accommodations were imposed as conditions of agreement — though, curiously, each was imposed by only one state. For example, even after all identifying information was removed, one state declined to provide gender on the theory that it prevented identification of individual students. Lacking that information in one state meant that this variable could not be included in the pooled model for any state, so this study was unable to control for gender effects.

Academic Quality , states vary in important ways in their treatment and support of charters. In recognition of this fact, this study sought to examine common elements of charter school performance while simultaneously recognizing that states may play a role in how these schools perform.

Others have suggested that state accountability tests differ, such that scores on a grade-level test in one state may not align with similar scores in another. The study design circumvented these potential difficulties by standardizing test results from each participating state for each grade and year and then calculating a one-year standardized growth score for each student. When scores are thus standardized into z-scores, every student is placed relative to his peers in his own state. A z-score of zero, for example, denotes a student at the 50th percentile in that state, while a z-score one standard deviation above that equates to the 84th percentile. Students who maintain their relative place from year to year would have a growth score of zero. Because both the FE and VCR methods pair control and treatment observations within states, structural problems with testing and scoring differences across states are minimized.

The pooled estimates of overall charter effectiveness, then, will be affected by the sampling fractions included in each state for each method. Table 1 displays, for each location, the total count of individual charter students who had one or more recorded growth scores at a charter school in that location. Also shown are the sampling fractions for the restricted models that will form the basis of the head-to-head test of the FE and VCR methods in Section 2 and the sampling fractions for the unrestricted models to be discussed in Section 3. [4]

As noted above, the head-to-head methods test necessitated that charter students were included only when they were TPS-then-charter sector switchers and a virtual control match

[4] Percentages and counts in the tables in this section are from the math dataset unless otherwise noted. Student subgroups were generally equivalent across the two test subjects.

could be found for them. The proportion of students who met the first criterion is shown in the third column in the table. The proportion of those students who also met the second criterion is displayed in the fourth column. These statistics give some context to the methods test discussion in the following section and also show that finding a VCR match for charter students who have previously attended TPS is slightly easier than finding a VCR match for the average charter student in each state.

Table 1. Unique Charter Students by Location and Sampling Fraction by Comparison Criteria.[5, 6]

State	Unique Charter Students	% Unique students meeting restricted FE criteria	% FE criteria students with VCR	% of Students by state (unrestricted models) FE	VCR
AR	3,184	35.65	90.13	50.16	87.88
AZ	46,372	17.23	85.37	29.91	82.29
CA	265,779	30.71	84.16	44.79	84.70
CO	38,258	21.88	87.74	39.19	90.03
DC	10,955	29.38	87.95	34.58	83.07
FL	84,922	29.38	92.00	50.09	92.64
GA	24,210	25.15	94.27	40.15	93.63
IL (Chicago)	11,842	18.59	95.00	23.93	94.04
LA	10,526	15.74	87.51	24.57	85.50
MA	19,670	22.44	87.34	33.60	80.49
MN	12,077	20.24	83.31	32.23	76.45
MO	8,538	20.06	89.14	24.70	84.22
NC	24,001	19.35	81.46	39.52	80.14
NM	6,873	30.51	80.73	44.06	75.34
NYC	8,453	23.34	92.07	32.00	83.76
NY (upstate)	11,026	14.94	84.48	32.13	84.15
OH	31,836	14.25	80.05	24.55	78.17
TOTAL	618,522	26.02	86.32	40.44	85.53

[5] Throughout this paper, upstate New York and New York City are reported separately. This is done because the student demographics and charter policy are vastly different, which could affect results.

[6] A majority of the states were followed from the 2004-05 through 2007-08 school years, but a handful of areas underwent major testing changes in spring 2006 and thus were followed from the 2005-06 through 2008-09 school years. These are California; Washington, DC; Chicago, IL; Massachusetts; Minnesota; Missouri; and New York City and state.

For the unrestricted models, the VCR method includes a much greater percentage of charter students in every location, ranging from a low of 75 percent in New Mexico to 94 percent in Chicago, IL. The FE method ranges from a 24 percent inclusion rate in Chicago, IL, to 50 percent in Arkansas. Even within states, FE inclusion rates were found to vary widely (due in part to the age of the charter law, the pace of growth in the sector and the number of consecutive grade levels included in the testing regimen). These findings suggest that the results of the FE analysis are highly temporally and location-sensitive. The findings also indicate that results from the VCR method, containing a higher proportion of all charter students, is the more generalizable of the two methods.

The state-by-state sampling fractions in the unrestricted models tell only part of the story about which students are or are not included in the model for each method, however. The whole population of unique charter students is disaggregated by characteristic below, and the proportions of charter students included in the unrestricted fixed effects and VCR models are compared to all charter students for each characteristic. The purpose of this exercise is to determine whether specific limits on generalizability for each method can be described.

Student traits

First, it is important to consider whether the FE and VCR models adequately represent the mix of ethnicities, income levels and program participants present in the charter student population. As seen in Table 2, the VCR population is most similar to the full charter population in terms of these student traits, although the FE population is also generally similar. Both models have roughly the same proportion of black students and students who are English learners as the full charter population. In the FE models, though, white students are underrepresented and Hispanic and poor students are overrepresented by a couple percentage points each. Special

education students and, to a lesser extent, Native Americans are underrepresented in the VCR data. These results suggest no particular strength or limitation for either method in demographic representativeness of the data.

Table 2. Proportion of Students by Trait.

Student Trait	All Charter Students	FE Model	VCR Model
Race/Ethnicity			
Asian	4.28	4.22	3.56
Black	24.63	23.86	24.66
Hispanic	29.64	32.85	30.05
Native Am	1.15	1.16	0.52
White	39.05	36.83	40.46
Multi/Other	1.26	1.08	0.76
Low Income	47.94	50.66	48.02
English Learner	8.61	8.95	7.93
Special Education	9.01	8.69	6.90

First observed score

Another measurable difference among students is their early performance on achievement tests. Because these scores encompass a host of unmeasurable student traits, such as earlier schooling and family and neighborhood backgrounds, alignment of the means of the first observed scores of all charter students and the students included in each model is important to note. Table 3 shows how each method divides the All Charter group into two groups: Those whose observations are included in the model and all remaining students.

Table 3. Mean First Observed Score by Method and Inclusion.

	Math		Read	
	Mean	Variance	Mean	Variance
FE Model				
Included	-0.113	0.94	-0.113	0.94
Not Included	-0.093	0.98	-0.050	1.01
VCR Model				
Included	-0.056	0.89	-0.023	0.90
Not Included	-0.372	1.32	-0.392	1.32

All Charter Students	-0.101	0.97	-0.076	0.98

Table 3 provides numerous insights both within and across the models. Within the FE model, the difference in the means of the students who are and are not included is much smaller than within the VCR model. Further, the students who are not included in the FE model have a higher average than the included group, which also contrasts with the findings for the VCR model groups. However, the difference in the means of the two FE groups is statistically significant in both math and reading. The implication is that the FE models do not accurately represent students from the higher end of the performance distribution.

Within the VCR model, the average first score for students who are not included is much lower in both math and reading than that of the charter students who are included, and the difference between the two groups is statistically significant. The implication here is that the lowest students in the charter school distribution do not have counterparts in the traditional public schools that feed their charter school. The reasons for this were found to be threefold. First, lower-achieving students are more likely to repeat a grade in charter schools than in TPS. These same students are less likely to be labeled as needing special education services in charters than in TPS. Third, there are very few very low performers in each school, regardless of sector, and thus matching is excessively difficult due to demographic heterogeneity. This phenomenon is evident with the highest performers as well; the unmatched group also contains a disproportionate number of the highest achievers, as evidenced by the high variance. An analysis of the match rate by first observed score revealed lower rates at both tails of the performance distribution than in the middle of it. However, the phenomenon affected only the four percent of charter students scoring two or more standard deviations above or below the mean. The VCR

method, then, appears to be more representative of students in the middle of the distribution than of the distribution as a whole.

Comparing only the included groups across the models, the average for the VCR group is significantly higher than the FE group in both subjects. Although t-tests of each model's included group average compared to the all-charter-student group average yielded p-values that were significant at the 0.01 level in every instance, the average for the FE model group is closer to the all-charter-student average than is the VCR group. This suggests that, in terms of first observed test scores, the FE models are more representative of the charter student population than the VCR models.

First observed year

One of the largest differences in the composition of the two analytic datasets is the distribution of students across years, which is seen in Table 4. As discussed earlier, charter students can be matched in the VCR method as soon as they have two consecutive test scores (one growth period), regardless of the sector in which the first score was attained. Thus, roughly 80 to 90 percent of students in each year are included, as was seen in Table 1 above. In contrast, students must have three consecutive test scores (two growth periods) to be included in the FE models; further, at least one growth score must be attributed to each sector. These additional requirements result in a dataset that is more heavily weighted toward the earlier years of available data. While it is possible that a longer time horizon would increase the percentages of students included in the FE models, the FE requirements will always fail to include large proportions of recent school entrants.

Table 4. Annual Student Count and Inclusion Rates.

		% of Students by Year	
First Observed Year	N	FE	VCR

2004-05	185,150	49.56	89.23
2005-06	305,048	47.45	85.22
2006-07	90,564	15.06	81.71
2007-08	37,734	0.00	78.98

The inability of the FE design to obtain estimates from recent years has serious implications. Given that available data is almost always a full year out of date, FE models can only provide reasonable estimates at a significant lag, shown here to be three or four years. This represents a significant drawback for policymakers, since the models lack information about the effects of the newest charter schools or recent quality shifts in existing charter schools. Results from the FE method will also fail for several time periods to adequately capture the most recent charter policy developments, such as the replication of successful charter schools or the proliferation of charter conversions from failing traditional public schools.

Attendance patterns

Results for charter student performance may also vary depending on students' attendance patterns into and out of the charter and traditional public school sectors. In the full dataset, about 43% of students attend only charter schools. About the same proportion, 46% attend traditional public schools before transferring to the charter sector. The remaining 11% attend charter schools before they transfer to TPS. Table 5 displays the percent of each group included in the VCR and FE subsets.

Table 5. Attendance Patterns and Inclusion Rates.

		% of Students by Attendance Pattern	
Attendance Pattern	N	FE	VCR
Always Charter	265,370	0.00	84.80
TPS then Charter	284,611	65.34	86.43
Charter then TPS	68,515	93.66	84.56

The VCR method includes about the same proportion of students from each attendance group. This is not the case for FE, however, since only sector switchers are eligible for this method. Interestingly, the two sector switcher groups have vastly different proportions of inclusion in the FE models. A deeper analysis of the students in the "TPS then Charter" group revealed a variety of attendance possibilities under this broad category, including in about one-third of the cases a one-year stint at a TPS, which is insufficient to obtain a TPS growth period. In contrast, the students in the "Charter then TPS" category nearly always obtain a growth score in each sector.

To fully unpack the implications of the differences in inclusion rates for each attendance pattern, it is useful to determine whether test scores differ for the three groups. Table 6 below displays the results of the mean first test score for all charter students as well as each attendance pattern. Breaking the full charter group into subsets by their attendance pattern, differences in starting scores become apparent. Students who have charter experience before attending a traditional public school have higher average scores in both math and reading than students who attend TPS before transferring to a charter, and t-tests show that these differences are statistically significant at $p<0.01$. In reading, the "Always Charter" and the "Charter then TPS" groups are also statistically different from each other at $p<0.01$.

Table 6. Mean First Observed Score by Attendance Pattern.

First Observed Score	All Charter Students	FE	VCR
Math			
All Students	-0.101	-0.113	-0.056
Always Charter	-0.060		-0.010
TPS then Charter	-0.149	-0.132	-0.104
Charter then TPS	-0.067	-0.061	-0.026
Read			
All Students	-0.076	-0.113	-0.023
Always Charter	0.004		0.055

TPS then Charter	-0.155	-0.138	-0.098
Charter then TPS	-0.037	-0.035	0.011

As discussed above, each of the three attendance patterns is adequately represented in the VCR models, although the VCRs have statistically higher first scores than the full population in each case. The FE models represent two of the three attendance patterns and do so more closely than the VCR subset for each pattern. For FE, only the "TPS then Charter" averages are statistically different than the full average for that pattern. In this case, then, the FE subset more closely resembles the average first score of the full charter student dataset. To the extent that charter students' scores diverge along attendance patterns over time, however, the FE method may not remain representative of the full charter population's starting score.

Summary of data description

Both methods include similar proportions for many of the various student demographics, although each method differs on at least one dimension. The VCR method is less likely to include students with a first observed test score at the very low or very high end of the distribution. The FE method has lower rates of inclusion at each geographic location than the VCR method, which is due in large part to the inability of the FE method to include students who are observed only in the charter sector. A further limitation of the FE method is that new charter schools and new entrants are not likely to be included in sufficient proportions for the results to be representative of the most recent policy changes. As the two approaches have different strengths and weaknesses, findings in areas of agreement between the methods' analytic results should be more robust than those found with only one method.

Test of Methods

At the heart of the debate of the FE and VCR methods is whether they produce equivalent results. Because the two methods include different charter students, it is necessary to construct a head-to-head comparison using the same charter students, i.e., TPS-then-charter sector switchers who receive a virtual control match. By using these restricted models, the control conditions of the FE and VCR methods can be directly compared.

(1) $\Delta z_{i,t}^{Ch} - \Delta z_{FEi,(t-1)}^{TPS}$ *or* $\Delta z_{i,t}^{Ch} - \Delta z_{FEi,(t+1)}^{TPS}$

(2) $\Delta z_{i,t}^{Ch} - \Delta z_{VCRi,t}^{TPS}$

In the FE method, z-score growth for charter student *i* in period *t* is compared to a separate growth period in the TPS sector for the same student, as seen in Equation (1). In a regular FE model, the TPS experience can precede (i.e., period *t-1*) or succeed the charter experience (i.e., period *t+1*). For the VCR method, however, the z-score growth for charter student *i* in period *t* is compared to the growth of the student's virtual control record in the same time period, as shown in Equation (2). In other words, the VCR method uses a differences-in-differences approach. To ensure time symmetry across the two methods, only charter students with preceding TPS experience were included in the models for this comparison.[7] Thus, the historical TPS experience of the charter student in the FE method will be compared with the contemporaneous TPS experience of the VCR.

For both methods, controls for the previous test score, year, grade level and an indicator if the growth score was measuring a repeated grade were included in the regression models. The use of each student as his own control obviates the need for additional demographic variables in the FE models. The VCR models include controls for time period, student demographics and

[7] About 8% of students whose first growth period is observed in TPS switch back to TPS after having growth in charter. In these cases, the average of the two TPS growth periods are used as the control condition in the FE calculations.

state indicators. For the marginal effects model in each method, variables were added to isolate the charter effect for each subgroup. The results of the regression models are shown in Table 7. Although the estimates are not useful to a general discussion about charter school effectiveness, the FE estimates can be compared to the VCR estimates to evaluate the two methods.

Table 7. Overall and Marginal Charter Effects by Method using Restricted Models.

	Math		Read	
	FE N=389,961	VCR N=351,616	FE N=428,018	VCR N=381,096
All Students	-0.074**	-0.063**	-0.034**	-0.024**
	(0.003)	(0.002)	(0.002)	(0.002)
Marginal Effects				
Race/Ethnicity				
Black	-0.035**	-0.040**	-0.035**	-0.035**
	(0.005)	(0.005)	(0.004)	(0.004)
Hispanic	-0.073**	-0.062**	-0.026**	-0.032**
	(0.004)	(0.004)	(0.004)	(0.004)
White	-0.174**	-0.165**	-0.095**	-0.075**
	(0.004)	(0.003)	(0.003)	(0.003)
Low Income	0.024**	0.051**	0.024**	0.034**
	(0.004)	(0.004)	(0.003)	(0.004)
English Learner	0.109**	0.048**	0.072**	0.042**
	(0.006)	(0.007)	(0.005)	(0.006)
Special Education	0.058**	0.024**	0.012*	0.017*
	(0.006)	(0.008)	(0.006)	(0.008)

Robust standard errors in parentheses; *p<0.05 **p<0.01
N is equal to the total number of observations included in the model.

Overall and for every subgroup, the FE and VCR models return estimates that are the same sign and have the same level of significance. Many of the coefficients across the two models are quite similar, such as those for black students. Further, neither method produces consistently larger or smaller coefficients. Since the charter students are exactly the same in the two methods, the differences seen in these coefficients are due to the differences in the control conditions: the charter students' historical TPS experience in the FE method versus the current TPS experience

of the controls in the VCR method. These differences are particularly evident in the English Learner and Special Education coefficients. The smaller coefficients in the VCR models relative to the FE models seems to indicate that the traditional public school sector is experiencing gains with these two student groups over time, which are not captured by the FE models.

These results indicate that the two methods produce reasonably similar estimations of charter school effectiveness when restricted to the same charter students. Based on these results, we can infer that differences in the estimates in the following section, which use unrestricted models, will be due to differences in the charter student populations included in each method.

Policy Analysis

The purpose of this section is to determine the policy insights that can be gained by comparing the differences in the estimations from the FE and VCR models. The models in this section are unrestricted; the FE models include all sector switchers regardless of whether they have a matching virtual control and the VCR models includes all charter students with a matching virtual control regardless of whether they were sector switchers. Both types of models take the general form

$$\Delta A_{i,t} = \theta A_{i,t-1} + \beta X_i + \gamma C_{i,t} + \varepsilon_{i,t} \quad (3)$$

where the dependent variable, growth in student achievement, is

$$\Delta A_{i,t} = A_{i,t} - A_{i,t-1} \quad (4)$$

when $A_{i,t}$ is the z-score for student *i* in period *t;* $A_{i,t-1}$ is the z-score for student *i* in period *t-1;* X is a set of control variables for student characteristics, states and years; C is an indicator variable for whether the student attended a charter in period *t;* and ε is the error term. Lagged achievement ($A_{i,t-1}$) is included as a control variable because growth is not necessarily equivalent

across the spectrum of prior achievement levels, due in part to the phenomenon of regression to the mean (Ladd & Walsh, 2002). As with the prior model, the FE models include only time-variant control variables, which are a subset of those included in the VCR models. For both FE and VCR models, ordinary least squares regressions were run for pooled models and for individual locations.

Results from the pooled models are displayed in Table 8. Both the FE and VCR models show that charter students, on average, have statistically lower growth than their TPS counterparts in math. The result for reading is split, however; the FE estimate is negative and significant, while the VCR estimate is positive and significant, though small. This finding is indicative of a larger pattern between the two methods: Overall and for virtually all the subgroups in both math and reading, estimates from the VCR method are more positive than those from FE. This suggests that the students who are excluded from the FE method have higher growth than those who are included.

Table 8. Overall and Marginal Charter Effects by Method using Unrestricted Models.

	Math		Read	
	FE	VCR	FE	VCR
	N=19,905,212	N=1,692,236	N=19,780,459	N=1,765,464
All Students	-0.065**	-0.016**	-0.049**	0.007**
	(0.001)	(0.001)	(0.001)	(0.001)
Marginal Effects				
Race/Ethnicity				
Black	-0.025**	0.008**	-0.037**	0.007**
	(0.003)	(0.002)	(0.003)	(0.002)
Hispanic	-0.057**	-0.020**	-0.040**	-0.001
	(0.003)	(0.002)	(0.003)	(0.002)
White	-0.131**	-0.062**	-0.095**	-0.013**
	(0.002)	(0.001)	(0.002)	(0.001)
Low Income	0.016**	0.029**	0.017**	0.019**
	(0.00234)	(0.002)	(0.002)	(0.002)
English Learner	0.049**	0.034**	0.041**	0.028**
	(0.00404)	(0.003)	(0.004)	(0.003)
Special Education	0.013**	0.019**	0.007	0.006
	(0.00379)	(0.004)	(0.004)	(0.004)

Robust standard errors in parentheses; *p<0.05 **p<0.01
N is equal to the total number of observations included in the model.

Both methods return positive results for charter schools for three groups: low income students, English learners, and special education students. In addition, the VCR estimates for black students in both subjects and all students in reading are small but positive and significant at p<0.01. The difference between the FE and VCR coefficients in these instances may indicate a positive trajectory for these measures, since the VCR models include the experience of more recent charter school entrants.

The estimates for English learner students do not fit the pattern of VCR estimates being more positive than the estimates from FE. The TPS controls in the VCR method are contemporaneous to the charter growth observations, a phenomenon that allows the performance of the traditional public school sector to adjust over time. It is possible, then, that the lower estimates from the VCR method reflect a positive change in performance for English learners in charters at the same time that English learners in traditional public schools are also posting positive gains.

Variation in results by location

The real insight from each of the models is that the effectiveness of charter schooling is widely varied. To illustrate variation across jurisdictions, the unrestricted models were re-estimated by location. Individual location results are displayed in Table 9, while the observations for each method at each location can be found in the Appendix. The estimates between the two methods are more similar for math than for reading in terms of sign and significance. Six of the 17 locations post positive and significant estimates in the FE method in math; all of those locations, in addition to Chicago, have positive and significant VCR estimates. In contrast, only one state has positive and significant estimates for both methods in reading. For this subject, the

VCR estimates indicate that charter students in eight locations have statistically higher growth than their TPS peers. More generally for both subjects, where there is agreement between the methods on sign and significance, the estimates from the VCR method paint a more favorable picture of charter school student performance. The differences reflect the inclusion in VCR estimates of students who have only attended charter schools and those who have only a brief period in TPS.

Table 9. Charter Effects by Location.

	Fixed Effects		VCR	
State	Math	Reading	Math	Reading
AR	0.051**	0.018	0.049**	0.017
	(0.012)	(0.012)	(0.011)	(0.010)
AZ	-0.102**	-0.055**	-0.046**	-0.013**
	(0.004)	(0.005)	(0.003)	(0.003)
CA	-0.047**	-0.018**	-0.026**	0.018**
	(0.002)	(0.001)	(0.001)	(0.001)
CO	-0.077**	-0.061**	-0.028**	-0.020**
	(0.003)	(0.004)	(0.002)	(0.002)
DC	0.043**	-0.008	0.105**	0.067**
	(0.011)	(0.012)	(0.007)	(0.006)
FL	-0.038**	-0.036**	-0.021**	-0.021**
	(0.002)	(0.003)	(0.002)	(0.002)
GA	-0.088**	-0.058**	-0.017**	0.005
	(0.006)	(0.006)	(0.004)	(0.004)
IL (Chicago)	-0.062**	-0.073**	0.012**	0.019**
	(0.008)	(0.009)	(0.004)	(0.005)
LA	0.060**	0.089**	0.043**	0.055**
	(0.011)	(0.011)	(0.007)	(0.007)
MA	0.017**	-0.021**	0.064**	0.027**
	(0.006)	(0.007)	(0.004)	(0.004)
MN	-0.087**	-0.034**	-0.028**	-0.010
	(0.008)	(0.008)	(0.006)	(0.005)
MO	-0.198**	-0.145**	0.000	-0.004
	(0.015)	(0.014)	(0.007)	(0.007)
NC	-0.133**	-0.049**	-0.042**	0.018**
	(0.005)	(0.005)	(0.004)	(0.004)
NM	-0.091**	-0.044**	-0.047**	-0.024**
	(0.009)	(0.010)	(0.007)	(0.008)
NYC	0.072**	-0.044**	0.113**	0.049**
	(0.009)	(0.010)	(0.005)	(0.005)
NY (upstate)	0.235**	0.019	0.149**	0.025**

	(0.011)	(0.014)	(0.007)	(0.006)
OH	-0.156**	-0.085**	-0.058**	-0.005
	(0.006)	(0.006)	(0.004)	(0.004)

Robust standard errors in parentheses; *p<0.05 **p<0.01

Individual schools

State or city differences do not account for all the variation in performance. Some of the variation occurs at the local market level. To isolate these effects, we define highly localized markets, consisting of the students in each charter school and the public schools from which it draws students. This highly localized analysis can provide the answer to the question: Do charter schools create student outcomes that are on par with their local competition?[8] Because the appropriate sampling fraction for each charter's competitors is unknowable in the FE models, that method is unable to support a school-by-school quality estimator.

The analysis employed market fixed effects to remove all the effects that are shared by a charter school and its traditional public school market, as represented by the charter's VCRs. This model takes the functional form

$$(5) \quad \Delta A_{i,t} = \theta A_{i,t-1} + \beta X_i + \gamma^c C^c_{i,t} + \rho^m M^m_{i,t} + \varepsilon_{i,t}$$

where γ^c is the marginal effect of charter *c* given the average performance of market *m*. These individual charter school coefficients convey the marginal performance of each school relative to its market. A summary of these coefficients is displayed in Table 10.

Table 10. Summary of Marginal Effects for Individual Charter Schools.

Subject	Number of Charter Schools	Positive and Significant	Not Significant	Negative and Significant
Math	2,242	19.1%	47.8%	33.1%
Reading	2,234	19.1%	59.7%	21.2%

[8] Note that this analysis is purely relative and gives no information about the overall level of each market's performance or whether the existing market performance is acceptable.

$p<0.05$

In both math and reading, about one-fifth of the charter schools have significantly higher average growth than their TPS market counterparts. The preponderance of charter schools in each subject are on par with the performance in their local market, however high or low that market performance may be.

Summary of unrestricted models

The results of the models, when taken together, indicate that the story of charter school performance is one of uneven quality on many dimensions. The pooled result for all students is negative in math, but there is cause for cautious optimism in reading. Positive and significant results are found for multiple subgroups of charter students have while negative and significant effects are found for other subgroups. This variation is also evident when results are disaggregated by location. Charter school students outperform their TPS counterparts in one-third to one-half of the geographic locations, depending on the subject tested and method used. (The picture is much more positive for charters in reading using the VCR method.) The proportion of positive results for individual schools in math and reading is smaller still.

Conclusions

The present study was aimed at two purposes: to construct a rigorous test of the attributes of student fixed effects models against those employing virtual control records, and to provide a comparison of the estimates of charter school effectiveness when employed in a policy analysis. While one is hesitant to draw generalizations from a single comparison of the two techniques, the insights that the analysis provides can help guide researchers and decision makers alike in using studies featuring each approach.

For charter school research, the student fixed effects methodology is the predominant quasi-experimental technique for minimizing selection bias in regression estimates. By using a

student as his/her own control, albeit temporally displaced, the fixed effect approach ensures that all time-insensitive observable and unobservable characteristics are equivalent in treatment and control conditions. The possibility of selection bias into one or the other group is thus minimized.

Alternatively, the VCR method uses all available observable characteristics and prior performance – which captures motivation and family supports -- to create a composite comparison record. An important attribute of the VCR method is that the period of comparison between charter students and their controls is contemporaneous; the control condition occurs in parallel with the year of charter school study. This feature could be important when studying jurisdictions that are undergoing significant and rapid change.

The head-to-head comparison of the FE and VCR methods used the same charter students to create a computational test. Since the charter students were the same in both methods, the comparison essentially tested for equivalence between the FE control (e.g., the charter student's own TPS experience) and the VCR. Under these conditions, the two methods produced highly similar results. Coefficients for the charter effect were identical in sign and significance and of the same general magnitudes. Further, the differences in size of coefficient did not consistently favor either method. Therefore, it can be concluded that the influence of selection bias in the VCR method is about the same as in the FE method.

In addition to selection bias, a chief concern about the usefulness of an analytic method is the generalizability of the results produced by it. Each of the two methods discussed here has some limitations in this area, as does any analytic method that uses samples of a population. The fixed effects methodology appears to have more limits on generalizability than the VCR methodology. The students covered by the FE method comprise about 40 percent of the charter

student population. FE is limited to sector switchers, which means its estimates can provide no insight into students who are educated solely in the charter sector. The proportion of students included in FE models is shown to be influenced by the maturity and growth rates of the charter sectors, as these factors affect the flow of students from one sector to the other. Further, the results of many of the newest entrants in the charter school market are not included in the FE estimates. In contrast, the VCR method includes about 85 percent of charter students. Students who first score in the extreme tails of the achievement distribution are difficult to match and are less likely to be included in the VCR method than those in the middle of the distribution. However, the match rates on the low end still exceed 60 percent, and the high end still has a match rate of 75 percent. On the low end, less than three percent of the students are affected; on the high end, about one percent of the students are affected. On balance, the generalizability of the VCR method is more favorable than occurs with FE methods.

This paper has shown that the strength of the FE method lies in selection bias minimization. The VCR method is shown to be equivalent with respect to selection bias and provide results that are more generalizable. However, no single analytic method can provide satisfactory answers to every research question. Therefore, the layering of evidence from multiple methods is always preferred. When that is not possible, it is important for researchers to select the most appropriate method for their research objectives and to be explicit about the strengths and limitations of the chosen technique. FE appears to be the preferred method for nontraditional students and when sector switchers are a sizable and representative proportion of the overall charter student population, while the VCR method seems more appropriate for students who attend only charter schools, charter-versus-feeder analyses and estimates of more recent charter school phenomena. Table 11 describes some possible research objectives and

notes whether fixed effects or virtual controls is likely best suited to accomplish each based on the findings in this paper. This table, while useful for researchers, can also be used by policymakers to help assess the research results presented to them, which in turn will help pave the way for more-informed policy decisions.

Table 11. Best Method by Research Objective.

Research Objective	FE	VCR
Provide best estimate of students with the lowest starting scores	X	
Provide best estimate of students with the highest starting scores	X	
Provide best estimate of students with a special education designation	X	
Provide estimate of "Always Charter" students		X
Include largest proportion of charter students possible		X
Provide best estimate when sector switchers are substantially different from non-switchers		X
Provide best estimate when sector switchers are representative of overall charter population	X	
Provide estimate of head-to-head charter effect versus their feeder market		X
Provide timely estimates of changes in charter school policy		X
Provide timely estimates of new charter school effectiveness		X

In the area of charter school policy, more information to policymakers cannot come too soon. Although the focus of this research was to compare two quasi-experimental methods that have been used to measure charter school effectiveness, the results also serve to reinforce what has become a familiar refrain in recent charter school research: Charter school quality is demographically and geographically uneven. For the charter school movement to survive and thrive, more needs to be done to improve and tighten the quality distribution. Because it is also clear that state charter school policies influence quality, research looking at the policy similarities and differences between high- and low-quality locations could help determine the drivers of high quality and point to specific policy improvements.

Acknowledgements

The authors are grateful for helpful comments from Eric Hanushek, Susanna Loeb, Sean Reardon, Edward Cremata and anonymous reviewers as well as support from colleagues at CREDO. Financial support from the Packard Humanities Institute is also gratefully acknowledged.

Appendix A

Table A1 below displays the number of unique charter students for each location as well as the observations by method. Although the fixed effects method appears to include many more observations than the virtual control method, this is not in fact the case. In the fixed effects models, all but the sector switchers drop out of the estimation of the charter effect, thereby bringing the effective observation count much closer to that of the VCR method.

Table A1. Observations by Location and Method.[9]

State	School Years Included	Grades Included	Unique Charter Students	FE	VCR
AR	2004-05 to 2007-08	3-8	3,184	479,669	8,124
AZ	2004-05 to 2007-08	3-12	46,372	989,716	116,494
CA	2005-06 to 2008-09	2-11	265,779	1,054,833	722,520
CO	2004-05 to 2007-08	3-10	38,258	1,100,752	115,498
DC	2005-06 to 2008-09	3-8	10,955	60,693	26,670
FL	2004-05 to 2007-08	3-12	84,922	3,867,055	251,274
GA	2004-05 to 2007-08	1-8	24,210	2,370,835	74,022
IL (Chicago)	2005-06 to 2008-09	3-8	11,842	383,007	36,798
LA	2004-05 to 2007-08	3-9	10,526	807,170	26,938
MA	2005-06 to 2008-09	3-8	19,670	1,033,531	54,396
MN	2005-06 to 2008-09	3-8	12,077	798,255	28,028
MO	2005-06 to 2008-09	3-8	8,538	60,583	22,308
NC	2004-05 to 2007-08	3-12	24,001	1,992,393	65,678
NM	2004-05 to 2007-08	3-9	6,873	365,767	14,918
NYC	2005-06 to 2008-09	3-8	8,453	977,980	32,162
NY (upstate)	2005-06 to 2008-09	3-8	11,026	1,950,978	24,036
OH	2004-05 to 2007-08	3-8	31,836	1,612,971	72,372
TOTAL			618,522	19,906,188	1,692,236

[9] Due to the inability to acquire the full set of California data, the fixed effects models for this state included all observations for sector switchers, all observations of students educated solely in charters and VCR observations standing in for the missing observations of students educated solely in TPS.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *SSRN eLibrary* (2007).

Abdulkadiroglu, Atila; Josh Angrist, Sarah Cohodes, Susan Dynarski, Jon Fullerton, Thomas Kane, Parag Pathak. "Informing the Debate: Comparing Boston's Charter, Pilot and Traditional Schools." Boston, MA: The Boston Foundation 2009.

Ballou, Dale, Bettie Teasley, and Tim Zeidner. "Charter Schools in Idaho." In *Charter School Outcomes*, edited by Mark Berends, Matthew G. Springer and Herbert J. Walberg. New York: Taylor & Francis Group, LLC, 2008.

Gleason, Philip; Clark, Melissa; Tuttle, Christina Clark; Dwoyer, Emily. "The Evaluation of Charter School Impacts: Final Report." Washington, DC: Institute for Education Sciences, 2010.

Hanushek, E. A., J.F. Kain, S.G. Rivkin, and G.F. Branch. "Charter School Quality and Parental Decision Making with School Choice." *Journal of Public Economics* 91 (2007): 823-48.

Hoxby, Caroline, and Sonali Murarka. "New York City's Charter Schools Overall Report." Cambridge, MA, 2007.

Hoxby, Caroline; Murarka, Sonali. "Charter Schools in New York City: Who Enrolls and How They Affect Their Students' Enrollment." NBER Working Paper No. 14852, 2009.

National Consensus Panel on Charter School Academic Quality. "A Framework for Academic Quality: A Report from the National Consensus Panel on Charter School Academic Quality." Building Charter School Quality Initiative, 2008.

Peterson, Paul E. *The Future of School Choice*. Stanford, Calif.: Hoover Institution Press, 2003.

Tuttle, Christina Clark, Philip Gleason, and Melissa Clark. "Using Lotteries to Evaluate Schools of Choice: Evidence from a National Study of Charter Schools." *Economics of Education Review* 31, no. 2 (2012): 237-53.

Zimmer, Ron, and Richard Buddin. "Charter Schools in California." In *Charter School Outcomes*, edited by Mark Berends, Matthew G. Springer and Herbert J. Walberg. New York: Taylor & Francis Group, LLC, 2008.

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