Charter School Growth and Replication

1/30/2013

Technical Appendix

CENTER FOR RESEARCH ON EDUCATION OUTCOMES

Table of Contents

Table of Figures

Table of Tables

Volume I: Technical Appendix

This study employs panel data sets containing longitudinally linked student records for each state. Because states vary in their treatment and support of charters, and because state tests do not align with one another, analyses that include multiple states must create a common basis for comparison. To this end, we have standardized test scores into standard deviation units, or z-scores, based on the statewide distribution of scores for each accountability test. Thus, every student is placed relative to her peers in her unique testing population (e.g. third grade reading, or seventh grade math in a given academic year).

These standardized results can be combined across years to estimate the progress each student makes relative to the average progress for the student's grade, subject and growth year. For purposes of this study, the outcome measure of interest was annual school-wide average student-level growth on math and reading accountability tests.

Fixed Effect Models School fixed effect studies take advantage of the longitudinal nature of the student-level data set. For every individual student who has two or more test score records, this technique can be deployed to discern changes in student performance that are time-variant by holding all the non-varying student characteristics constant over time. School fixed effect techniques therefore isolate the value-added impact, or growth, of each school by treating all unchanging school- and student-level characteristics as constant.

To estimate school-by-year fixed effects, we used an errors-in-variables regression (the *eivreg* command in STATA) with robust standard errors. This model factors in variations in test reliability and student clustering within the charter population. Reliability was calculated using states' own reported standard errors of measurement (SEMs) on state accountability tests. The SEM is a measure of test score error reported in standard deviation units for each test. A student's observed score should be interpreted as an estimate of her true score, and SEMs define the standard error band around this estimate. Conditional standard errors of measurement, or CSEMs, refer to SEMs that are conditioned on students' observed ability levels. CSEMs are unique to each year/test/score combination. Because ability is more difficult to measure accurately at the highest and lowest levels of achievement, the CSEM distribution resembles an inverse test score distribution curve, with higher values at the tails and the lower values at the mean. Wherever possible, we used CSEMs to calculate average test score reliability for the population studied. In some cases, CSEMs reported for key cut points were used to interpolate values for the entire distribution of scores. In the very small number of states where CSEMs were not available in any form, we substituted overall reliability estimates for the entire test as reported in states' annual Technical Reports.

Worth noting is that, because performance in the charter student population did not always – or even often – concentrate at the statewide mean, our calculated estimates of test reliability were frequently much lower than statewide estimates. The *eivreg* command incorporates this reduced reliability into its estimate of variable error, resulting in modified, more precise effect size coefficients, larger standard error estimates, and an increased threshold for statistical significance.

Our regression also controlled for starting score, student-level characteristics, and grade-level concentrations of poverty and mobility. A specific list of independent variables is as follows:

- **Prior Performance:** Z-score (achievement level) in prior year, with error-in-variable reliability estimate
- **Race/Ethnicity:** dummy variables indicating that student is Caucasian/White, Black, Hispanic/Latino, Asian/Pacific Islander, American Indian, or Multiracial
- **Gender**
- **Special Program Participation:** dummy variables indicating Special Education status, English Language Learner status, and participation in the federal free or reduced lunch program (a standard proxy for poverty)
- **Peer Effects**: Percentage of students receiving free/reduced lunch by grade for each school in each year, and percentage of new student enrollment by grade for each school in each year

Excluded Observations Our overall school counts included 3,613 unique charter schools in math and 3,616 schools in reading. The National Alliance for Public Schools estimates that there were 4,913 charter schools in operation in the 2009-10 school year ("Public Charter Schools Dashboard," 2012), suggesting that our analysis represents approximately 74% of charter schools nationwide. This estimate may need to be adjusted downward slightly, because our sample includes schools that were closed prior to the 2009-10 school year, but conservative estimates still place the represented proportion at over 70% of charter schools nationally.

There are two exclusion rules for the sample:

1) Schools with fewer than 30 student observations were omitted. If a unique school in a given year did not have 30 students with prior scores, it was excluded on the grounds that we could not accurately calculate an average student effect size for that school. This constraint may affect new charter schools disproportionately, as they often have small starting classes.

2) Schools with no tested grades in a year were excluded. If school enrollment in a given year (school age) does not include any tested grades, it was per force excluded from the analysis for that year. This constraint affects high schools and elementary schools in slightly different ways. In most states, high school students are not tested in every grade, which makes the estimation of year-by-year performance more susceptible to shifts in student characteristics. We did include measures of 8th-to-10th grade growth from states that do not test their 9th graders, but high schools in states that only test in 11th or 12th grade, with no intervening test administrations, were excluded from the sample. In the elementary school case, the inclusion of so-called "slow grow" elementary schools is delayed until these schools, which grow one grade at a time, mature to the point that students are enrolled in grades that are covered by the state accountability program. This typically occurs in year 4 or year 5 of operation, as elementary schools that open with K and K-1 are both considered slow-grow. Our concern about relative differences in impact between schools that open one grade at a time ("slow grow") and schools that open with all grades ("full grow") was sufficient to warrant a secondary exploratory analysis. See Appendix B of Volume I of the report: Sensitivity Tests on Growth Strategy for full account of methodology and results.

Selection of Comparison School School-by-year effect sizes in each state were calculated relative to a synthetic charter reference school. This school mirrored the average charter school demography in the state, in order to avoid biasing estimates in our model. Prior score and achievement level independent variables, however, were created using statewide means; school-by-year effect sizes are therefore based on each state's mean for all schools. The statewide five-year grand mean was used in order to establish a single common benchmark, or "zero line," across all charter maturation levels regardless of calendar year.

Because the statewide mean performance was used to ground estimates, school-by-year effect sizes represent charter school quality relative to the statewide average, and not quality relative to the local market average. Thus, a concentration of negative effect sizes in a given state may be attributable to charter sector location restrictions or differentiation, rather than – or in addition to – pure differences in quality. For example, charters located in mainly urban areas are being compared to the statewide, and not urban, average.

Decile Comparisons

In volume 1 of this report, we include tables which give the conditional probability for a school in any given starting quintile to move into either the low performing level (quintiles 1 -2) or the high performing level (quintiles 3-5) in the next period. The tables which follow include the number of schools assigned to each starting decile for each of Tables 9 – 20 in volume 1. The headings below give in parentheses the number of the table from volume 1 to which the values apply.

Table 1: Number of Schools per Starting Quintiles by Age, Reading (Table 9)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	117	129	128	143	132
2	86	99	165	173	193
3	71	132	189	180	190
4	86	94	151	197	163
5	76	98	150	187	128

Table 2: Number of Schools per Starting Quintiles by Age, Math (Table 10)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	110	122	131	135	136
2	88	111	157	155	201
3	77	116	166	203	195
4	81	87	149	180	176
5	81	104	154	158	134

Table 3: Number of Schools per Starting Quintiles by Age, Reading (Table 11)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	18	11	25	24	47
2	19	14	33	34	69
3	21	15	30	42	43
4	27	17	30	53	35

Table 4: Number of Schools per Starting Quintiles by Age, Reading (Table 12)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	22	19	30	26	53
2	20	22	34	41	74
3	22	20	38	52	60
4	26	18	32	42	54

Table 5: Number of Schools per Starting Quintiles by Age, Elementary Reading (Table 13)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	22	21	27	29	18
2	10	18	28	29	31
3	10	21	25	23	44
4	11	12	20	34	30
5	11	12	29	35	15

Table 6: Number of Schools per Starting Quintiles by Age, Elementary Math (Table 14)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	11	16	25	35	29
2	6	15	18	32	46
3	1	7	22	43	50
4	5	6	23	33	40
5	2	14	22	41	22

Table 7: Number of Schools per Starting Quintiles by Age, Middle School Reading (Table 15)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	14	26	31	31	32
2	10	21	25	35	40
3	3	25	31	29	29
4	1	9	33	28	22
5	3	9	16	28	14

Table 8: Number of Schools per Starting Quintiles by Age, Middle School Math (Table 16)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	13	26	23	33	39
2	11	18	27	25	51
3	9	17	24	35	33
4	2	7	23	25	36
5	1	8	15	28	18

Table 9: Number of Schools per Starting Quintiles by Age, Multi-Level Reading (Table 17)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	53	74	54	58	36
2	49	49	90	88	79
3	40	67	104	103	84
4	42	57	83	100	92
5	34	66	78	109	80

Table 10: Number of Schools per Starting Quintiles by Age, Multi-Level Math (Table 18)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	60	51	61	53	40
2	36	58	86	86	86
3	41	62	86	109	97
4	37	51	79	108	92
5	41	65	92	80	90

Table 11: Number of Schools per Starting Quintiles by Age, High School Reading (Table 19)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	28	8	16	25	46
2	17	11	22	21	43
3	18	19	29	25	33
4	32	16	15	35	19
5	28	11	27	15	19

Table 12: Number of Schools per Starting Quintiles by Age, High School Math (Table 20)

Age of School	Starting Q1	Starting Q2	Starting Q3	Starting Q4	Starting Q5
1	26	29	22	14	28
2	35	20	26	12	18
3	26	30	34	16	15
4	37	23	24	14	8
5	37	17	25	9	4

Sensitivity Tests

We conducted a number of sensitivity tests during the preliminary phase of analysis in order to assess and refine the model. These tests were primarily conducted on large states with large charter populations. In order to remain compliant with our state partner agreements, we do not present results directly; however, we do our best to summarize the relevant findings below. This section presents findings in the following format: the guiding concern for each sensitivity check, the test performed, the results of this test, and any follow-up action taken.

Test #1

Guiding Concern A charter-only dataset, while much easier to manipulate due to its substantially smaller size, could potentially bias our baseline estimates and limit our ability to control for major sectoral differences (and changes in those differences) between the traditional public schools (TPS) and charter sectors. Of particular concern were potential differences in charter impacts on prior scores and poverty.

Sensitivity Test We ran an exploratory baseline model on statewide data sets that included all students, with coefficients on prior score and poverty stratified by attendance at TPS or charter. We were thus able to explore potential differential impacts in these two variables.

Result No evidence of meaningful differential treatment impacts was discovered overall or by grade.

Action Taken We felt confident in our decision to limit the model to the charter student population only.

Test #2

Guiding Concern Structural shifts in the charter sector over time may limit the validity of using a single common baseline of comparison across school ages and calendar years.

Sensitivity Test We calculated mean starting scores of new students separately by age *and* calendar year, thus enabling a search for patterns in the data indicative of problematic structural shifts.

Result There were often steady structural differences between charters and TPS (for example, charters generally pulling from one end or the other of the statewide distribution), but we failed to find systematic evidence of major sectoral shifts over time.

Action Taken We settled upon Year 1 quintile cut points as a common yardstick for comparison.

Test #3

Guiding Concern Student-level controls, in combination with school-by-year fixed effects, may fail to control adequately for relevant differences among students and schools due to classroom-based peer effects or other exogenous shocks.

Sensitivity Test Using our baseline fixed effect model estimates, we created a post-estimation data set of mean effect sizes by concentration of race/ethnicity or program participation (SPED, ELL, Free/Reduced).

Result There was no evidence of systemic error. There was no clear relationship or clustering of performance after controlling for student-level means. There were some statistically significant differences at the tails of the performance distribution, as described in the body of the report, but these were quite small in magnitude.

Action Taken Our continuing concerns about an increasingly dynamic charter market in some states led us to include controls for grade-level concentrations of poverty and new non-entry grade students, despite a lack of clear evidence for their necessity. Poverty-by-grade is consistently not statistically significant; new-student-concentration by grade is typically significant and negative, but extremely small in magnitude.

Test #4

Guiding Concern Again, sectoral shifts in new schools over time could bias school-by-year fixed effect estimates in unpredictable ways. To isolate the impact of maturation properly, was it necessary to limit the school fixed effects model only to those schools for which there is data from their first year of operation forward ("year 1-forward"), or is it preferable to include all schools in each state, even those that are far past the critical first five year period ("all-in")?

Sensitivity Test We performed side by side comparisons of models that were all-in and year 1-forward only.

Result The coefficients did not differ dramatically between the two models, but they were slightly smaller in magnitude in the all-in model. This difference can likely be attributed to the much larger number of observations in this version of the model, which in turn leads to a more precise variance-covariance matrix.

Action Taken We used our all-in model in order to calculate more precise estimates of effect size.

Test #5

Guiding Concern Post-estimation analysis of the performance trajectory of new schools could be impacted by the inclusion of our all-in fixed effect estimates. Schools that first appear in our data set at later ages, including so-called "slow grow" schools and schools that opened prior to 2005/06, may affect our ability to discern performance trajectory accurately due to structural differences among such schools or changes in the charter sector over time. Our concern was greatest at the elementary school level, due to the systematic exclusion of slow-grow schools.

Sensitivity Test We created post-estimation probability tables for year 1-forward schools only – the purest, most uncomplicated group – and then compared trajectories between slow-grow and non-slow grow schools, broken out even further by grade span. Then, we compared these tables to our all-in tables.

Result It did appear that slow grow schools may have better starting positions and slightly "stickier" trajectories going forward. The inclusion of schools for which we do not have Year 1 data for reasons other than slow-grow status (typically, due to the natural limits set by having data from 2005-2010 only) did not appear to impact estimates of performance trajectory.

Action Taken Due to continuing concerns about underrepresentation of slow-grow schools in our data set, we undertook a separate, exploratory analysis of slow-grow schools in five representative urban communities. The results of this analysis point to a generally stronger and tighter performance among slow-grow schools, particularly at the middle school level. A fuller account of this sensitivity test appears as a separate appendix.

Volume II: Technical Appendix

Consolidating Student Data from Multiple States

This study combines parallel data from multiple states into a single pooled dataset. The model controls for state baseline characteristics which allows for comparisons by school types both at the state and national level. By comparing values on a common standardized scale, z-scores, this study removes many of the comparability issues common with multi-state analyses.

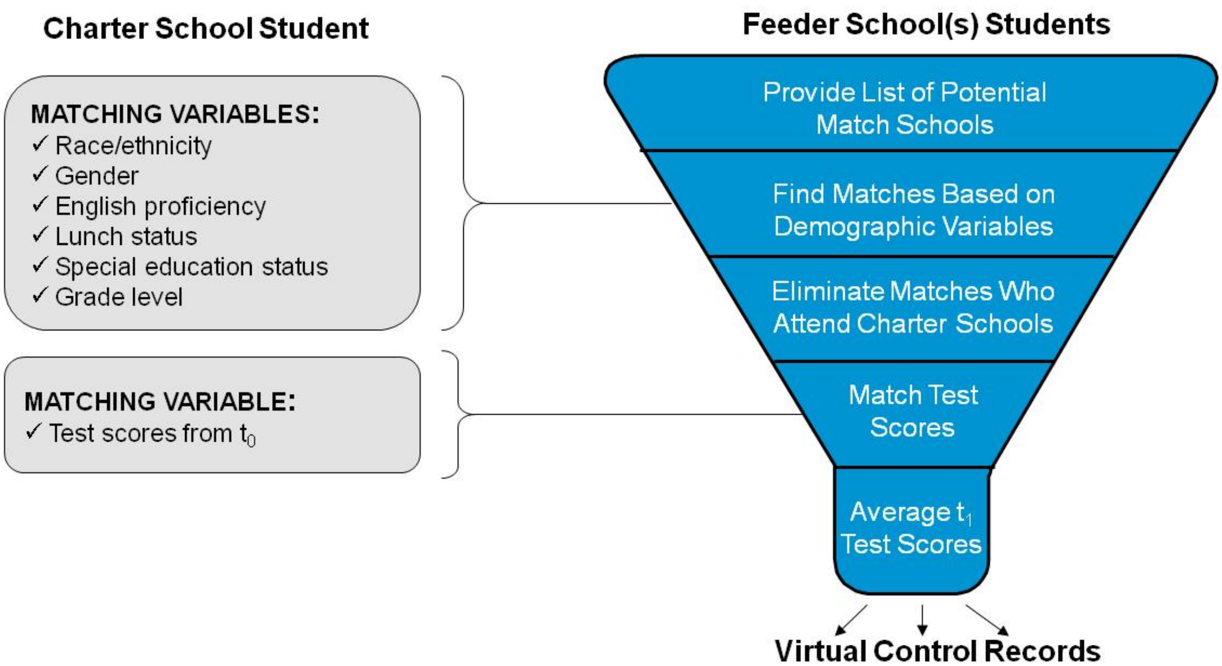
While efforts were made to standardize the test scores used in this analysis, the reliability of each state's test can vary greatly. Therefore, while the methods used make the results generally comparable, it is worth noting that additional analyses which use additional states or additional test years may lead to different outcomes. Further, making comparisons across states assumes that grade level in each state represents mastery of similar materials and concepts; thus, a student scoring near the mean in performance in one state would likely perform similarly in another state.

On a practical note, this study required an approach that meets 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 is to negotiate agreements with state education agencies for permission to use administrative datasets with student-level records. 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 is unable to control for gender effects.

Selection of Comparison Observations. To create a reliable comparison group for our study, we attempted to build a Virtual Control Record (VCR) for each charter school student. The goal of the matching process is to create a virtual twin for each treatment student in the data set. By combining the characteristics of the TPS students selected on exact matches of observed characteristics, we generate a virtual twin with unobserved characteristics highly likely to be similar to the charter students' characteristics. The VCR approach builds on work from Harvard University (Abadie, Diamond et al. 2007).

To reduce concern about selection effects and the accompanying bias, students in this study were matched to students from the same "feeder schools", TPS schools previously attended by students in each particular charter. Limiting the possible control matches to the "feeder schools" essentially limits the comparisons to being between the school the charter student actually attends and the schools the charter's students would have attended in absence of school choice. This method provides a stronger counterfactual as residential school assignment commonly used to place students in schools will increase the likelihood that the students will have similar backgrounds, similar knowledge of school choice programs, and similar school choice options.

Figure 1: CREDO VCR Methodology



CREDO's approach is displayed in Figure 1. We identify all the TPS that have students who transfer to a given charter school; these are the "feeder schools." Once a school qualifies as a feeder school, all the students in the school become potential matches for a student in a particular charter school. All the student records from all the feeder schools are pooled – this becomes the source of records for creating the virtual match. Using the records of the students in those schools in the year prior to the test year of interest, CREDO selects from the available records those with exact characteristic matches to each charter school student based on the following factors:

- Grade-level
- Gender[1]
- Race/Ethnicity
- Free or Reduced Price Lunch Status
- English Language Learner Status
- Special Education Status
- Prior test score on state achievement tests

[1] Gender is used as a match factor in all states except Florida.

Each VCR is thus a composite made up from as many as seven matches. For charter students with more than seven exact matches, seven TPS students are randomly selected to make up the VCR.

The scores from the test year of interest are then averaged and a VCR is produced. That record is completely masked, because there is no trace of from which specific feeder school the contributing records originated. The VCR produces a score for the test year of interest that corresponds to the expected value results of matching techniques used in other studies, such as propensity matching. Table 13 shows the percentage of charter students from each state in this report who were matched under CREDO's matching schema.

Table 13: Percent of Charter School Students with Matches

State	Reading	Math
AR	89.6%	81.6%
AZ	84.4%	83.9%
CA	89.1%	82.6%
CO	88.8%	88.6%
DC	82.8%	80.2%
FL	91.8%	91.9%
GA	91.6%	88.5%
IL*	86.9%	87.6%
IN	84.7%	82.9%
LA	87.3%	88.1%
MA	76.7%	81.1%
MI	83.9%	85.5%
MN	78.0%	71.0%
MO	81.0%	81.7%
NC	80.9%	74.6%
NM	74.4%	74.0%
NY	82.3%	81.0%
NYC	85.0%	84.2%
OH	78.3%	78.7%
OR	80.3%	81.6%
PA	85.0%	84.9%
RI	77.0%	73.4%
TN	96.0%	94.4%
TX	89.6%	90.1%
UT	91.5%	85.5%
Pooled Average	87.2%	84.9%

*Data is limited to Chicago Public Schools students.

Because the VCR match process pairs each charter student with a virtual twin, the number of TPS comparison records is always identical to the number of charter students. Because the VCR virtual twin has identical characteristics to the charter student (with the exception of charter enrollment), the composition of restricted samples for subpopulation analyses will always remain one-to-one matches between charter and VCR records. Since each charter student and their VCR are identical on observable characteristics, this also means the treatment and control samples will always have identical racial/ethnic, socio-economic, and urbanicity characteristics regardless of how the data is subset.

VCRs are re-examined in every subsequent test period to ensure that the match conditions still persist – namely that the students included in the VCR record are still enrolled in traditional public schools and have not left the state. Where the conditions are violated, the VCR is reconstructed to delete the disqualified TPS student records.

A number of things can contribute to a charter school student not finding a match. Students who are new to a community and have no prior history will not be matched for the first year. For some students, all the initial matches are invalidated in subsequent time periods due to school changes among the TPS students. The tight limits that are placed on starting scores also create a hindrance to matches; however, this has the positive effect of greatly reducing the number of students with values extreme enough to qualify as an outlier.

Data Set Properties

For this study, CREDO used student level test data in math and reading from 23 states, New York City, and Washington DC. The test data used was from the years 2006 through 2010. This five year span produced four growth periods which are show in tables 14 and 15 below. Table 14 gives the number of math charter students included in the analysis by year and state.

For the purposes of this report, the time period denoted “2007” covers growth between the 2005-2006 and 2006-2007 school years. This period can also be thought of as the growth from the spring 2006 test to the spring 2007 test. The time period denoted “2008” corresponds to the year of growth between the 2006-2007 and 2007-2008 school years and so forth for the remaining years. In other words, the label refers to the post-test spring term of each growth period, not the spring of the initial testing year.

Table 14: Charter School Observations by State and Year, Math

State	2007	2008	2009	2010	Total
AR	1,747	2,264	3,054	3,470	10,535
AZ	19,134	22,816	26,423	27,827	96,200
CA	94,978	118,487	135,808	153,067	502,340
CO	18,862	22,011	24,161	24,979	90,013
DC	3,787	5,935	6,697	6,561	22,980
FL	41,538	45,192	50,957	58,512	196,199
GA	11,734	16,401	16,477	20,667	65,279
IL*	4,913	5,899	7,612	0	18,424
IN	3,252	3,984	5,974	6,789	19,999
LA	4,159	7,550	10,141	12,476	34,326
MA	8,548	10,466	10,765	11,414	41,193
MI	31,302	33,919	35,280	35,247	135,748
MN	3,754	4,574	5,916	6,965	21,209
MO	2,645	3,967	4,799	5,428	16,839
NC	10,207	12,197	11,238	13,573	47,215
NM	2,692	2,152	2,717	0	7,561
NY	3,478	4,252	4,802	5,312	17,844
NYC	4,134	5,846	7,875	10,291	28,146
OH	13,363	14,382	16,014	19,189	62,948
OR	2,282	4,199	5,048	6,398	17,927
PA	0	18,460	20,970	22,127	61,557
RI	654	779	879	841	3,153
TN	1,260	1,611	2,001	2,758	7,630
TX	44,216	52,119	61,364	72,157	229,856
UT	8,967	11,119	11,780	15,394	47,260
Total	341,606	430,581	488,752	541,442	1,802,381

*Data is limited to Chicago Public Schools students.

Table 15 gives the number of reading charter students included in the analysis by year and state.

Table 15: Charter School Observations by State and Year, Reading

State	2007	2008	2009	2010	Total
AR	1,471	1,880	2,588	2,998	8,937
AZ	19,226	22,501	25,323	28,071	95,121
CA	112,739	138,864	157,682	176,706	585,991
CO	18,814	21,945	24,127	24,969	89,855
DC	3,885	6,086	6,850	6,918	23,739
FL	42,874	46,853	52,948	60,284	202,959
GA	11,220	15,725	17,032	20,135	64,112
IL*	4,912	5,832	7,513	0	18,257
IN	3,305	3,943	5,963	6,060	19,271
LA	4,138	7,484	10,022	12,398	34,042
MA	7,735	9,770	10,402	11,000	38,907
MI	30,584	33,268	34,702	34,533	133,087
MN	3,746	6,927	8,601	9,715	28,989
MO	2,614	3,954	4,804	5,292	16,664
NC	10,307	12,103	11,272	13,126	46,808
NM	2,735	2,128	2,739	0	7,602
NY	3,322	4,126	4,407	4,777	16,632
NYC	3,983	5,603	7,467	9,401	26,454
OH	13,233	14,360	15,963	19,118	62,674
OR	2,453	4,229	4,952	6,029	17,663
PA	0	18,409	20,945	22,823	62,177
RI	667	793	865	891	3,216
TN	1,502	1,843	2,255	3,245	8,845
TX	44,203	52,288	61,102	71,977	229,570
UT	11,099	13,413	14,608	18,077	57,197
Total	360,767	454,327	515,132	568,543	1,898,769

*Data is limited to Chicago Public Schools students.

This study uses a value-added approach by measuring student achievement growth on state achievement tests in both reading and math while imposing controls for student demographics and eligibility for categorical program support such as free or reduced-price lunch and special education.

The challenge of all quasi-experimental analysis of charter school performance is to create comparison samples which as much as possible reduce differences between charter school students and their TPS counterparts. We have made efforts in both the matching process and through the use of statistical modeling to produce an unbiased comparison sample which will serve as a strong counterfactual to charter school attendance.

Table 16 shows the number of charter schools in the country and our data set by year. As we did not expend the man hours required to assign CMO and EMO values to schools in states not included in this study, we cannot estimate what percentage of the national CMO or EMO populations are included in our dataset. Hopefully in the future we will be able to secure data agreements with all 50 states and Washington DC which would allow us to compute figures for all CMO and EMO schools.

Table 16: Number of Charter Schools in US and CMO Study by Year

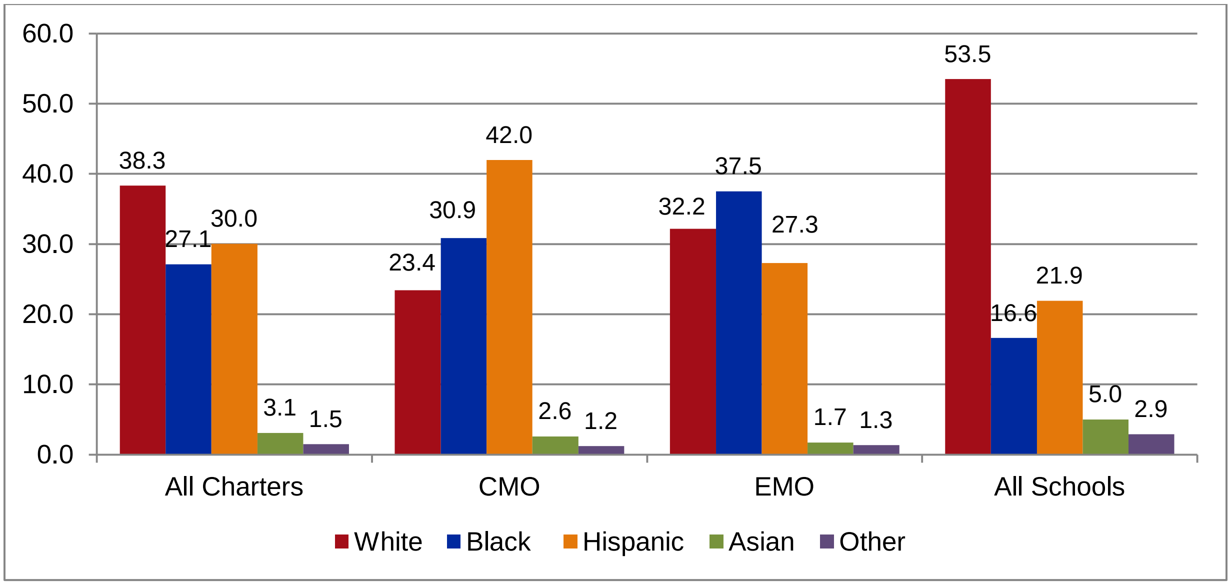
Year	2006/07	2007/08	2008/09	2009/10
US Total	4258	4511	4789	5063
CGAR Total	2450	2865	3082	3112
Percentage	57.5%	63.5%	64.4%	61.5%

Student Characteristics. The following section will provide descriptive information for the three charter types in our analyses. These charter types are all charter school students (including CMO and EMO students), CMO students, and EMO students. Additionally, we include data for all schools in the US where applicable.[2] This will permit the reader to gauge the extent to which the samples represent the general population of students. While these data are interesting and provide insight as to the ability to generalize our findings to all schools, it is worth noting that our statistical models for these analyses included controls for these characteristics. Therefore, differences between the groups should not affect the outcomes of this study as those characteristics have already been controlled for in the statistical models.

Among the students in our data set, we see differing enrollment patterns based on the type of charter school the student is attending (see Figure 2). Of particular interest is the much smaller percentage of CMO enrollment made up by white students as opposed to all charter schools in our dataset. The large drop in white students as a percentage of enrollments in CMOs is offset by a slight increase in black students and a large increase in Hispanic students as a percentage of CMO enrollments. For EMOs, the differences in enrollment are due to an increased in the enrollment of black students who make up the largest percentage of EMO enrollment.

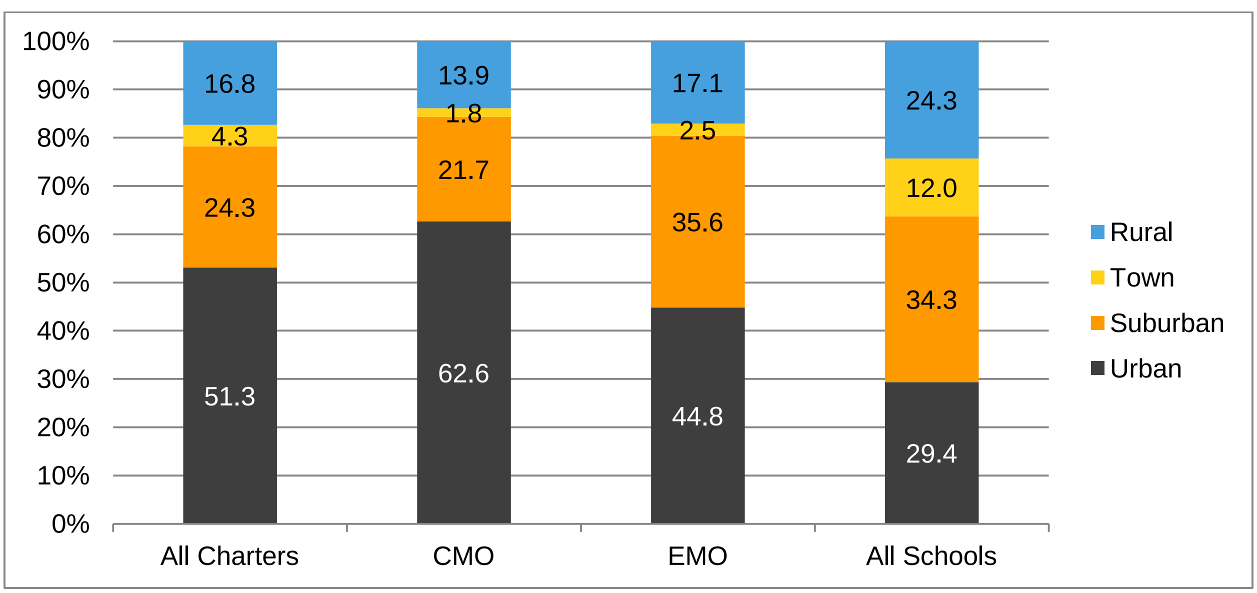
[2] The all schools information is based on data from the National Center for Educational Statistics (NCES) and includes both charter and TPS students across all grades.

Figure 2: Race/Ethnicity as a Percentage of Total Observations by Charter Type



For this study, we used the National Center for Education Statistics definition of locale to classify students by urbanicity. In all three types of charter schools, urban students made up the largest group by enrollment. Figure 3 shows that charter schools serve far more urban students than all schools combined (charter and TPS) US schools population. We found that suburban students make up a much larger portion of the student body for EMO schools than for CMO or charters in general. Figure 3 shows that EMO charter schools have a suburban enrollment much more in line with all schools.

Figure 3: Urbanicity of Charter Students by Charter Type



Because urban students make up such a large proportion of charter school students, we felt it would be informative to exam the racial/ethnic makeup of students across the urban sector to determine if the urban population in charter schools differed from the urban student population in general. Figure 4 shows the racial/ethnic makeup of the urban sector by school type. We find that charter schools regardless of type serve a much larger proportion of black students than all schools combined. Further white urban students attend CMO network schools in our sample at a rate half that of urban schools in general and black students attend EMO network schools in our sample at over twice the rate of all schools combined.

Figure 4: Race/Ethnicity as a Percentage of Total Observations by Charter Type, Urban Students Only

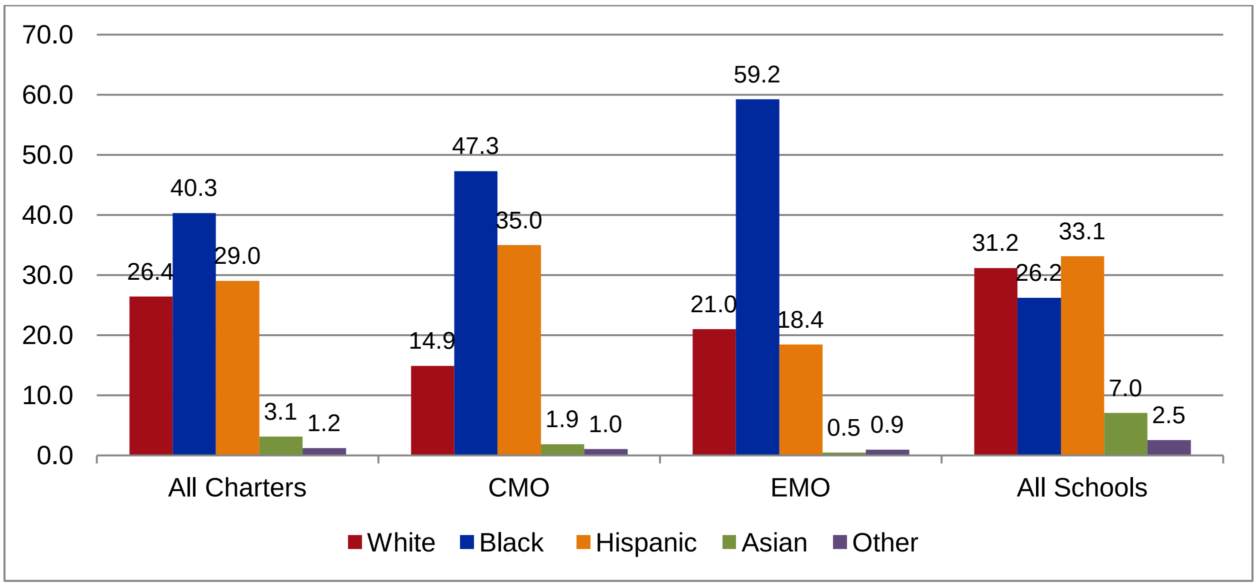
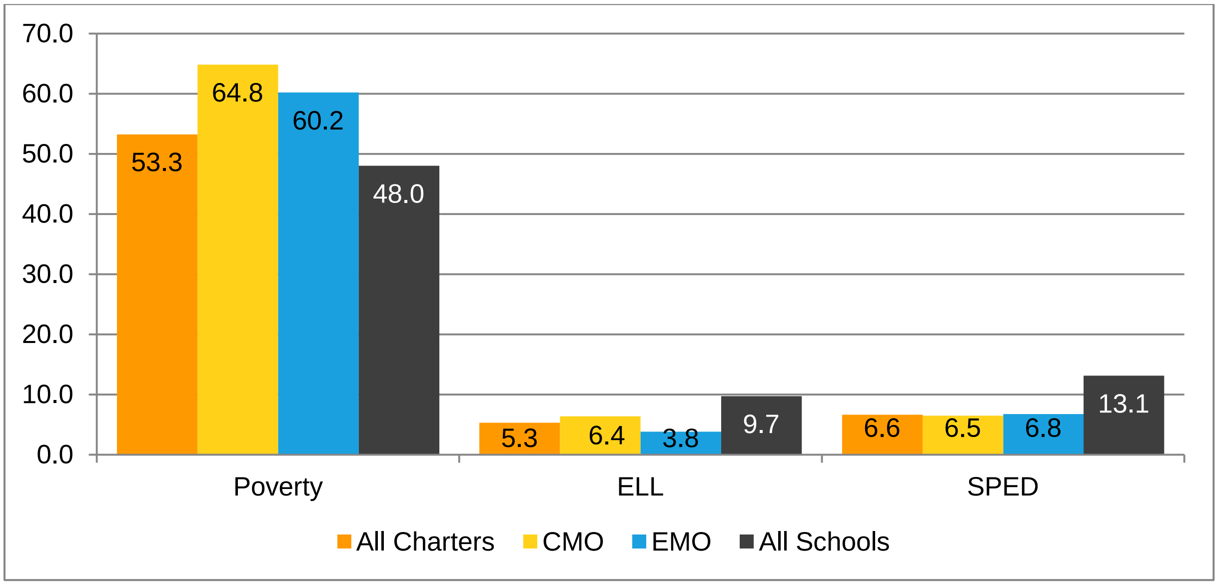


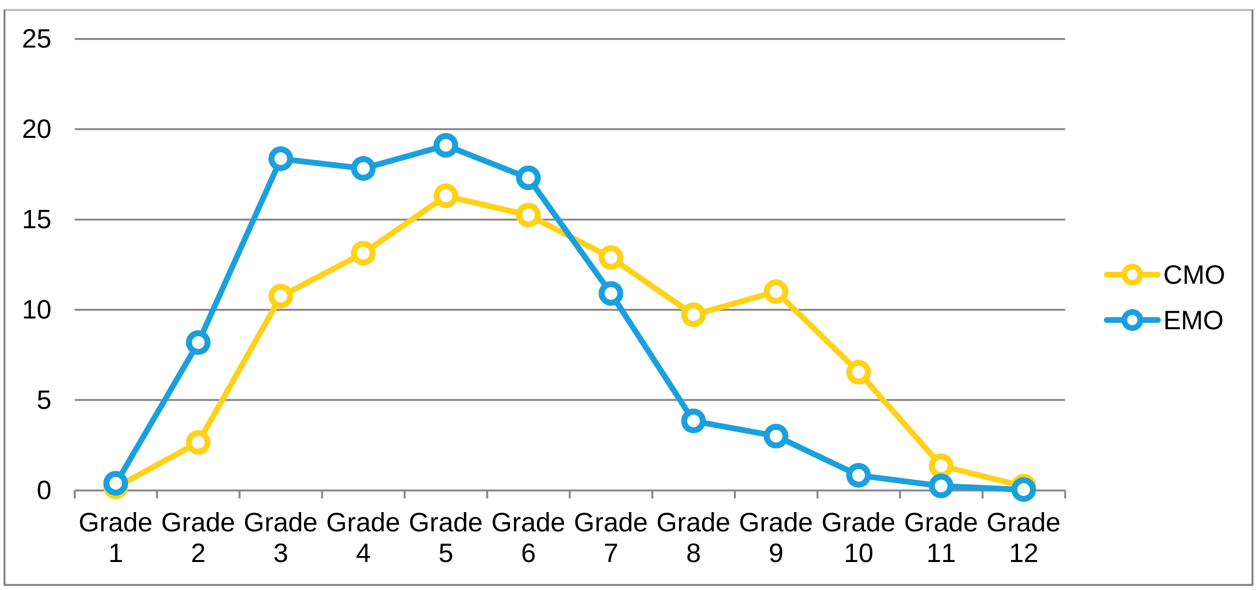
Figure 5 shows the percentage of students in each type of charter school who belong to categorical subpopulations of students. Both CMOs and EMOs serve a population which has 5 to 10 percentage points more students in poverty than the typical charter and all three types of charters serve more students in poverty than the national average. Additionally, CMO schools serve a slightly higher percentage of ELL students than other charters. This difference is to be expected since Hispanic students make up a larger percentage of CMO schools' enrollment. All three categories of charter schools had similar levels of SPED enrollment. Charter schools in our sample served a lower percentage of ELL and SPED students but a higher percentage of students in poverty than all US schools combined.

Figure 5: Percentage of Observations of Students in Categorical Subpopulations



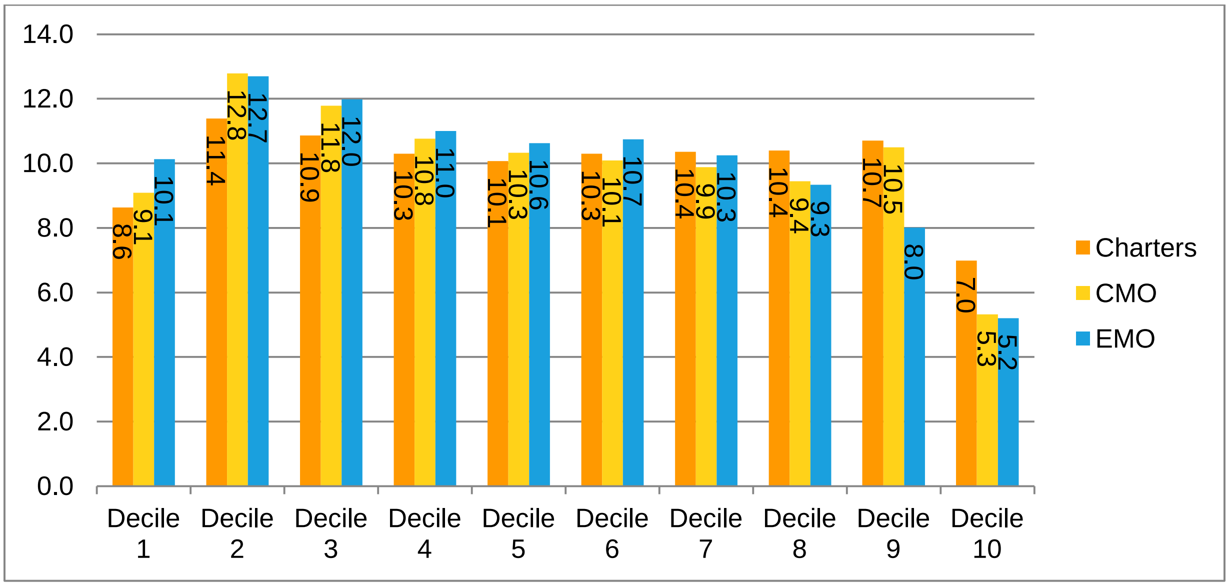
In this sample, a higher percentage of senior high charter students attended CMO network schools than attended EMO network schools. This is the opposite of the trend in elementary and early middle school where a higher percentage of charter students enroll in EMOs. Both types of charter school networks enrolled a higher percentage of elementary students than high school students.

Figure 6: Percentage of <u>Charter Observations</u> Enrolled in CMO and EMO Schools by Grade



Additionally, it appears that CMO and EMO schools also have more lower-performing students than charters or TPS. Figure 7 shows the percent of students in each decile by type of charter school. All Charters, the blue bar, has a distribution pattern which we would expect to see with the values for all deciles being near 10%. CMO (yellow bars) and EMO (blue bars) have a slight positive skew. This means there are more lower-performing students enrolled in CMO and EMO network schools than in all charters or in TPS schools. This is represented in Figure 7 by the longer yellow and blue bars in the lower deciles and the shorter yellow and blue bars in the higher deciles.

Figure 7: Charter School Enrollment by Overall Student Achievement Deciles



School Characteristics

The distribution of schools by levels differs by school type. Figure 8 shows that across all charter schools there are fewer elementary level schools than among TPS schools. This is in part due to the existence of more multilevel charter schools than multilevel TPS schools. We find similar distributions among CMO network schools. However, the opposite is true of EMO network schools. We find more elementary level schools within the EMO networks as well as more multilevel schools. This fits with the student grade data presented in Figure 6 above which shows a majority of EMO network students are in grades from the elementary school level.

Figure 8: Distribution of School Levels by School Type, 2009-2010

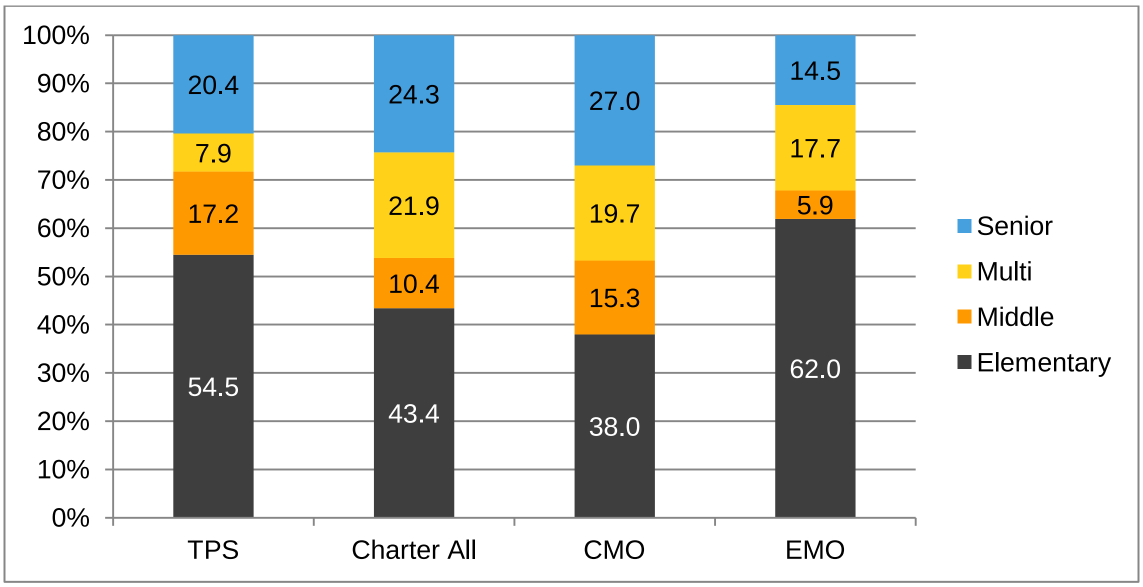


Table 17 lists the number of schools per state by type of school. California had the most charter schools followed by Texas, but Texas had more CMO charter schools. Three states New Mexico, Rhode Island, and Utah did not have any CMO charter schools. Florida and Michigan had the most EMO charter schools. These figures show that the distribution of CMO and EMO affiliated charter schools is not the same as the distribution of charter schools in general. For example, while Michigan has only 6% of the total charter market, Michigan schools make up 23% of the EMO market. For this reason, all our statistical models in this study will include controls for the state. In our models, we omit New Mexico as it has a smaller number of students and average student growth values in both reading and math which were very near the national average student growth value.

Table 17: Number of Charter Schools per State by Type

State	Number of Charters	% of Charters	Number of CMO	% of CMO	Number of EMO	% of EMO
AR	31	0.73%	9	0.85%	0	0.00%
AZ	465	11.00%	147	13.95%	38	9.25%
CA	860	20.35%	193	18.31%	17	4.14%
CO	169	4.00%	21	1.99%	9	2.19%
DC	81	1.92%	37	3.51%	8	1.95%
FL	424	10.03%	113	10.72%	111	27.01%
GA	78	1.85%	9	0.85%	4	0.97%
IL	45	1.06%	20	1.90%	7	1.70%
IN	55	1.30%	19	1.80%	4	0.97%
LA	72	1.70%	23	2.18%	6	1.46%
MA	65	1.54%	1	0.09%	3	0.73%
MI	254	6.01%	25	2.37%	96	23.36%
MN	180	4.26%	14	1.33%	4	0.97%
MO	39	0.92%	9	0.85%	7	1.70%
NC	98	2.32%	2	0.19%	5	1.22%
NM	58	1.37%	0	0.00%	0	0.00%
NY	36	0.85%	3	0.28%	5	1.22%
NYC	63	1.49%	14	1.33%	8	1.95%
OH	298	7.05%	91	8.63%	57	13.87%
OR	102	2.41%	6	0.57%	1	0.24%
PA	115	2.72%	12	1.14%	5	1.22%
RI	7	0.17%	0	0.00%	0	0.00%
TN	22	0.52%	2	0.19%	0	0.00%
TX	535	12.66%	284	26.94%	6	1.46%
UT	74	1.75%	0	0.00%	10	2.43%

Table 18 includes counts for the number of schools associated with EMO companies in our dataset. These counts include the number of EMO charter schools per state as well as a breakout by the non-profit status of those charter schools. The majority of EMO charter schools are run as for-profit ventures.

Table 18: Number of EMO Schools per State by Non-Profit Status

State	Number of EMO Schools	Number of Non-Profit EMO Schools	Number of For-Profit EMO Schools
AZ	37	10	27
CA	17	5	12
CO	9	0	9
DC	7	6	1
FL	68	3	65
GA	4	0	4
ID	2	0	2
IL	2	2	0
IN	8	5	3
KS	1	0	1
LA	5	0	5
MA	3	0	3
MI	106	1	105
MN	3	0	3
MO	5	0	5
NC	5	0	5
NV	1	0	1
NY	18	0	18
OH	56	16	40
OR	1	0	1
PA	5	0	5
SC	2	0	2
TX	16	12	4
UT	10	0	10
WI	3	0	3
Total	394	60	334

Analytic Models

While fixed effects models have become the standard practice for quasi-experimental studies of charter school effectiveness, a fixed effect model would have been a particularly poor model for this study. Instead, this study used the VCR method. The primary reason for this choice is that VCR models preserve a much larger data sample than fixed effects models. The fixed effect model only detects effects from students who switch from one sector to another such as from TPS to charter. However, in this study we also wished to compare the growth of students who attend CMOs or EMOs to charter and TPS students. As the number of students who switch from TPS to charter and from independent charters to CMO or EMO charters is limited, a fixed effects model would have greatly reduced the sample size of the analyses. Further, previous research (Davis and Raymond 2012) has found VCR models to be comparable to fixed effects models for analyzing charter school impacts.

The analyses in this report relied primarily on multiple regression models. Because we were using state level test scores with known errors, we conducted the regressions using the error in variables regression command from Stata 12.1. This regression method includes an estimate of the standard error associated with each state's test instrument.

All of our regressions included controls for student characteristics such as race, socio-economic status, ELL status, and SPED status. We did not include controls for gender as this data was not provided by one of the states. All regressions also included state dummy variables to control for differences between states. The omitted value was New Mexico. New Mexico was chosen because it had a value very near the national average in both math and reading.

The results of our network analysis did not include a Bonferroni adjustment to the critical values. Typically, when an analysis includes multiple tests for significance such as the individual regression results by network, we would adjust down the p-value for each analysis to account for the fact that we are running multiple tests and expect five percent of them to be significant by chance. However, with the high number of networks in this analysis, the Bonferroni adjustment would set the p-value necessary for significance unreasonably low.

We also did not adjust our results for clustering. This is typically done in studies which aggregate individual data up to a higher level such as the classroom- or school-level. Clustering must be done to address issues caused by having a high and positive inter-cluster correlation between the data points. However, evaluations of results obtained under the VCR method have shown the data points to have low and slightly negative inter-cluster correlations. As a result, applying clustering techniques to our VCR results would have the opposite effect from what is expected. Clustering would more data points into the tails thus increasing the likelihood of finding significant values. We have therefore determined that clustering is not an appropriate procedure to apply to these analyses.

Analysis Results

Table 19 shows three coefficients for each categorical subpopulation. The top coefficient is the difference between members of the subpopulation and the typical TPS student. The middle coefficient is the marginal difference between the categorical eligible students who attended a CMO affiliated school and the categorical eligible TPS students. The full effect of comparing the categorical eligible CMO student to a non-eligible TPS student in a single number requires adding the two marginal coefficients. This is shown in the third row for each section below.

Table 19: Difference in Growth by Exceptional needs Category and Charter Attendance, CMO

		Math	Reading
Poverty	Poverty	-0.097**	-0.100**
	CMO Poverty	0.060**	0.050**
	Full Effect (CMO compared to TPS)	-0.037**	-0.050**
ELL	ELL	-0.146**	-0.237**
	CMO ELL	0.068**	0.079**
	Full Effect (CMO compared to TPS)	-0.078**	-0.158**
SPED	SPED	-0.237**	-0.262**
	CMO SPED	0.025**	0.038**
	Full Effect (CMO compared to TPS)	-0.212**	-0.224**

** p<0.01, * p<0.05

Table 20 summarizes the results of the poverty by race analyses. The values shown are relative to white non-poverty students attending a TPS school.

Table 20: Differences in Growth by Ethnic/Poverty Subpopulation and CMO Charter Attendance

		Math	Reading
Black	Black Poverty	0.047**	0.029**
	Black Non-Poverty	0.010**	0.001
Hispanic	Hispanic Poverty	0.050**	0.035**
	Hispanic Non-Poverty	-0.014**	-0.008**

** p<0.01, * p<0.05

Table 21: Difference in Growth by Exceptional needs Category and Charter Attendance, EMO

		Math	Reading
Poverty	Poverty	-0.097**	-0.099**
	EMO Poverty	0.016**	0.011**
	Full Effect (EMO compared to TPS)	-0.081**	-0.088**
ELL	ELL	-0.143**	-0.237**
	EMO ELL	0.058**	0.070**
	Full Effect (EMO compared to TPS)	-0.087**	-0.167**
SPED	SPED	-0.236**	-0.261**
	EMO SPED	0.010*	0.004
	Full Effect (EMO compared to TPS)	-0.226**	-0.257**

Table 21 shows that while the EMO coefficients for categorical subpopulation students were generally smaller than those for categorical subpopulation students in CMO charter schools, the coefficients for ELL students were medium sized in both math and reading.

Table 22 below shows the full effect for EMO charter and non-EMO charter students relative to white TPS students.

Table 22: Differences in Growth by Ethnic Subpopulation and Charter Attendance, EMO

		Math	Reading
White	EMO Charter	-0.038**	-0.014**
	Non-EMO Charter	-0.058**	-0.018**
Black	TPS	-0.175**	-0.153**
	EMO Charter	-0.139**	-0.134**
	Non-EMO Charter	-0.161**	-0.135**
Hispanic	TPS	-0.068**	-0.061**
	EMO Charter	-0.058**	-0.042**
	Non-EMO Charter	-0.084**	-0.072**

Table 23 shows the regression results for the number of schools per network. For this analysis, we regressed student growth in math and reading on a continuous variable of the number of schools in a network in a given year. The results of these analyses were significant, but as can be seen in Table 23 the values were too small to be meaningful. Each additional school within a network led to a -0.0003 standard deviation decrease in growth in math and -0.0001in reading. We conducted a specification check of this model by also running a model which used dummy variables of the number of schools per network rather than the continuous variable and a model with bins grouped by number of schools (3-10, 11-20, 21-30, etc). The results of these alternative specifications showed no discernible pattern which would indicate a correlation between number of schools in a network and network quality.

Table 23: Regression Results: Number of Schools per Network per Year

Variable	Math	Reading
z_origin	-0.318** (0.001)	-0.337** (0.001)
charter	0.014** (0.001)	0.013** (0.002)
Number of schools per network per year	-0.0003** (0.00003)	-0.0001** (0.00003)
ch_cmo	-0.011** (0.002)	-0.001 (0.002)
Constant	0.089** (0.002)	0.083** (0.002)
Observations	1,291,382	1,359,438
R-squared	.181	.226

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

We also conducted preliminary analyses of the geographic size of the charter organization. These analyses included continuous variables for the dispersion of schools within a network. These analyses found no evidence of a significant relationship between the geographic distribution of schools within the charter organization and student growth.

Another analysis which was not carried past the initial stages was an investigation into the relationship between network performance and variance. We created scatter plots which had the network coefficients from Appendix A on the horizontal axis and the network's variance on the vertical axis. These scatter plots (Figures 9 & 10) show no indications of the existence of a systematic relationship between network performance and variance around the network performance value.

Figure 9: Scatter Plot of Variance by Network Coefficient, Reading

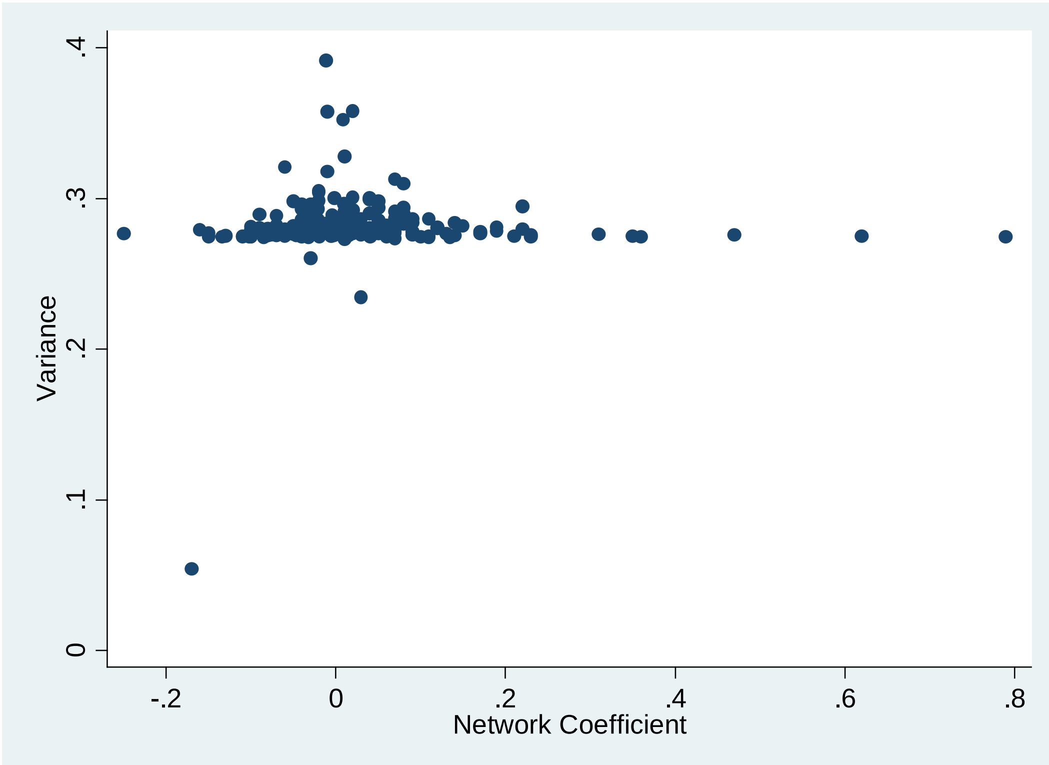


Figure 10: Scatter Plot of Variance by Network Coefficient, Math

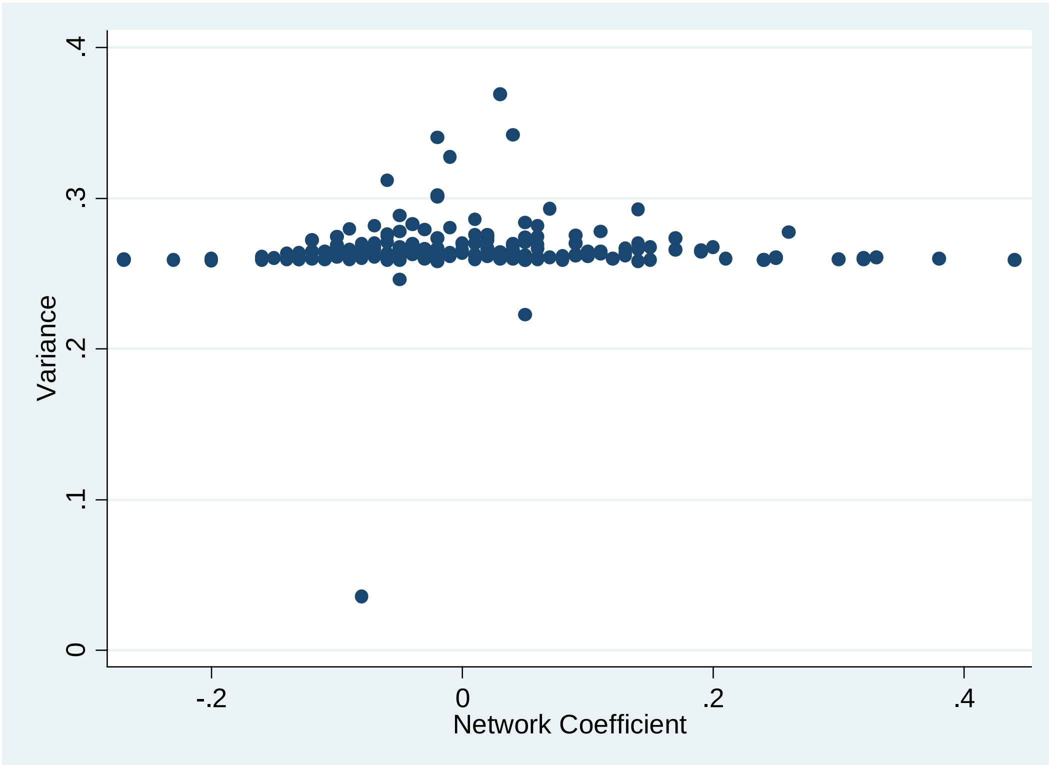


Table 12 shows the results of our analysis of multistate charter organizations. This analysis included a dummy variable to indicate if a charter school belonged to a charter organization with schools in multiple states. We found that this value was significantly negative for all charter organizations combined and for CMO networks. The coefficient was however, positive and significant for EMO charter schools. These results were consistent across both subjects.

Table 24: Regression Results: Multistate Networks, Math

Variable	Charters	CMO	EMO
z_origin	-0.304** (0.0004)	-0.304** (0.0004)	-0.304** (0.0004)
ch_multistate	-0.028** (0.001)		
ch_net_multistate		-0.119** (0.002)	0.040** (0.001)
ch_net_nonmultistate		0.041** (0.001)	0.004* (0.002)
charter	-0.007** (0.0006)	-0.013** (0.0004)	-.0141** (0.0006)
Constant	0.111** (0.004)	0.118** (0.004)	0.115** (0.004)
Observations	3,603,372	3,603,372	3,603,372
R-squared	.180	.182	.180

** p<0.01, * p<0.05

Table 25: Regression Results: Multistate Networks, Read

Variable	Charters	CMO	EMO
z_origin	-0.304** (0.0004)	-0.304** (0.0004)	-0.304** (0.0004)
ch_multistate	-0.016** (0.001)		
ch_net multistate		-0.059** (0.002)	0.017** (0.001)
ch_net_nonmultistate		0.013** (0.001)	0.006** (0.002)
charter	0.008** (0.0006)	0.006** (0.0006)	0.005** (0.0006)
Constant	0.145** (0.004)	0.118** (0.004)	0.147** (0.004)
Observations	3,795,768	3,603,372	3,603,372
R-squared	.189	.182	.189

** p<0.01, * p<0.05

We are also including results from the analysis of the impact of being part of a Charter School Growth Fund (CSGF) partner network. For this question, we ran two models for each subject. The first model was the marginal model. In this model we included separate dummy variables for being a charter school, for being a CMO network school, and for being a CSGF partner network. This allows for the evaluation of each of these levels individually. The full model included only interaction dummy variables for being a charter CMO CSGF school or being a charter CMO school which was not in CSGF. As can be seen in Table 26, adding the marginal coefficients together produces the same value as the interaction dummy variable for being a CSGF school.

Table 26: Regression Results: Charter School Growth Fund, Math and Reading

	Math		Reading	
Variable	**Marginal**	**Full**	**Marginal**	**Full**
z_origin	-0.306** (0.0004)	-0.306** (0.0004)	-0.296** (0.0004)	-0.296** (0.0004)
CSGF	0.241** (0.002)		0.139** (0.002)	
charter	-0.013** (0.001)		0.006** (0.001)	
ch_cmo	-0.022** (0.001)		-0.018** (0.001)	
ch_cmo_CSGF		0.206** (0.002)		0.128** (0.002)
ch_cmo_nonCSGF		-0.035** (0.001)		-0.011** (0.001)
Constant	0.115** (0.004)	0.115** (0.004)	0.146** (0.004)	0.146** (0.004)
Observations	3,603,372	3,603,372	3,795,768	3,795,768
R-squared	.183	.183	.190	.190

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

Limitations. While we feel confident in presenting our results, all studies have limitations. In these analyses, the national composite values can be considered average measures for one particular time span only. They do not provide the reader information about the student growth over the full range of a charter school's existence. Further, our data set while extensive does not contain the universe of charter schools across the country but is limited to our 23 states, New York City, and Washington DC. Charter schools operating outside of these areas will not be included. This means some of the network analyses may include only a portion of the schools in a multistate network.

Another limitation of any analyses based on state test scores is the difference in absolute achievement and growth from state to state. A one standard deviation changer in growth from one state will not necessarily equate exactly to a one standard deviation change in another state. Additionally, measures in this study are of growth not achievement. The reader is cautioned to not assume a state with larger growth has higher achievement. Absolute achievement is a different measure than growth.

Abadie, A., A. Diamond, et al. (2007). "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." SSRN eLibrary.

Davis, D. and M. Raymond (2012). "Choices for Studying Choice: Assessing Charter School Effectiveness Using Two Quasi-Experimental Methods." Economics of Education Review **31**(2): 225-236.