



REVIEW OF *CHARTER SCHOOL PERFORMANCE IN MICHIGAN*

Reviewed By

Andrew Maul

University of Colorado Boulder

February 2013

Summary of Review

The Center for Research on Education Outcomes (CREDO) at Stanford University conducted an analysis of the differences in student performance at charter schools and traditional public schools in the state of Michigan. In contrast to the majority of prior evidence regarding charter effects in the U.S., which tends to show no impact, the study finds an overall small positive effect of being in a charter school. As with CREDO's previous reports on charter schools, the study employs a large and comprehensive dataset and fairly solid analytic methods, making this study a contribution to the literature on charter school effectiveness. However, there are significant reasons for caution in interpreting the study's results. In particular, the plausibility of any causal inferences depends on the premise that the seven matching variables are sufficient to account for all meaningful differences between students in charter and traditional public schools. Additionally, important details of the analytic methods are missing from the report, making it unclear whether the researchers have addressed previously identified issues with their approach concerning, e.g., the stability of the student population used as a comparison group. Finally, even setting aside issues with the study's methods, the actual magnitudes of the effects reported are extremely small.

Kevin Welner

Project Director

William Mathis

Managing Director

Erik Gunn

Managing Editor

National Education Policy Center

School of Education, University of Colorado

Boulder, CO 80309-0249

Telephone: (802) 383-0058

Email: NEPC@colorado.edu

<http://nepc.colorado.edu>

Publishing Director: Alex Molnar



This is one of a series of Think Twice think tank reviews made possible in part by funding from the Great Lakes Center for Education Research and Practice. It is also available at <http://greatlakescenter.org>.

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.

REVIEW OF *CHARTER SCHOOL PERFORMANCE IN MICHIGAN*

Andrew Maul, University of Colorado Boulder

I. Introduction

Over the past several years, the Center for Research on Education Outcomes (CREDO) at Stanford University has produced a series of reports on the performance of charter schools relative to traditional public schools. These reports seek to inform ongoing conversations amongst policymakers and researchers regarding whether charter schools are likely to generate better outcomes than traditional public schools overall, and whether these effects might be especially pronounced for members of particular subgroups—particularly, students from minority racial/ethnic backgrounds and students from less socioeconomically advantaged backgrounds.

CREDO's latest study, *Charter School Performance in Michigan*,¹ employs a methodological approach highly similar to that employed in the center's previous reports. In contrast to the majority of evidence on the topic nationally, including earlier work from CREDO itself, this study finds that charter schools in Michigan perform better on average than traditional public schools in terms of raising student test scores.

II. Findings and Conclusions of the Report

Based on data collected between 2005 and 2011 and furnished by the Michigan Department of Education, the report presents a large number of conclusions, including the following 13:

- On average, it was estimated that students in charter schools in Michigan experience more academic growth than students in traditional public schools. Overall, this difference was estimated to be 0.06 standard deviations for both reading and math. This is equivalent to saying that about a tenth of one percent of the variation in academic growth is associated with school type.
- Estimated differences in growth were greater on average for students in the city of Detroit compared to other areas of Michigan. However, it was also reported that estimated differences were greater on average for students in rural schools compared to schools in urban or suburban neighborhoods and towns.

- The apparent advantage of charter schools was estimated to be greater on average for those that did not belong to a charter management organization (CMO), and those that did belong to an Education Management Organization (EMO).
- Estimated differences were greatest for elementary schools, and were close to nonexistent for multi-level schools (i.e., those that combine middle school with elementary, high, or both).
- The advantage of being enrolled in a charter school appeared to increase as a function of the number of years a student was enrolled in a charter school.
- On average, the gap in academic growth between White students and Black and Hispanic students was estimated to be smaller in charter schools than in traditional public schools.
- Similarly, on average, the gap in growth between students in poverty (defined as being eligible for subsidized school meals) and those not in poverty was estimated to be smaller in charter schools.
- Similarly, on average, the gap in growth between students who are held back a year and those who were not was estimated to be smaller in charter schools.
- Conversely, the gap in growth between students in special education programs and other students was greater in charter schools.
- The gap in growth between English Language Learners and other students was greater in charter schools on reading tests, and equivalent on math tests.
- The apparent overall advantage of charter schools was present regardless of students' estimated achievement at baseline. There was no evidence of a larger or smaller impact of being in a charter school for students of exceptionally low or high achievement.
- When considering school averages rather than individual students, 35% of charter schools in Michigan were estimated to have greater average growth than comparable public schools in reading, and 42% were estimated to have greater average growth in math. Two percent of charter schools were estimated to have lower average growth in reading, and 6% were estimated to have lower average growth in math.
- 65.8% of charter schools were estimated to have lower absolute achievement than the state average, possibly due to overall differences in populations served. (In Michigan, the proportions of students in poverty and of ethnic minority background are significantly higher in charter schools than in traditional public schools.) However, the report concludes that, given that growth rates are higher on average in charter schools, "if [charter] schools continue their trends of positive academic growth, their achievement would be expected to rise over time" (p. 36).

III. The Report's Rationale for Its Findings and Conclusions

Like previous CREDO reports of this nature, this is an empirical report. The conclusions are based primarily on analysis of a dataset furnished by the Michigan Department of Education, which is stated to include 85,650 charter school students from 273 charter schools. (It is stated that there are a total of 110,904 students in 297 charter schools in

The use of a large dataset represents a major strength of this study. However, this review has noted a number of reasons for concern regarding the methodology.

Michigan.) It also apparently includes a large number of students in traditional public schools (of which it is stated that there are 1,507,621 total, of whom 564,251 are in “feeder” schools from which charter students transfer), though it is not stated exactly how many of these were included in the dataset.

Full details on the methods are not included in this report, and there is no accompanying technical report. However, given that the methods employed in the present study appear to be highly similar to an earlier major report by CREDO² (henceforth referred to as “the 2009 study”), we can look to the methodology described in the technical appendix accompanying that earlier study.

The analytic conclusions are based solely on the findings from the Michigan data set. No conclusions about implications for policy are explicitly stated in the report. In contrast, the press release for the Michigan CREDO report states, “these findings show that Michigan has set policies and practices for charter schools and their authorizers to produce consistent high quality across the state.”

IV. The Report's Use of Research Literature

As with previous state-level CREDO reports on charter school data, the contents of the report focus on new findings. The report does not contain a literature review and contains minimal reference to other evidence. It seems reasonable to assume that the report was written with the intent of adding to CREDO's list of state-level CREDO reports on charter school data and is thus not meant to be taken in isolation.

V. Review of the Report's Methods

Miron and Applegate³ made a variety of comments regarding technical issues about the methods employed in the 2009 CREDO study. It does not appear that CREDO has altered

their methodology in light of those comments. Here I address three concerns with the present study, which overlap somewhat, and differ somewhat, with the comments made by Miron and Applegate.

The Virtual Control Record (VCR) Approach

When a causal inference is desired in the absence of a controlled experimental design, it must be argued that observational data can be used to provide an estimate of the counterfactual (i.e., what would have happened to these students had they attended a traditional public school rather than a charter school). CREDO's argument depends on the construction of a "Virtual Control Record" for each student in a charter school, which is obtained by averaging together up to seven students in "feeder" public schools (i.e., those schools whose students transfer to charters) with the same gender, ethnicity, English proficiency status, eligibility for subsidized meals, special education status, and grade level, and a similar score from a prior year's standardized test (within a tenth of a standard deviation) as the specified charter student. Thus, for each charter student, the method attempts to construct a "virtual twin" by averaging together scores of up to seven similar students enrolled in "feeder" schools.

Given the presence of a continuous covariate (prior year test score), it remains unclear why scores were not matched using a propensity method.⁴ It is reported that virtual matches were constructed for 86% of the tested charter school students, but no information is reported regarding the success of the matching procedure in eliminating selection bias (as measured by the matching variables).

The larger issue with the use of any matching-based technique is that it depends on the premise that the matching variables are sufficient to account for all relevant differences between students; that is, once students are matched on the aforementioned seven variables,⁵ there remain no meaningful unobserved differences between students in charter and traditional public schools (other than their school type). School-choice systems always implicate unobservable differences among parents (i.e., parents of charter school students are necessarily sufficiently engaged with their children's education to actively select a charter school). To the extent to which a reader finds it implausible that the seven variables have captured these and all other important differences, she will be unconvinced that these methods can provide true estimates of causal effects.

The Multilevel Structure of the Data

The data analyzed are hierarchical in the sense that students are nested within classrooms and schools, and it is likely that observations are not conditionally independent—that there is considerable within-school shared variance. As Miron and Applegate also commented, multilevel modeling seems like it would have been the natural choice for such a data structure. This seems especially relevant given that the present report describes results at both the student level and the school level; it analyzes data separately at the student and school levels. Whether the student-level analyses appropriately adjusted for

the hierarchical data structure is not clear. In considering this issue, it may be inferred that the methods are identical to those used in the 2009 study, in which case standard regression with robust standard errors was applied to the student-level data. This approach may be sufficient, but the report is missing details regarding the implementation of this procedure that could have been used to evaluate its appropriateness.

More importantly, at the school level, it appears that student data were simply aggregated up to schools, and significance tests were performed on the resulting school mean scores. This approach implicitly treats the school means as fixed, rather than randomly sampled from a theoretical population of schools, which would seem to be the preferred interpretation (given that, as displayed in the press release, CREDO's desire is to make an inference not only to the *actual* charter schools studied, but to charter schools *in general*). This could easily be handled via multilevel modeling by the inclusion of a random intercept for schools (interacting with a variable for school type). The present report contains no statements about differences between overall mean growth in charter schools and traditional public schools (such statements are made only at the student level); instead, statements about schools are made in terms of the percentages of schools of different types with above- and below-average levels of growth. In and of itself, such statements can be made on the basis of fixed effects; however, again given that it seems that the target of inference is the difference between charter schools and traditional public schools *in general*, it would be more appropriate to estimate differences in mean growth using a model with random effects. Thus it seems there is some degree of disconnect regarding the connection between the analyses and the desired inferences.

Additionally, it could be argued that the mean growth in each school as reported in the study should actually be treated as an *estimate* of the true (i.e., long-run) mean growth in that school, rather than as a fixed value. If understood this way, the true mean growth of the school is fallibly measured by the growth of the students actually in the school at the time of testing. Given the large number of students within each school, the extent to which estimates are biased by not modeling this kind of measurement error is probably minimal. However, there is a related concern that could be more serious. The approach used in the study demands the exclusion of schools with a lower number of tested students. The researchers report that 23 percent (61 out of 269) of the state's charter schools were excluded from analyses on the basis of having an insufficient number of tested students to "provide a fair test" of the school impact (p. 32).⁶ This exclusion would not be necessary if error in the measurement of school means were explicitly modeled (as would occur in the aforementioned random-intercept model). This could be problematic if these 61 excluded schools were systematically different from others, which seems plausible given that it is implied that they are small and new (p. 32).

The Estimation of Growth

As with previous reports, findings are described in terms of "growth," estimated via average year-to-year gains on state standardized tests expressed in standard deviation units. As Miron and Applegate noted, making inferences to longitudinal growth of

individual students' levels of achievement leans on certain (unstated) assumptions, most notably that the group of students used to construct the Virtual Control Records is itself stable. Given that the researchers had access to individual student records, changes at the level of the individual could have been modeled directly using a multilevel framework.

VI. Review of the Validity of the Findings and Conclusions

As with previous CREDO reports on charter schools, the use of a large dataset obtained from the state Department of Education represents a major strength of this study. However, this review has noted a number of reasons for concern regarding the methodology.

It seems clear that inferences are desired both about the population of students in Michigan schools and about schools in general. Given the availability of models directly suited to simultaneously address research questions at multiple levels of analysis, it is unclear why CREDO seems to have decided to stick with the analysis of group-averaged gain scores using standard regression.

Even given CREDO's analytic framework, the actual magnitude of the differences discussed are generally quite small (e.g., 0.06 standard deviations). The very large sample size guarantees that nearly any predictor will be statistically significant; however, in practical terms, many of the differences may be regarded as trivial. Even the largest effect sizes reported (e.g., the estimated effect of charter schools in urban Detroit, as set forth on page 14 of the report) are less than a tenth of a standard deviation, and many (such as the estimated advantage of charter schools for students in poverty, p. 25) are on the order of 0.02 or 0.03 standard deviations. An effect size of a tenth of a standard deviation would mean that one quarter of one percent of all the variation in student growth could be explained by whether the student was in a charter or traditional public school. As another point of reference, Eric Hanushek has described an effect size of 0.20 standard deviations for Tennessee's class size reform as "relatively small" compared to the nature of the intervention.⁷

The report occasionally expresses effects in terms of "months of learning," but admits that such interpretations are imprecise (p. 15).⁸

Finally, and perhaps most importantly, the conclusions of the study require acceptance of the premise that the seven matching variables employed to create Virtual Control Records are sufficient to estimate a true counterfactual. Thus, these variables must be sufficient to capture all other sources of heterogeneity between charter and traditional public school students. Especially given the small effect sizes reported in this study, even small unmodeled differences could play a decisive role. At a minimum, it seems that the claim that a true counterfactual is estimated here should be carefully examined.

VII. Usefulness of the Report for Guidance of Policy and Practice

The size and comprehensiveness of the dataset analyzed, and the fact that a positive effect of charter schools was found even using the same methodology that has usually found no difference between charter and traditional public schools, make this report an interesting contribution to the charter school research base. At minimum, it suggests that it may be worth attempting to more deeply understand the differences between charter policies and conditions in Michigan and other states. However, this review has noted reasons for caution when making inferences to a true causal effect of charter schools. As such, it is advised that the findings of this report not be regarded as definitive evidence in favor of the effectiveness of Michigan's charter schools or of that state's charter-related policies.

Notes and References

1 Center for Research on Education Outcomes (CREDO) (2013, January). *Charter School Performance in Michigan*. Palo Alto: CREDO, Stanford University. Retrieved February 10, 2013, from http://credo.stanford.edu/pdfs/MI_report_2012_FINAL_1_11_2013_no_watermark.pdf.

2 Center for Research on Education Outcomes (CREDO) (2009, June). *Multiple Choice: Charter School Performance in 16 States*. Palo Alto: CREDO, Stanford University.

3 Miron, G. & Applegate, B. (2009). *Review of "Multiple choice: Charter school performance in 16 states."* Boulder, CO: National Education Policy Center. Retrieved February 10, 2013, from <http://nepc.colorado.edu/thinktank/review-multiple-choice>.

4 Propensity-based score matching is an increasingly common way of attempting to reduce bias in the estimation of causal effects from observational data due to the confounding influence of variables that predict whether or not a student receives a treatment. Such techniques predict the probability of treatment based on a set of conditioning variables, which can be either continuous or categorical, and then match subjects in the two groups based on similarity in this probability; thus exact matches are not required, and therefore neither are arbitrary cutoffs for continuous covariates.

5 Studies using propensity-based methods frequently use very large numbers (e.g., 70 or greater) of variables to match students, and even then there is debate concerning whether the matches can be thought of as true counterfactuals.

6 It is stated that the criteria were that the school have at least 60 matched charter students over a two-year period, or at least 30 matched students in the case of charter schools that have only one year of data. The justification for these particular cutoffs is not explained.

7 Hanushek, E. A. (2002). Evidence, politics, and the class size debate. In L. Mishel & R. Rothstein (eds.), *The Class Size Debate*. Washington, DC: Economic Policy Institute, 37-65; 56. Retrieved February 10, 2013, from http://borrellid64.com/images/The_Class_Size_Debate.pdf.

8 The synthesis of the report states that "...a majority of charter schools have academic growth that is above the average for all public schools in Michigan... Should these trends continue, the share of schools which currently lag the state average for absolute achievement would be expected to decline" (p. 45). This same statement is contained in the press release for the report, but it's a curious statement. First, it is not clear what "trends" are being referred to here, as it has not been argued that growth rates have increased over time in charter schools (equivocal evidence on p. 16 of the report notwithstanding). Moreover, even if students in charter schools continue to experience higher gains than those in public schools, it is not clear that this would ever translate into a higher level of absolute achievement for charter schools, given student turnover and the aforementioned differences in populations served by charters and traditional public schools.

DOCUMENT REVIEWED:

Charter School Performance in Michigan

AUTHOR:

Center for Research on Education Outcomes (CREDO)

PUBLISHER:

CREDO

DOCUMENT RELEASE DATE:

January 11, 2013

REVIEW DATE:

February 12, 2013

REVIEWER:

Andrew Maul, University of Colorado Boulder

E-MAIL ADDRESS:

andrew.maul@colorado.edu

PHONE NUMBER:

(303) 492-7653

SUGGESTED CITATION:

Maul, A. (2013). *Review of "Charter School Performance in Michigan."* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-charter-performance-michigan>.