



Recently Gary Miron and Christopher Shank released a review of CREDO's Charter Management Organizations for the Nation Education Policy Center. While we appreciate neutral, professional critiques of our work, and agree with some of the comments in Miron and Shank's review, several claims are skewed in their presentation or simply incorrect.

### **Insufficient Literature Review**

One of the touted criticisms of the Charter Management Organizations study (referred to from here as CMO study) was "not taking into account of acknowledging the large body of charter school research beyond CREDO's own work"(Miron and Shank 2017). Miron and Shank are correct in that CREDO did not include a literature review in the report. This is not because we do not take into account the work of other researchers or because we do not wish to acknowledge the long history of charter school research. There is a significant body of work which documents the history of charter school research readily available to readers in journals and accessible via the internet. CREDO does draw and build on that work. But spending 10 to 15 percent of a report going over out-of-date findings which no longer represent the performance of the charter sector does not seem to be a good use of the readers' time in a report intended to give policy makers a view of the current status of the charter sector.

For example, Miron and Shank specifically note the absence of a study by Mathematica(Gleason 2010) which Miron and Shank describe as "the most expensive and rigorous study to date". We did not reference the Mathematica study because the data used in that study was from the 2004-2005 and 2005-2006 school years, thus the results in that study do not reflect the current state of the charter sector. The CMO study was intended to evaluate the current state of the charter sector; therefore, we did not see the relevance of discussing results from 13-year-old data. Even our references to CREDO's previous studies were merely to make the point that the performance of the charter sector has changed greatly over time.

### **Methods Criticisms**

In criticizing CREDO's methods, Miron and Shank cite a number of reviews all issued by the National Education Policy Center (NEPC) the same organization which published this current critique. CREDO has addressed the comments within these previous reviews on our website. Links to our responses are available here: <http://credo.stanford.edu/research-reports.html>.

Miron and Shank state "CREDO's Virtual Control Record (VCR) approach to quasi-experimental control deviates from widely-accepted propensity matching methods and is not a legitimate substitute for true random assignment to experimental

conditions.” This statement is false. CREDO researchers (Davis and Raymond 2012) completed a peer-reviewed evaluation of the VCR method. Davis and Raymond found the analyses using the VCR method produced results which were similar in significance, direction, and magnitude to a fixed effects model using the same data. Fixed effects models are a well-accepted and commonly used quasi-experimental method for analyzing the impact of charter school programs.

Additionally, a 2012 National Center for Education Evaluation and Regional Assistance (Fortson, Verbitsky-Savitz et al. 2012) report tested an exact matching process based on CREDO’s VCR approach against an experimental model, the widely-accepted propensity matching model, simple OLS regression model, and a fixed-effects model. The comparison study found CREDO’s VCR method produced results which were not significantly different from the experimental model and that the VCR method arguably outperformed the standard propensity score method.

Finally, a working paper from the Harvard Kennedy School Program on Education Policy and Governance (Ackerman and Egalite 2015) conducted an independent evaluation of CREDO’s VCR method. Ackerman and Egalite found results from a VCR model to be unbiased and consistent with findings from the comparison instrumental variables (IV) model. Ackerman and Egalite go on to point out that a significant number of charter schools do not have waiting lists; thus, these schools cannot be evaluated using randomized control trials. Ackerman and Egalite state that matching models such as CREDO’s VCR method can be an useful tool for measuring performance in this large sector of charter schools.

Miron and Shank also state CREDO’s outcome variable is flawed because we are comparing improvement on one state’s test to that on another state’s test. This could be a valid critique if CREDO were using raw or scale scores to measure growth or were comparing the growth of students in one state to that of students in another. However, that is not the case. The CMO study technical appendix on page 93 defines the growth measure as “ $\Delta A_{i,t} = A_{i,t} - A_{i,t-1}$  where  $A_{it}$  is the state-by-test z-score for student  $i$  in period  $t$ ;  $A_{it-1}$  is the state-by-test z-score for student  $i$  in period  $t - 1$ ”. Because we are measuring the change in z-scores, we are measuring the change in each student’s position relative to their in-state peers in year  $t$  to the position of the student relative to their in-state peers in the previous year. This comparison has nothing to do with the relative level of the average student in one state as compared



to the average student in another state. It is simply saying that on average, students who attend a charter school move  $x$  amount relative to their peers each year.

Another criticism of Miron and Shank is that CREDO should not generalize our findings to the theoretical population of all charter schools. Unlike randomized control trial studies which generally include only a few schools or even a single school, CREDO's data for the CMO study included states which enroll 91 percent of all charter students in the United States. The strong study match rates and large size of CREDO's sample relative to the entire population make the results of this study highly generalizable to the entire population of charter students.

Miron and Shank critique the choice of categorical breakouts of schools by operating practices. They complain that the categories are not "conventional in either research or practice." Miron and Shank are correct in stating these categories are not conventional in research. CREDO's stance is that they should be. As the CMO study shows, there are differences in the performance of charter school by management structure. Exploring this possibility was the purpose of the study. Likewise, Miron and Shank criticized that CREDO did not spend enough time focusing on the question of for-profit vs. non-profit operators. The question of the performance of for-profit vs. non-profit operators was not the primary focus of this study. We agree questions about the performance of for-profit vs non-profit operators still hold much interest for research, so much so that that question should constitute its own study.

### **Trivial Effect Sizes**

Miron and Shank criticize that most of the effect sizes found in the CMO study do not merit attention. They then cite Cohen's example benchmarks of 0.2, 0.5, and 0.8 for small, medium, and large effect sizes respectively. However, right after introducing these thresholds, Miron and Shank then point out that Cohen himself states those thresholds were only examples and are not appropriate for all areas of study. In fact, the Institute of Education Sciences (IES) report (Lipsey 2012) cited by Miron and Shank states, "Cohen's broad categories of small, medium, and large are clearly not tailored to the effects of intervention studies in education." Rather, the IES report points to an example effect size of 0.15 on reading achievement scores as "an effect quite likely to be judged to have practical significance in an elementary school context." Further, the report goes on to point out that a similar effect at the high school level would be huge. Since it is reasonable to presume Miron and Shank read the entirety of the report they

cited, it is left to conclude they are fully aware of the misleading nature of their claim that the effect sizes found by CREDO are trivial simply because those sizes fall below Cohen’s suggested “small” effect size.

It is correct to say CREDO’s large sample size ensures we have the ability to estimate very small effects. This is why we list all the effects and allow readers to decide if they feel the effects are worth the effort, expense, and disruption needed to achieve them. For some populations, the answer should be obvious. For others, each policy maker should read the report and make that decision based on the needs of their particular jurisdiction.

### **Days of Learning**

Miron and Shank continue to quote the above IES report’s suggestion effect sizes be described “relative to normal student academic growth.” One method of aligning effect sizes relative to normal student academic growth mentioned in the IES report is to use a nationally normed standardized exam to produce a baseline estimate of normal student growth. This is exactly what the conversation to days of learning attempts to do. The days of learning conversion uses the estimate of growth on the NAEP, a national representative exam, to establish a baseline growth rate for the typical American student. As the majority of the data in the CREDO study comes from students in grades 4 through 8, using the growth of students on NAEP from grades 4 to 8 is a reasonable choice. That being said, CREDO is continuing to examine other examples of nationally normed and representative exams to improve the days of learning conversion.

One of the concerns brought up by Miron and Shank was about differential rates of growth by grade and state. We too recognized these could be an issue. To address concerns about differential growth by grade and state, the regression models CREDO applies include controls for student grade and state. This was included in the model descriptions in the technical appendix.

On page 12 of the CMO study, CREDO states, “We wish to emphasize that the days of learning translation is only **meant to be a loose approximation** of the effect size to provide a sense of scale to aid the reader in interpreting the effect sizes. The effect sizes are the mathematically computed measures produced by the statistical models and should be the basis for policy decisions.” [emphasis added] Additionally



throughout the report, we use words such as “approximately”, “about”, “around”, etc. to remind the reader of the approximate nature of the days of learning.

### **Statements about the Usefulness of CREDO’s Report**

Miron and Shank state “CREDO approached this work as if in an academic vacuum.” That is entirely untrue. We approach our work with an expectation that policy makers can and do read the work of researchers other than ourselves. CREDO has never stated our work was “a definitive study” nor that other works should be ignored. We approach our work with the idea that information is readily available and readers are intelligent consumers of information. Thus we do not need to spend a great deal of time pointing out other studies to the readers.

Miron’s and Shank’s direct accusation of academic dishonesty on the part of CREDO is not only untrue, but it is also unprofessional and offensive. It is true CREDO receives funding from a variety of organizations some of which have a specific idea about the direction of public education. However, it is always stipulated in the contracts between Stanford University and the funders that CREDO will retain full control over the content of our work. CREDO has never based the findings of any study on anything other than the evidence from the data. Throughout the years, CREDO has consistently advised states and authorizers to hold schools accountable for student achievement gains and close down failing schools. In fact, in 2015 CREDO release a highly damning report about the poor performance of online charter schools.

### **References**

Ackerman, M. and A. J. Egalite (2015). assessing Tradeoffs between Observational and Experimental Designs for Charter School Research. Program on Education Policy and Governance Working Papers Series, Harvard Kennedy School.

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.

Fortson, K., N. Verbitsky-Savitz, et al. (2012). Using an experimental evaluation of charter schools to test whether nonexperimental comparison group methods can replicate experimental impact estimates, NCEE 2012-4019 U.S. Department of Education.



Gleason, P. C., Clark, M., Clark Tuttle, C. and E. Dwyer. (2010). The Evaluation of Charter School Impacts: Final Report. Washington, DC, Institute for Education Sciences.

Lipsey, M., Puzio, K., Yun, C., Hebert, M., Steinda-Fry, K., Cole, M., Roberts, M., Anthony, K., and Busick, M. (2012). Translating the Statistical Representation of the Effects of Education Interventions Into More Readily Interpretable Forms. I. f. E. Sciences. Washington DC, National Center for Education Statistics.

Miron, G. and C. Shank (2017). NEPC Review: Charter Management Organizations 2017 (CREDO, June 2017). NEPC Review. W. Mathis. Boulder Colorado, University of Colorado, Boulder.