



June 2015

## CREDO Response to Maul and Gabor

In March 2015, CREDO released a study of the performance of charter schools in 41 urban regions. The Summary Report and individual regional reports can be found on the [credo.stanford.edu](http://credo.stanford.edu) website.

Two critiques of the study have been issued since the release of the study. We are grateful to Andrew Maul and Andrea Gabor for taking the time to read and review our work. CREDO always strives for the highest degree of empirical rigor and transparency in our work; reviews of all kinds are welcome. Our goal is to produce policy-neutral evidence about important issues under debate as an impartial contribution. For these reasons, we value the exchange and appreciate the opportunity to reflect and respond.

This memorandum provides the major criticisms from each reviewer and our responses. Criticisms are summarized in bold; our responses immediately follow. Because several of the issues were common across the two reviews, we have combined the two for efficiency's sake.

### **Matching method critiques**

**Method not well described.** We prefer to issue reports that are prepared for a non-technical audience. This is a choice that aims at providing clarity about the issues and our findings, but in no way is intended to obfuscate our approach. We prepared a Technical Appendix that explained the new methods that were developed for the Urban Charter School Study, referenced our prior technical documents about the matching methodology, and provided references to the peer-reviewed articles in several academic journals that test the VCR method against other approaches. The Summary Report explicitly describes the Technical Appendix and includes a hyperlink directly to it on the [urbancharters.stanford.edu](http://urbancharters.stanford.edu) website. We consider this approach to strike a good balance between keeping the report non-technical and being transparent about our work. Dr. Maul proves the approach is workable when he includes a perfectly adequate description of the VCR technique in his review. For this reason, the claim of opacity cannot be supported.

**Method fails to account for all differences between charter and traditional public school (TPS) students.** Maul raises the question of whether the matching approach completely accounts for all possible differences between charter school students and TPS students. This critique has already been addressed in earlier work [here](#) and [here](#) and need not be repeated here, except to once again defend traditional public school parents' capacities and willingness to choose the environment for their children they feel offers the right mix of instruction, social milieu, school culture, and location.

**Use of feeder schools as source for Virtual Control Records (VCRs) criticized.** This critique reveals a fundamental misconception about the research hypothesis. Dr. Maul and others may wish for a head-to-head sector-to-sector study, but that is not the study we intended to perform. Our hypothesis is a student-level question: how well are charter students progressing? Our VCR method is built in specific recognition that students who enroll in charter schools are not a mirror sample of students in their communities' traditional public schools. Our demographic comparisons, included in every report, lay out these differences in clear terms.

We claim only to mirror the students in charter schools with our VCRs. Including schools in the match pool that none of the charter students has never attended would advantage the counterfactual, since the excluded schools are generally higher performing than the schools charter students leave. We have never asserted that the VCRs are representative of the full range of traditional public school students because we know they are not. Again, that is not the question we aim to study as such raw comparisons add little value to the knowledge of public schools both traditional and charter.

We use feeder schools specifically because they are the most appropriate source of students to develop a counterfactual for testing our hypothesis. As stated in the Summary Report, urban charter schools are not evenly distributed across metropolitan geographies. Explanations for this phenomenon include facilities constraints, explicit operator missions to locate in urban locations that have not had access to high quality education options, and considerations of transportation and access. Thus, limiting the range of schools from which we draw students for the VCR to only those schools that lose students to charter schools gives us the ability to estimate what the charter students' education outcomes would have looked like had they remained in the very same schools they or their peers come from.

**Use of eligibility for Free and Reduced Price Lunch (FRL) criticized as an imperfect measure of socio-economic status.** With this critique, we agree to a point. For research on education equity, FRL eligibility remains the standard indicator of poverty status, but we agree it is a blunt instrument and not as sensitive a measure as one might desire. Efforts for over a decade to develop better measures of socio-economic status have not gained any traction in education policy circles. This leaves us with two alternatives. Assuming infinite resources and time, one might be able to launch new data collections over many states – but it is easy to see how administratively infeasible this option is for any serious and ongoing program of research that involves dozens of states and thousands of schools. The other option is to weigh the options of keeping the variable in its current state or excluding it from our model. A decision to exclude socio-economic status would lead to additional criticism for failing to control, however crudely, for economic inequality. For this reason, we include FRL eligibility in our model; it is the single most appropriate indicator of economic status at this time.

**Over 80-percent match rate criticized as insufficient.** Claim is that propensity matching would have resulted in higher match rates. The parameters of propensity matching have been explored in full against the VCR method by an uninterested research group and was found to be inferior. See [Gleason et al.](#)

**The unmatched students are .43 standard deviations (s.d.) below the average of matched students.** Dr. Maul misinterprets the implication of unmatched charter students. The charter students who were unmatched were excluded because there were no matches for them in the traditional public schools in the feeder pool.

We have examined the unmatched students and find that their matches fail not because of their prior academic performance (especially in communities with small fractions of all students enrolled in charter schools), but because they present unusual combinations of the remaining factors. Of these factors, charter school students repeat grades far more often than students in traditional public schools, and this causes many of the match disqualifications that arise.

**Days of Learning conversion not explained.** We utilize and cite the work of [Hanushek, Peterson and Woessman](#) for performing the transformation of standard deviation units of growth into days of learning. The calculations are not presented in scientific notation since that is not the style of report we prefer to release. However, they are described for the lay reader's understanding, and we direct expert readers to the primary source.

**Selection of urban regions not clearly explained.** We are unclear on this criticism. The Technical Appendix details the methods used to select the 41 regions under study.

### **Critique of modifications to VCR method**

**Method “violated” in study of New Orleans.** Gabor asserts that CREDO “violated their own methodology” in our study of New Orleans, which creates a tainted set of results. Any study of New Orleans student academic performance has to manage the unique contours of that education landscape. Specifically, over the past 10 years, the Recovery School District has closed or transitioned all the public schools in its portfolio. Currently, four schools are directly operated by the Orleans Parish School Board. The rest are charter schools overseen by the Orleans Parish School Board or the Louisiana Board of Elementary and Secondary Education. New Orleans is considered the nation’s first all-charter district.

Since the basic VCR method seeks to find students in schools that charter students transferred from or would otherwise have attended, the lack of traditional public schools in New Orleans requires a modification of the VCR approach. We developed a list of schools in other cities in Louisiana that were similar in student populations as the schools that historically had served as feeder schools. Those schools were then treated as proxies for the feeder schools for matching students to New Orleans charter school students and the protocol for VCR development proceeded from that point.

New Orleans is the only district where this adjustment was made. Failure to annotate the difference in the Technical Appendix was merely an oversight, which we have since corrected. We appreciate Gabor’s note to this effect, but reject the charge that we intentionally concealed “violations.”

CREDO does not consider the substitution of similar schools to be detrimental to creating sound VCRs. The methodology of VCR creation is easily adapted to fit a variety of circumstances, and we have used variations of the VCR to examine other research questions. We will continue to use the basic approach and adapt it as circumstances require.

In this specific instance, the counterfactual VCRs have the advantage since their systems remained largely intact after Hurricane Katrina and students were not as

drastically displaced. In addition, the other cities in Louisiana have less mature charter sectors with much smaller proportions of students attending them, rendering assertions of systematic migration of high-performing students to charter schools unsustainable. Nevertheless, the oversight has been noted, and we will endeavor to be clearer in the future.

### **Bonferroni Correction critique**

**No Bonferroni Correction applied to adjust for multiple comparisons.** The Bonferroni Correction and other adjustments for multiple comparisons are meant to account for the fact that if you make enough comparisons, at least one of them will be significant simply due to chance. Researchers select the threshold for chance before beginning their research. CREDO, like most researchers, use a threshold of 5-percent chance. This means if we make 20 comparisons in which none of the differences are actually real, we will likely still find one of the 20 comparisons is different. The Bonferroni Correction is meant to prevent this by shirking the size of the acceptable error for every comparison so the total chance of error over all of the comparisons remains at 5 percent. This correction does not apply to CREDO's studies because we are not saying there is an effect if ANY one comparison is significant. We instead report the number and degree of significant findings. If the findings in CREDO's urban study were due to this "Bonferroni error," there would be far fewer significant results that what were found. Even when we overcorrect by applying the Bonferroni Correction, the vast majority of significant results found in the urban study remain significant.

Furthermore, results found by random chance should themselves be random. While results from CREDO's various studies have changed some over the years as new years of data have been added and old years have been dropped, the results have changed in predictable ways. The differences have taken the form of trends that are gradual and consistent over time. If CREDO's findings were the results of a random "Bonferroni error," we would have findings popping in and out of our results, not changing gradually over time.

For a more technical discussion of why the Bonferroni Correction is not needed, see [CREDO's website](#).

### **Effect size critique**

**Reported effect sizes are small and not substantial.** Maul claims that the effect sizes reported in the study are small, explaining “well under a tenth of one percent of the variance in test scores.” First, the author is confused about the outcome of interest and misdirects the criticism. CREDO looks at academic growth, not academic status. The outcome of interest is the change in each student’s test score from one year to the next; it is not the test score itself. Second, multiple individuals both at CREDO and elsewhere have tried to recreate the statistic that Maul reports and have been unable to do so.