



School of Education, University of Colorado Boulder
Boulder, CO 80309-0249
Telephone: 802-383-0058

NEPC@colorado.edu
<http://nepc.colorado.edu>

Rejoinder to Response by Will Flanders and CJ Szafir to a Review of *Bang for the Buck*

*Casey Cobb, University of Connecticut
July 2016*

Original Report: <http://tinyurl.com/j7c6jxk>

Original Think Tank Review: <http://nepc.info/node/8128>

Flander's and Szafir's Reply: <http://tinyurl.com/zepqdtm>

Cobb's Rejoinder (this document): <http://nepc.info/node/8128>

A report titled, *Bang for the Buck: Which public schools in Milwaukee produce the best outcomes per dollar spent?*¹ was released by the Wisconsin Institute for Law & Liberty in May 2016. The report generates “efficiency scores” by dividing test scores by per pupil spending scores for each public school in Milwaukee and draws conclusions about the relative efficiency among charter and traditional public schools operating in that city. I conducted a [review of the report](#)² for the NEPC Think Twice review project on July 12, 2016. I concluded that the claims drawn from the report were based on the construction of efficiency scores that were of questionable value, and further, that the research design was simply inadequate to justify many of the report's conclusions and recommendations for policy.

The co-authors of *Bang for the Buck* responded to my review of their report, with the response published in the Wisconsin Institute for Law & Liberty's (WILL) [Blog](#)³ on July 14, 2016. As an academic, I appreciate that the authors took the time to respond to my review. Education policy should be guided by sound education research, and it is toward that end that I conducted the original review and now offer my [rejoinder](#)⁴ to their criticism of that review.

The co-authors of *Bang for the Buck*, Dr. Will Flanders, Education Research Director at WILL, and CJ Szafir, Vice President for Policy, critique the five summary claims listed in the abstract of my review. Below, I respond to their critiques of these five claims.

My review's claim, as cited by the WILL Blog on July 14:

1. "Test scores do not comprehensively represent the purposes of schools."

The report's authors actually agree with this statement, but reject the notion of including other important measures of school success. In the body of my review, I wrote:

A second problem lies on the output side of the efficiency equation, which relies on test scores in math and science in a single year, across only tested grades, to capture an entire school's performance. Doing so ignores many other valued outcome measures such as graduation rates, post-secondary rates, student extracurricular success, and student achievement in the arts, reading, writing, and social studies. Further, it is grossly oversimplistic to assume that a test score in a single year represents a school's unique contribution to student achievement in those subjects.

In their critique, the authors state that it is "very hard, if not impossible, to quantify 'extracurricular' activities" but do not speak to the possibility of capturing the other outcome measures I suggested (i.e., graduation rates, post-secondary rates, student achievement in other subject areas such as the arts, reading, writing, and social studies). While it can be a challenge to measure success in extracurricular activities, it doesn't mean it's not worth doing. At the very least, quantifying extracurricular offerings and student participation in those activities would speak to a broader role of schools. Schools that are smaller in size may not be able to offer extensive extracurricular options such as robotics, math team, jazz club, or any number of athletic teams; certainly this is worth noting when making overall claims about the performance of schools.

The larger point here is that the WILL report's use of a narrow set of test scores within a non-causal research design does not serve as an adequate basis to make definitive claims about the effectiveness of any school type. That is, whether or not one thinks the study should have included these other things, the narrow approach is a clear limitation, and this should have been explained to readers. Nor should it lead to such strong policy guidance as found in the report's conclusion, which reads, "...given the growing evidence of charter school effectiveness from this study and others, it is important for the state to increase the access of Wisconsin's children to these schools." This study was simply not designed, in its data or its methods, to reveal school effects.

Finally, the authors take a jab at my references to other studies that used test scores as a measure of school performance. They write, "A point that Dr. Cobb, ironically, makes by *citing research that uses test scores to measure achievement*." The authors fail to put this into its appropriate context, which is within the section of my review that critiques the report's use of the research literature (Section IV, p. 4). Here, I wrote,

The report is selective in its use of research literature on charter schools. For instance, it states "[m]ost existing research has found that public charter schools earn better outcomes than traditional public schools" (p.1) but does not offer any citations for the claim. Such a claim is misinformed given the contested research terrain comparing charter and traditional public school effectiveness. A heavy volume of studies reveals mixed findings or point to the relative ineffectiveness of charters when compared to traditional public schools [see endnote 7

of my review for a list of studies]. (p. 4)

Summarizing the prior research on a topic helps orient the reader to the present study and provides a basis for researchers to launch their own investigations. That's what social scientists do; they build upon previous research to situate their own studies, test theories, and advance the knowledge base. Providing a grossly incomplete or selective research summary introduces a bias that is unhelpful and unnecessary.

My review's claim, as cited by the WILL Blog on July 14:

2. "The report does not address threats to the validity of its assumption that there is uniform financial accounting across schools and types."

The report's authors assert that I provide "no actual evidence to support [this] claim that charter schools share services with Milwaukee Public Schools." I am puzzled by this account, for in my review I cited the 2014-15 Wisconsin Charter Schools Yearbook, which is published by the state's Department of Public Instruction (highlighted below).

First, per-pupil funding does not take into account how such funds were used or parsed for what purpose. In addition to funds that directly support instruction, a portion of funds can be also used for administration, counseling, transportation, or special services to students with disabilities. The per-pupil funding for non-instrumentality and independent charter schools may be lower than that of MPS and instrumentality schools, but the figure does not necessarily account for other costs absorbed by sponsoring districts or expended through other contractual agreements. **The 2014-15 Wisconsin Charter Schools Yearbook, published by the Wisconsin Department of Public Instruction, indicates that in those instances where a charter school functions with less money [than its sponsoring district], it "can happen if a charter school shares an existing district facility, and shares management costs with the school district, participates in district services such as co-curricular activities, special education, psychological services, and food service" (p.7).** [see endnote 11 of my review]. Given the likelihood that non-instrumentality charters appreciate some of these district economies of scale, it would be inappropriate to assume per-pupil funding averages as used in the report represent the complete account of public funds expended in those schools.

The statement highlighted above explains how charter schools with lower per pupil funding relative to its sponsoring district can be offset by district-shared costs. But the larger and more pressing point is that the WILL study did not attempt to determine how public funds were used and for what purpose. Given the statement provided above by the Wisconsin Department of Public Instruction and statements cited in the WILL Blog by "Sean Roberts of Milwaukee Charter School Advocates," there appears to be enough uncertainty about how funds are spent and on what—enough that it deserves some validation by a study focused on fiscal efficiency.

My review's claim, as cited by the WILL Blog on July 14:

3. "The analytic description of the study is incomplete, making interpretation difficult."

The main point of my claim was that the methods were insufficiently described. For instance, it was not evident in their second efficiency score analysis—specifically the first-step regression analysis—that any weighting was used to account for variation in school enrollments. I wrote in the review that weighted regression would be the preferred approach; however the description of methods did not mention any weighting, nor did their response in the blog.

Further, interpretation of the efficiency scores, particularly those generated by the second set of analyses, was difficult to put into practical context without further description. I described the two-stage efficiency score analysis in my review and concluded:

Here again, the efficiency scores generated by school type are somewhat challenging to interpret. It is not entirely clear what the scores represent in Tables 5 and 6 of the report. As the statistical analyses were incompletely described, whether proper analytic techniques were used was not possible to fully evaluate. (p. 5)

The authors claim I misunderstood the method applied in their study because I cited research that critiqued multiple regression alone as a means of assessing efficiency. I understood that the authors employed multiple regression as a first step in a two-stage procedure to generate efficiency scores, and described this in detail in the report. Further, I questioned why more sophisticated and accepted methods of measuring efficiency were not used, such as stochastic frontier analysis (SFA). In their blog critique, the authors defend their non-use of SFA and cite a study they say is critical of its application. However, the study they cite actually *encourages* the use of a more advanced SFA model:

Our approach extends the SFA model, allowing [the researcher] to disentangle inefficiency and skewness and nesting, as a particular case, the traditional SFA model. ... Therefore, the model we propose enriches the toolbox of researchers for performing efficiency analyses with parametric SFA models. (p. 18).

Hafner, C.M., Manner, H., & Simar, L. (2016): The "wrong skewness" problem in stochastic frontier models: A new approach. *Econometric Reviews*, DOI: 10.1080/07474938.2016.1140284

So the authors of *Bang for the Buck* reference this article as a justification for not using SFA, even though this article proposes an extended version of it. It is common practice to justify the use of one's method in the social sciences. But their original report made no mention of SFA (or the extended SFA model offered by the article above) or other established methods of efficiency analysis cited in contrast to the methods they chose to employ in their study.

In their critique, the authors seem to imply since I did not contact them, I was not conducting an honest review. I take issue with this implication. It is my right as a reviewer to critique what is in front of me. My determination that the methods were insufficiently described represents a valid criticism of a published report. This was not a manuscript subjected to peer scrutiny, in which there would be opportunities for a back and forth dialogue to critique and improve the study; rather, this was a final and published report. I reviewed what they

presented to the public.

My review’s claim, as cited by the WILL Blog on July 14:

4. “Autonomy” is never really defined—it is just used as a loose term implying independence—so autonomous behavior is assumed by virtue of their charter status. The report then makes strong but unmeasured claims about the superior “efficiency” of charter schools based on their having this greater autonomy.’

The authors claim that autonomy was defined on the first page of the Executive Summary (and page 6 of the full report). Yet simply noting the authorizing entity and hiring capabilities of various school types does not constitute, in my view, a very complete or useful portrayal of school autonomy. The table on the first page of the Executive Summary categorizes “autonomy” separately from “authorizer” and “employees,” which further confuses the issue. On page 6, the authors speak to the “ability to experiment with curriculum” as an additional distinguishing feature among charter school types in Milwaukee. They note that non-instrumentality charters “have more freedom...[and]...have a greater ability to experiment with curriculum and make changes based on the state-of-the-art in the teaching discipline.” Certainly non-instrumentality charters can avail themselves of these practices (although don’t all schools have the ability to make changes based on “state-of-the-art” practices?); however, how do the authors know this to be the case with their sample of non-instrumentality charter schools? And wouldn’t the reader want to know what this autonomy looks like in practice, particularly if such autonomy leads to positive changes? The authors make serious claims about charter school types that they believe enjoy more autonomy than others (scored rather crudely as either “limited,” “high,” or “none”). For such claims to have merit, there should be some substantiating evidence that the sample of schools, in fact, behaved as expected. That is, what did the schools do that was an exercise in autonomy and that other schools should learn from and also do? Moreover, all traditional (MPS) schools should not be assumed to behave without autonomy or without “complete control over the administration of their school.” Did all the studied charters engage in innovative practices, and if so what were those practices?

As a researcher, I was expecting to see a description of the authors’ conceptual framework for autonomy. Conceptual frameworks help define and can also serve as a basis to operationalize social science concepts. I encourage the authors in future studies to develop a more comprehensive conceptual framework for autonomy, such as the one found here:

Hanushek, Link, & Woessmann. (2013). Does school autonomy make sense everywhere? Panel estimates from PISA. *Journal of Development Economics*, 104, 212-232.

The larger issue here is a lack of validation of school autonomy measures (or categories of high, limited, none) in their study. My use of the wording “unmeasured claims” in the claim above is in reference to that omission. Better phrasing may be, “claims without basis.”

My review's claim, as cited by the WILL Blog on July 14:

5. “While the report’s analysis controls for some school demographic characteristics, it does not appear to adjust for selection effects; effects that could prove fatal to their conclusions.”

In their critique, the authors claim that “selection bias was accounted for as extensively as possible” through the use of statistical controls. I repeat what I wrote in the review:

Further, it is grossly oversimplistic to assume that a test score in a single year represents a school’s unique contribution to student achievement in those subjects. Schools obviously affect student learning, but there are myriad other influences on student learning that cannot be accounted for by statistical controls. Students are not randomly distributed across charter and public schools, and most statistical controls are unable to attenuate such selection bias. As parents of children choosing a charter school are different from non-choosers, it is unclear as to whether the differences in test scores are due to selection effects, the type of school or some other “third factor.” The correlational research design employed here is simply insufficient to lend support to any causal statements about effectiveness or efficiency in performance (of which this report offers many). (p. 7)

Conclusion

In the spirit of academic rigor, I appreciate that the authors took the time to respond to my review. However, their response does not acknowledge any shortcomings of their study pointed out in my review. Unfortunately, the report continues to offer far-reaching claims about Milwaukee public schools based on an overly simplistic logic and fatally weak research warrant.

Notes and References

- 1 Flanders, W., & Szafir, CJ (2016). *Bang for the buck: Which public schools in Milwaukee produce the best outcomes per dollar spent?* Milwaukee, WI: Wisconsin Institute for Law & Liberty. Retrieved August 10, 2016, from <http://nepc.colorado.edu/thinktank/review-WILL>
- 2 Cobb, C.D. (2016). *Review of "Bang for the Buck."* Boulder, CO: National Education Policy Center. Retrieved August 10, 2016, from <http://nepc.colorado.edu/thinktank/review-WILL>
- 3 Flanders, W., WILL Blog: *WILL Responds to Criticism From Colorado Group.* Milwaukee, WI: Wisconsin Institute for Law & Liberty. Retrieved August 10, 2016, from <http://www.will-law.org/will-blog-will-responds-criticism-colorado-group/>
- 4 Cobb, C.D. (2016). *Review of "Bang for the Buck."* Boulder, CO: National Education Policy Center. Retrieved August 10, 2016, from <http://nepc.colorado.edu/thinktank/review-WILL>