



A WEAK DEFENSE OF A USELESS REPORT



In April, the National Education Policy Center published a [review](#) of a Mackinac Center for Public Policy report entitled [The Michigan Context and Performance Report Card: High Schools 2018](#). The report is the fourth in a series ranking Michigan high schools based on their test scores while controlling for the percentage of students eligible for free school lunches. These Mackinac reports have been [used](#) to compare schools and assess educational quality within the state. The Mackinac Center is a libertarian think tank based in Michigan.

In his April review of the Mackinac report, NEPC Fellow [John T. Yun](#), an associate professor at Michigan State University, raised multiple concerns about the models underlying the school ranking. Among these concerns are the following:

- The report does not offer any conclusions or explanations about the reasons why schools might attain high versus low rankings.
- The analysis combines the results of multiple different tests that were used by the state during the four-year period covered by the report. Yet it makes no effort to equate these exams to ensure they are comparable or account for any differences.
- The report does not provide a rationale for using free lunch as the only measure of school context.
- The use of a single predictor (percentage of students qualifying for free lunches) over-simplifies and biases the estimates.

As a result, the data and analytic approach used in the report do not warrant the claim that the schools can be ordinally ranked in a reliable and precise way. Yun concluded by advising that:

[T]he rankings presented in this report should be given no weight in any discussions of policy or practice. In fact, this report does a disservice by introducing questionable information in an easily readable form that is not substantiated by any credible analysis.

The week after the publication of Yun's review, the report's authors [fired back](#) on the Mackinac Center's blog. In the response below, Yun addresses that blog item, focusing on the authors' contention that their model has the ability to accurately and reliably rank schools.

**

A Response to “Critique of CAP Report Card Fires Blanks”

By John T. Yun

Ben DeGrow and Michael Van Beek have published a response to my NEPC review of the Mackinac Center's latest “Context and Performance [CAP] Report Card.” In their response, they take issue with a several points made in the review. I will not address their points one by one, since many of them are minor and rest in the realm of reader interpretation. However, one main point that I would like to address lies at the very heart of the critique: *the ability of their model to accurately and reliably rank schools using their CAP Score metric.*

Putting aside all the measurement issues with the CAP Index that are highlighted in my review, consider the authors' claim that they made a choice for “simplicity” by using percent of free-lunch eligible 11th grade students as the sole predictor of a school's predicted CAP Index. Consider also the authors' claim that in previous versions of the report card, “When testing the impact of adding these other variables, we found that ‘the improvements in predictive power [of the model] are marginal, and including the additional variables would have only increased the model's complexity for little measurable gain.’”

In this statement, the authors are clearly conflating model fit with the reliability of a model's predictions. Given the key variables that were available and not included (urbanicity, school size, racial composition, per/pupil expenditures, percent special education students, percent English language learners, availability of advanced courses, etc.) it is likely that model fit would have been significantly improved, but more importantly the inclusion of different variables would likely have yielded different predicted scores for many of the schools.

For example, under the Mackinac Center's model, a small rural school with low minority and high special education enrollments that had the same percentage of free-lunch eligible students as a large, urban school with high minority enrollments would receive the same predicted score on the CAP Index. This would *not* happen if those additional variables were included in the model. The result of that new model would likely be a very different CAP Score for these specific schools—even if the overall model was only marginally more predictive. In addition, depending on the specific variables used (or the specification of the model) the predicted scores are likely to change from model to model. Thus, the school rankings are likely to shift from model to model as well leading to very unreliable rankings at the school level. My review's critique, therefore, covered the specific Mackinac model in addition to the usefulness of using any available model to generate these sorts of school-level rankings.

In their response, the authors of the Mackinac Center report seem to suggest that simply acknowledging the limitations of their approach and appealing to simplicity justifies the publication of their ranked results. My position (and the position of most academic researchers) is that the limitations of data limit the use to which you can put those data. Given that the authors of the report—and previous reports—do not demonstrate that their rankings are at all robust to different model specifications, and given that they themselves recognize the serious limitations in the data that they use, it should be very clear that their approach for ranking schools in this very precise manner (e.g., School X scores 98.0 and is therefore ranked higher than School Y at 97.9) is simply outside the ability of the methods and the data that they are using. This is the bottom line: the data and analytic approach used by the Center do not warrant the claim that the schools can be ranked reliably and precisely enough to publish them in this way.

If the Mackinac authors wanted to appeal to simplicity, a conclusion that would in fact be supported by this simple approach is that the share of free-lunch eligible students powerfully predicts their CAP Index of Michigan test scores, and the higher the percentage of students on free lunch, the lower their predicted CAP Index. This conclusion is consistent with a large body of prior research that argues student poverty predicts performance on standardized test scores. But any attempt to then extend these findings to tell us more about the relative performance of specific schools is unwarranted and misleading.

NEPC Resources on High-Stakes Testing and Evaluation

This newsletter is made possible in part by support provided by the Great Lakes Center for Education Research and Practice: <http://www.greatlakescenter.org>

The National Education Policy Center (NEPC), housed at the University of Colorado Boulder School of Education, produces and disseminates high-quality, peer-reviewed research to inform education policy discussions. Visit us at: <http://nepc.colorado.edu>