Supplement to Expert Reports of C. Kirabo Jackson (Revised)

In response to the rebuttal expert reports of Steven G. Rivkin and Jay P. Greene, C. Kirabo Jackson hereby supplements his expert reports:

1. The Greene report comments on differences in how the results of studies are classified in my 2018 working paper and my expert report. (Greene, p. 7) Greene appears to have not understood that each study reports on different (but closely related) aspects of the studies listed. Because both of my studies explicitly report on different aspects of papers, they are somewhat different. These differences are by design, and by no means an error. Specifically, there are "apparent" differences because the 2018 summary paper indicates the conclusions of the papers as they were reported by the authors. In contrast, the expert report performs a formal meta-analysis and is based on the results of calculations I made using the numbers reported in the papers. The 2018 paper is a literature review that summarizes the findings of papers in the past few years, and the report conducts a formal meta-analysis which involved additional calculations to ensure that all papers can be compared to each other. This should be clear to anyone who reads both.

There is a well-known distinction between a literature review and a formal meta-analysis. A quick google search of "literature review vs meta-analysis" yields the following. "*A literature review cites the conclusions of previous studies in order to provide a historical overview of a particular field of research. A meta- analysis also contains a summary of previous research, but in addition it compiles data from multiple studies, performs statistical analyses on this aggregate data, and then draws new conclusions from the results of this analysis.*" (https://www.quora.com/Whats-the-difference-between-a-literature-review-and-a-meta-analysis#:~:text=A%20literature%20review%20is%20a,of%20evidence%20on%20a%20question.).

To be more specific about differences, the classifications may and do differ because authors sometimes highlight and base their stated conclusions on a result that is significant for a subgroup (e.g. positive effects on math) although a summary based on all subjects might not be statistically significant. The 2018 literature overview paper is based on the authors' claims (i.e. the results highlighted by the authors), while the expert report is based on a summary measure of impacts reported in the paper for the full population and across all outcomes. The meta-analysis, as conducted for the report, is the more rigorous approach.

As a side note, it is important to keep in mind that Working Papers are not publications. They are versions of papers that are released to the public to facilitate feedback and debate before final revision. As such, there are almost always differences between a working paper and the

final version as methods are updated and models improved. It is also important to note that because some of the papers summarized are working papers themselves, the results in some of these papers have changed between 2018 and 2020. As the unpublished papers are updated, we also update our meta-analysis. As such, numbers can (and do change) for completely good reasons.

2. The Rivkin rebuttal report errs when it criticizes my expert report for not including studies using value-added models because value-added models are designed for answering very different kinds of questions from those addressed in my report. I will also note that no paper was excluded from the analysis because it used value-added models. Studies were excluded if they did rely on exogenous variation and could not demonstrate that the differences in school spending being examined were unrelated to other policy changes, or underlying differences in student populations.

It is also important to highlight that value-added models are in ill-suited to understand the effect of school spending (as distinguished from changes in spending) on outcomes. The reason is that value-added models always have a baseline test score (typically 3rd grade) that is supposed to "*account for student differences."* The problem is that if one wishes to understand the effect of school spending on test score growth, the baseline test score is itself affected by school spending – making it an inappropriate control variable for understanding the effect of school spending. Note that if the baseline score was from before kindergarten entry (i.e. a true baseline from before being exposed to the schooling system), then this value-added approach might be feasible. However, almost all value-added models use a baseline score established after school entry.

To make this clear, consider a world in which all students start out with a score (at birth) of 0. There are two schools A and B. School A raises baseline scores by 50 in kindergarten and 5 in first grade. School B raises baseline scores by 5 in kindergarten and 10 in first grade. Between birth and the end of first grade school A raises scores by 55 points, while school B raises them by only 15 – school A is clearly the better school at raising test scores. However, a value-added model that used the end of kindergarten score as the "baseline" would find that school A only raises scores by 5 while school B raises scores by 10. This would lead to the erroneous conclusion that school B is more effective than school A, although the opposite is true. This example highlights that the value-added models (as they are currently used) are often ill-suited to addressing questions about schools spending.

My work using value-added models do not suffer from this problem because I use these models to estimate the effect of high schools on outcomes and use 8th grade scores as the baseline. Using this approach, one could look at test scores in 8th grade (before entering high school) to

those in 9th grade (after having spent one year in a new school) and arguably isolate the effect of the school on 9th grade scores. This value-added approach can work for identifying the effect of high schools on 9th grade score because 8th grade scores are unaffected by the high school, while 9th grade scores are. This is exactly what I do in the working paper in question, which is now forthcoming at the American Economic Review: Insights.

3. The Rivkin rebuttal report errs when it criticizes the research and analysis in Jackson (2016), for example in its discussion of the implications of geographic mobility . The report disregards or overlooks our use of well recognized social science techniques.

4. When Dr. Greene characterizes the vote count analysis I use as being "virtually unheard of" (Greene p. 32) he is overlooking common statistical practice. He is apparently unaware of the Fisher Exact Test.

5. The funnel plot discussion in the Greene expert report (Greene pp. 35-38) is not is evidence of bias in my expert report for two reasons:

First, Greene presented a graph but did not present a formal test. To assess whether there is systematic relationship between a study estimate and the precision of the estimate one should regress the reported effect on measures of the studies' precision. When I do this, I do not find a robust statically significant relationship between the study estimate and the precision of the estimate – that is, formal statistical tests fail to ~~reject~~ show consistent evidence of publication bias.

Second, *even if the formal tests had indicated a relationship (which they do not)*, Funnel plot analysis does not in and of itself indicate publication bias. This is discussed in Egger, Smith, Schneider & Minder (1997) referred to in the Greene expert report, and Sterne, et. al. (2011), BMJ 2011;343:d4002.

6. Dr. Greene's use of a description from my 2016 paper (Greene p. 4) disregards that given publication lags the 2016 paper would have been written in 2013. At that time, there were very few other papers using credible research designs to estimate the effect of school spending. Our statement was accurate about the state of the literature at the time. A lot has changed since 2013.

/s/ C. Kirabo Jackson

Dated: 6/22/20 show consistent