The Heritage Foundation

SEPTEMBER 13, 2023

New Yorkers for Students' Educational Rights, et al. v. The State of New York: Expert Report of Jay P. Greene, PhD

INTRODUCTION

Lindsey M. Burke, PhD

Two years ago, Heritage Foundation Senior Research Fellow Jay Greene was retained by the state of New York to write an expert report as part of its defense in *New Yorkers for Students' Educational Rights, et al. v. The State of New York*.[1] The Heritage Foundation is now able to publish that report, enabling Americans to see the evidence debunking the claim that increasing education spending generally leads to improved student outcomes.

This claim has become almost a matter of consensus among education policy researchers, more than 450 of whom signed a group letter stating, "Research is abundantly clear that money matters for student achievement and other important life outcomes, and this is especially the case for low-income students."[2] That sentence contains four citations, all of which refer to research conducted by Kirabo Jackson, an economics professor at Northwestern University. Jackson also served as an expert witness in *New Yorkers for Students' Educational Rights, et al. v. The State of New York*, but on behalf of the plaintiffs. The report reproduced below by Jay Greene is a rebuttal of Jackson's claims about the effects of increasing school spending.

SERIOUS FLAWS IN SCHOOL SPENDING RESEARCH

Greene's expert report reveals that Jackson's evidence is plagued by errors and characterized by p-hacking, or specification-shopping, techniques with which researchers alter their analyses, deliberately or unconsciously, to ensure that they yield desired results. The practice of p-hacking or specification-shopping renders the results unreliable.

Because Jackson produced six different versions of his meta-analysis, or systematic review, of the evidence on the causal relationship between school spending and student outcomes for different legal proceedings, it is possible to see how Jackson changed the analysis to engineer more favorable results. For example, Greene's expert report documents:

> There are 37 studies that Dr. Jackson at one time or another thought appropriate to consider when reviewing the evidence on how additional spending affects student outcomes. Dr. Jackson has changed his mind about whether to include the study or how to characterize the direction or significance of the results for 22 of those 37 studies. For some of those studies, he has changed his characterization of the results more than once across these six reviews in the last two years.

Since most studies report multiple results, meta-analyses or systematic reviews must apply consistent rules for selecting which results are considered or how multiple results should be combined. On more than one occasion, Jackson changed those rules and then inconsistently applied them to select positive outcomes and avoid negative ones.

In a flagrant example of p-hacking observed in real time, Jackson switched an analysis that produced a null finding reported in the November 2020 version of his meta-analysis so that in the March 2021 version he found a statistically significant effect with a p-value of .0497, just below the standard threshold of .05 for classifying a result as statistically significant. Changing analyses to achieve p values below a critical threshold is the very definition of p-hacking.

Jackson also failed to follow standard procedures for identifying the full set of studies he should include in his meta-analysis. Instead, he simply did a Google search and asked followers on Twitter for suggestions. Not surprisingly, the set of studies he found was skewed to exclude those with negative results.

In another study, Jackson sought to demonstrate that cuts in education spending during the Great Recession led to worse outcomes for students in states where education funding is more dependent on state revenues that are more sensitive to economic cycles than local property taxes. There are multiple data sources for information on the share of education spending that comes from state revenues. There is also a difficulty in deciding what the "state" share of education spending is in the District of Columbia, where the state and local government are one and the same.

Jackson made the indefensible decision to assign the District a value of 0 for the state share of education revenue even though his data source listed the information for DC as "not applicable." He also chose to rely solely on Census data for the state share of education spending rather than also considering data from the U.S. Department of Education's National Center for Education Statistics. Had he made different decisions, he would have found a null result in most alternative specifications. Had he excluded DC from the analysis, which seems like the most reasonable solution, 23 of 30 possible permutations of data sources for his two outcomes would have yielded null results.

In addition, Jackson chose to divide states into three groups based on whether they have high, medium, or low shares of education spending from state revenue. Had he chosen slightly different cut-offs for classifying states into these groups, he would have found null results in almost all cases. In fact, of the 120 permutations that Greene examined of how one might classify the states and which data sources to use, Jackson would have found null effects in 104 of those analyses. Only by making very specific and insupportable decisions about model specification and data sources is he able to find significant, positive effects for school spending. This is the very definition of specification-shopping.

As Greene's expert report states, "The practice of making or altering research decisions after results are known in a way that supports one's argument is known as cherry-picking, p-hacking, or specification-shopping. These practices are fundamentally inconsistent with generally accepted scientific principles. It is false to claim a bull's-eye if the circles are drawn after the arrow lands."

INTELLECTUAL CORRUPTION IN SOCIAL SCIENCE

Despite all the obvious defects of Jackson's research, his meta-analysis was recently accepted for publication in a leading economics journal[3] and his analysis of the effects of cuts in education spending during the Great Recession was earlier published in another leading economics journal.[4] The Biden Administration also just announced the appointment of Jackson to the White House's Council of Economic Advisors in part because of his research on the benefits of increased school spending.[5]

How did the journal editors, the articles' peer reviewers, more than 450 education policy researchers, and the staff who vet presidential appointments all miss the academic dishonesty in Jackson's research on this issue? Scientific processes, such as peer review, are intended to detect these kinds of problems and ensure the quality of published research. In addition, well-trained social scientists should approach research with skepticism and should not collectively embrace findings as “settled science” without overwhelming evidence and careful scrutiny.

Social scientists have abandoned the standards of their disciplines and betrayed the rigor of their training for the allure of advancing political goals and receiving social acceptance and career advancement within their professional communities.

Many social scientists want to believe that additional school spending would improve student outcomes, regardless of what the evidence shows. In addition, Jackson is a high-status academic who is an editor at two top economics journals. In fact, one of those journals published his Great Recession paper. He may have followed the formal process of recusing himself from the decision to publish that article, but the close personal connections and the dangers of crossing someone in a powerful position would nevertheless undermine the quality control that editors and the peer-reviewers they select are supposed to ensure.

Economics, in particular, once stood out among the social sciences for having greater methodological rigor and a combative culture that would facilitate careful scrutiny of research. But like many institutions that once served as checks on concentrated power, from journalism to professional associations, social science has been corrupted by the allure of that power.

That is why it is so important that The Heritage Foundation publish Greene’s expert report. It not only documents the falsehood of specific claims about the benefits of additional school spending, it reveals the ways in which social science was unable to detect and reject shoddy research. Nothing about this report is out of date and no subsequent revisions in Jackson’s research can undo the defects revealed here. Once a study is marred by p-hacking, that bell cannot be un-rung. The results have been compromised by selecting favorable results after doing too many analyses, which cannot be corrected by doing yet more analyses.

Now that Greene’s report is available to the public, everyone can see those defects and how they were ignored.

ENDNOTES

1. Greene’s expert report was prepared with assistance from Josh McGee, who also presents some of the work he did for this report in the following working paper: Jessica Goldstein and Josh B. McGee, “Did Spending Cuts During the Great Recession Really Cause Student Outcomes to Decline?” EdWorkingPaper 20-303, 2021, https://doi.org/10.26300/qzrd-0323 (accessed September 7, 2023).
2. A Letter from Education Researchers, “How Schools Can Help Children Recover from COVID School Closures, August 3, 2020, https://docs.google.com/document/d/1Pi-u7TJVIMVPIaGZ9QhJuT9VJ6URuvW3ki8diELwf7o/edit?pli=1 (accessed September 7, 2023).
3. C. Kirabo Jackson and Claire L. Mackevicius, “What Impacts Can We Expect from School Spending Policy? Evidence from Evaluations in the U.S.,” *American Economic Journal: Applied Economics*, forthcoming. https://www.aeaweb.org/articles?id=10.1257/app.20220279&&from=f (accessed September 7, 2023).
4. C. Kirabo Jackson, Cora Wigger, and Heyu Xiong, “Do School Spending Cuts Matter? Evidence from the Great Recession,” *American Economic Journal: Economic Policy*, Vol. 13, No. 2 (May 2021), https://www.aeaweb.org/articles?id=10.1257/pol.20180674&from=f (accessed September 7, 2023).
5. Kevin Mahnken, “Four Top Takeaways from the School Research of New Biden Advisor Kirabo Jackson,” The 74, August 14, 2023, https://www.the74million.org/article/four-top-takeaways-from-the-school-research-of-new-biden-advisor-kirabo-jackson/ (accessed September 7, 2023).

Expert Report of Jay P. Greene, Ph.D.

New Yorkers for Students' Educational Rights, et al. v. The State of New York

May 21, 2021

Introduction

I have been retained by the New York State Office of the Attorney General to evaluate and respond to a number of the conclusions and opinions set forth in the Expert Report of C. Kirabo Jackson. Dr. Jackson's report argues that scientifically valid research provides "overwhelming" (p. 3) support for the claim that, in general, additional school spending improves student outcomes and cutting school spending, as occurred during the Great Recession, harms student outcomes. Applying these findings to Schenectady and New York City, he asserts with near certainty that "school spending cuts halt progress, lower student outcomes, and cause lasting negative educational harm, particularly for the most vulnerable students." (p. 5) To make these arguments Dr. Jackson relies on three pieces of research on which he is a co-author and which are attached to his report: two versions of a systematic review, or meta-analysis, of what he claims to be a complete and unbiased set of methodologically appropriate studies regarding the relationship between additional school funding and student outcomes; and a study on the academic effects of school spending cuts that occurred during the Great Recession.

There are a number of serious flaws in the evidence and arguments Dr. Jackson makes, which are enumerated in the Key Claims below. There is also a pattern connecting many of these flaws. Across the three studies on which he relies, Dr. Jackson has choices he can make in selecting among data sources, applying inclusion and classification criteria, choosing model specifications, and picking outcomes to examine. His results are highly sensitive to those choices. If he had made different decisions among options that were indistinguishable a priori, his results would undermine the conclusions he has drawn. On some occasions, we can see that Dr. Jackson changed his choices over time. On other occasions, Dr. Jackson may have tried different selections before publicly reporting the results of any of them.

The practice of making or altering research decisions after results are known in a way that supports one's argument is known as cherry-picking, p-hacking, or specification-shopping. These practices are fundamentally inconsistent with generally accepted scientific principles. It is false to claim a bull's-eye if the circles are drawn after the arrow lands.

The sensitivity of Dr. Jackson's claims to these research decisions and evidence of his altering decisions after results are known make his report more a work of advocacy than expert opinion. This unscientific pattern, in addition to the specific flaws described below, prevent Dr. Jackson's report from offering reliable information about the likely consequences of changes in school spending in Schenectady and New York City.

Based on my reading of Dr. Jackson's report, the underlying research on the subject, as well as my experience, I have the following opinions, to a reasonable degree of professional certainty:

Key Claim 1 – Dr. Jackson's report and the three studies he appends to it are so riddled with errors, inconsistencies, and ambiguities that they are not reliable evidence for predicting what might happen if school spending were changed in Schenectady and New York City.

Key Claim 2 – Dr. Jackson's Great Recession analysis is not robust to reasonable changes in model specification, data source, or outcome selection. The sensitivity of his results to numerous, very specific choices renders the work unreliable for predicting the effects of changes in school spending in Schenectady and New York City.

Key Claim 3 – Dr. Jackson's reviews of the research, or meta-analyses, on the relationship between additional school spending and student outcomes are not based on a complete and unbiased set of studies. They are therefore uninformative regarding how possible education spending changes might affect student outcomes in Schenectady and New York City.

Key Claim 4 – Dr. Jackson's reviews of the research, or meta-analyses, focus on a set of studies that he claims "employ quasi-experimental methods to isolate causal impacts." (p. 7) While the research designs Dr. Jackson prioritizes can approximate the causal estimates derived from actual experiments if certain assumptions are strictly met, in the set of studies Dr. Jackson considers, those assumptions are routinely violated and their results should not be considered causal. In addition, Dr. Jackson's own analyses of spending and student outcomes in Schenectady and New York City do not even attempt to use credibly causal research designs.

Key Claim 5 – The particular context, including initial spending levels and political circumstances, differs dramatically between most of the studies Dr. Jackson includes in his reviews of the literature and the current context in Schenectady and New York City.

The documents and data upon which my opinions are based are cited throughout this report. Because Dr. Jackson presented similar opinions in a recent school funding case in Delaware, I have drawn upon some of the same language I used for my expert report in that case for this report.

Key Claim 1 – Dr. Jackson's report and the three studies he appends to it are so riddled with errors, inconsistencies, and ambiguities that they are not reliable evidence for predicting what might happen if school spending were changed in Schenectady and New York City.

A) Dr. Jackson has conducted a review or meta-analysis of research on the relationship between additional spending and student outcomes six times in the last two years. The set of studies he considers appropriate to include as well as his characterization of the results has changed in each of these six reviews. Exhibit 1 presents a summary of those six reviews and how his characterization of the same studies has changed. Two of those reviews, dated November 30, 2020 and March 3, 2021, are appended to Dr. Jackson's expert report as Exhibits C and E. While those two versions of the review, both contained in the same expert report, do not differ from each other in their classification of the direction and statistical significance of results, they do frequently differ in how they describe the magnitude of effects in each study, as will be discussed below.

There are 37 studies that Dr. Jackson at one time or another thought appropriate to consider when reviewing the evidence on how additional spending affects student outcomes. Dr. Jackson has changed his mind about whether to include the study or how to characterize the direction or significance of the results for 22 of those 37 studies. For some of those studies, he has changed his characterization of the results more than once across these six reviews in the last two years.

In Exhibit 1, each row represents a study that Dr. Jackson deemed appropriate to include in at least one of his reviews of evidence on the relationship between additional school spending and student outcomes. The grouping of columns represents each of Dr. Jackson's reviews over the last two years. Within those groupings are columns that indicate whether Dr. Jackson thought a study met his criteria to be included in that review as well as columns indicating how Dr. Jackson characterized the direction and statistical significance of each study's results. "Pos." is an abbreviation for Dr. Jackson describing a study's results as finding a positive relationship between increasing school spending and improving student outcomes, while "Neg." is an abbreviation for Dr. Jackson describing a study as having a negative relationship between additional spending and outcomes. "Sig." is an abbreviation for Dr. Jackson describing the results of a study as being statistically significant. Statistical significance generally means that a researcher is reasonably confident the results of a study differ from zero in the direction indicated by "Pos." or "Neg." "Not Sig." is an abbreviation for not statistically significant and generally means that a researcher cannot be reasonably confident that the result is distinguishable from zero. The yellow highlight marks instances where Dr. Jackson's characterization of results differs from that found in the current versions of his review of literature appended to his New York report.

Exhibit 1 – Summary of Dr. Jackson's Reviews of the Literature Over the Last Two Years

	Jackson (2018)				Jackson, Wigger, & Xiong (2018)			Jackson Delaware Report (March 2020)		Jackson Delaware Revised Report (July 2020)			Jackson New York Reports (November 2020 and March 2021)			
Study	Pos. & Sig.	Not Sig.	Neg. & Sig.	Meets Criteria for Inclusion	Pos. & Sig.	Pos. & Not Sig.	Neg & Not Sig.	Pos. & Sig.	Pos. & Not Sig.	Pos. & Sig.	Pos. & Not Sig.	Meets Criteria for Inclusion	Pos. & Sig.	Pos. & Not Sig.	Neg & Not Sig.	Meets Criteria for Inclusion
Abott, Korgan, Lavertu, & Peskowitz (2019)				NA		Yes		Yes		Yes		Yes		Yes		Yes
Baron (2019)				NA				Yes		Yes		Yes		Yes		Yes
Biasi (2019)	Yes			Yes				Yes		Yes		Yes	NA			Yes
Brunner, Hyman, & Ju (2019)	Yes			Yes		Yes			Yes		Yes	Yes	Yes			Yes
Candelaria & Shores (2019)	Yes			Yes					Yes	Yes		Yes	Yes			Yes
Card & Payne (2002)	Yes			Yes				Yes		Yes		Yes	NA			Yes
Carlson & Lavertu (2018)				NA					Yes		Yes	Yes		Yes		Yes
Cascio, Gordon, & Reber (2013)	Yes			Yes					Yes		Yes	Yes	Yes			Yes
Cellini, Ferreira, & Rothstein (2010)		Yes		Yes				Yes		Yes		Yes	Yes			Yes
Clark (2003)		Yes		Yes		Yes			Yes		Yes	Yes		Yes		Yes
Conlin & Thompson (2017)	Yes			Yes					Yes		Yes	Yes		Yes		Yes
Downes & Figio (1997)	Yes			Yes								No				No
Gigliotti & Sorensen (2018)	Yes			Yes	Yes			Yes		Yes		Yes	Yes			Yes
Goncalves (2015)		Yes		Yes					Yes		Yes	Yes			Yes	Yes
Guryan (2001)	Yes			Yes		Yes		Yes			Yes	Yes		Yes		Yes
Holden (2016)	Yes			Yes					Yes		Yes	Yes				No
Hong & Zimmer (2016)	Yes			Yes					Yes		Yes	Yes		Yes		Yes
Hoxby (2001)		Yes		Yes								No				No
Husted & Kenny (2000)	Yes			Yes								No				No
Hyman (2017)	Yes			Yes				Yes		Yes		Yes	Yes			Yes
Jackson, Johnson, & Persico (2016) Johnson & Jackson (2019)	Yes			Yes				Yes		Yes		Yes	Yes			Yes
Jackson, Wigger, & Xiong (2018)	Yes			Yes				Yes		Yes		Yes	Yes			Yes
Johnson (2015)	Yes			Yes				Yes		Yes		Yes		Yes		Yes
Kogan, Lavertu, & Peskowitz (2017)	Yes			Yes		Yes		Yes		Yes		Yes		Yes		Yes
Kreisman & Steinberg (2019)				NA	Yes			Yes		Yes		Yes	Yes			Yes
LaFortune & Schonholzer (2018)	Yes			Yes				Yes		Yes		Yes	Yes			Yes
LaFortune, Rothstein, & Schanzenbach (2018)	Yes			Yes	Yes				Yes		Yes	Yes		Yes		Yes
Lee & Polachek (2018)	Yes			Yes				Yes		Yes		Yes	Yes			Yes
Martorell, Stange, & McFarlin (2016)		Yes		Yes					Yes		Yes	Yes		Yes		Yes
Matsudaira, Hosek, & Walsh (2012)		Yes		Yes			Yes		Yes		Yes	Yes				No
Miller (2018)	Yes			Yes	Yes			Yes		Yes		Yes	Yes			Yes
Neilson & Zimmerman (2014)	Yes			Yes					Yes		Yes	Yes		Yes		Yes
Papke (2008)	Yes			Yes				Yes		Yes		Yes	Yes			Yes
Rauscher (2019)				NA					Yes		Yes	Yes		Yes		Yes
Roy (2011)	Yes			Yes				Yes		Yes		Yes	Yes			Yes
Van der Klaauw (2008)		Yes		Yes								No				No
Weinstein, Stiefel, Schwartz, & Chalico (2009)		Yes		Yes			Yes		Yes		Yes	Yes			Yes	Yes

Indicates discrepancies or differences between Jackson's earlier studies and reports and Jackson's New York Reports (November 2020 and March 2021).

The changes summarized in Exhibit 1 only capture whether Dr. Jackson altered whether the study was appropriate to be included in the review, whether the results were positive or negative, and whether those results were statistically significant or not. If Dr. Jackson altered his characterization of the magnitude of effects reported by studies but that change did not switch that study from being classified as positive or negative or significant or not, then we could not observe those changes in Exhibit 1. We can, however, compare Dr. Jackson's description of the average effect for each study in Table 1 of his Exhibit C to the figures he included in his Delaware report to see whether he has changed their estimated effects, even if they have not been shifted between categories marking the direction and significance of results. We can also compare the description of the average effect for each study in Table 1 of Exhibits C and E, the two different versions of Dr. Jackson's meta-analysis that he includes in his expert report.

A summary of the changes in Dr. Jackson's description of each study's effects can be seen in Exhibit 2. Of the 29 studies for which Dr. Jackson describes results across these three versions of his meta-analysis produced in the last year and presented in expert reports for school finance litigation, the results have changed for 17 of them at least once. That is, within the last year Dr. Jackson has altered his description of the average effects for a majority of studies in his meta-analysis without acknowledgement or justification. The majority of these changes have resulted in attributing larger effects to studies.

The numbers from Dr. Jackson's Delaware report are based on a visual estimate of his figures, so they may be imprecise by as much as .02 of a standard deviation. But it is clear that for 11 of the 29 studies for which Dr. Jackson describes results in both his Delaware and New York reports the results have been substantially changed. In the November 2020 analysis presented in his New York report, Dr. Jackson finds larger benefits for nine of those 11 studies for which he has changed his description of their effects, relative to what he stated in his Delaware report. For the two versions of the meta-analysis (November 2020 and March 2021) that are both included in Dr. Jackson's expert report, the results differ for nine of the 29 studies. Seven of those nine studies have larger effects in the March 2021 version of the meta-analysis than in the November 2020 version.

The average absolute value of the change in Dr. Jackson's description for those studies where the effects have been altered is 12% of a standard deviation. The average effect of all 29 studies as reported by Dr. Jackson in his March 2021 version of the meta-analysis is 17% of a standard deviation, so the magnitude of the revisions across these three versions of the meta-analysis is quite large. To put the magnitude of Dr. Jackson's revisions of results further in perspective, in his report Dr. Jackson claims that "on average, increasing school spending by $1000 per pupil in 2018 dollars (and sustained over four years) would increase test scores by roughly 4.4 percent of a standard deviation." (p. 4) The average revision in Dr. Jackson's description of study effects is almost three times as large as the average benefit he claims from spending increases. These large changes in how Dr. Jackson describes the effects of studies in his meta-analysis occurred in less than a year and were all presented to courts as expert testimony, two of which are contained in the same, current expert report.

Exhibit 2 – The Change in Dr. Jackson's Descriptions of Effects

Study	Effect Size from Revised Delaware Report (July 2020)	Effect Size from New York Report (November 2020)	Effect Size from New York Report (March 2021)
Abott Kogan Lavertu Peskowitz (2020)	0.14	0.10025	0.10025
Baron (2020)	0.03	0.2198	0.1176
Biasi (2019)			
Brunner Hyman Ju (2020)	0.06	0.0531	0.0531
Candelaria Shores (2019)	0.15	0.1435	0.1435
Card Payne (2002)			
Carlson Lavertu (2018)	0.09	0.0902	0.0902
Cascio Gordon Reber (2013)	0.17	1.1837	1.1837
Cellini Ferreira Rothstein (2010)	0.10	0.212	0.212
Clark (2003)	0.01	0.0148	0.0148
Conlin Thompson (2017)	0.02	0.0159	0.0323
Gigliotti Sorensen (2018)	0.06	0.0424	0.0424
Goncalves (2015)	0.01	-0.005	-0.0048
Guryan (2001)	0.02	0.0165	0.0281
Holden (2016)			
Hong Zimmer (2016)	0.11	0.2545	0.3265
Hyman (2017)	0.12	0.1109	0.1109
Jackson Johnson Persico (2015), Jackson Johnson (2019)	0.11	0.1897	0.1897
Jackson Wigger Xiong (2020)	0.05	0.09335	0.09335
Johnson (2015)	0.35	0.3417	0.3417
Kogan Lavertu Peskowitz (2017)	0.01	0.0219	0.019
Kreisman Steinberg (2019)	0.15	0.0912	0.0912
Lafortune Rothstein Schanzenbach (2018)	0.02	0.0123	0.0164
Lafortune Schonholzer (2019)	0.08	0.0798	0.233
Lee Polachek (2018)	0.25	0.4703	0.4778
Martorell Stange McFarlin (2016)	0.03	0.0304	0.0304
Matsudaira Hosek Walsh (2012)			
Miller (2018)	0.10	0.0941	0.0941
Neilson Zimmerman (2014)	0.02	0.0248	0.0248
Papke (2008)	0.16	0.1652	0.1652
Rauscher (2020)	0.01	0.0166	0.0286
Roy (2011)	0.14	0.3804	0.3804
Weinstein Stiefel Schwartz Chalico (2009)	0.07	0.1625	0.1625

Indicates discrepancies or differences > .02 between Jackson's Delaware (July 2020) and New York (Nov. 2020) reports and any differences between New York Nov. 2020 and March 2021.

Dr. Jackson tells a positive story from all of these versions of his meta-analysis, but the ever-changing story renders it fundamentally unreliable and unscientific. In his Delaware report, Dr. Jackson claimed that the odds of the positive pattern he described from his review of the literature being produced by chance were "one in 8,589,934,592." (p. 12) In his New York report, Dr. Jackson has revised that estimate to say, "If there were no effect, the likelihood of observing 29 or more positive effects out of 31 studies is 1 in 4,320,893." (p. 9) Both of these claims of near-certainty are highly misleading because they depend on the false assumption that each of the studies is a truly independent observation, when in fact they have overlapping sets of authors, examine overlapping interventions, and use common methods that require overlapping assumptions. Leaving this aside, it is striking that Dr. Jackson's quantification of his near-certainty decreased by nearly 2,000-fold in less than a year. It is not credible to be nearly certain and to change the magnitude of that certainty so dramatically.

B) Most studies included in Dr. Jackson's meta-analysis report multiple findings, derived from different analytical models and examining different outcomes over different time-periods. Quite often, within the same study those multiple findings point in different directions and have different levels of statistical significance. A proper meta-analysis needs to clearly articulate and consistently apply criteria for selecting which results within studies to feature. Dr. Jackson fails to do this.

A clear example of this inconsistent, and self-serving, selection of results to feature in the meta-analysis can be seen in how Dr. Jackson summarizes the results from Kogan, et al. (2017). For all of the non-capital studies, including Kogan, et al., Dr. Jackson says that he focuses on the effects after four years of additional spending. This is confirmed by the fact that in the worksheet Dr. Jackson provided, "SchoolSpendingPapers," column D lists "Years_Expose" as "4" for almost all studies, including Kogan, et al. Yet, when we look at the table and model specification from which Dr. Jackson draws the result from Kogan, et al., it is clear that Dr. Jackson has taken the result that is "3 years after" rather than "4 years after." Keeping in mind that Kogan, et al. examine the effect of a school tax levy failing, a negative coefficient would represent higher test scores from additional spending and a positive one would represent lower test scores. Dr. Jackson selects the favorable result highlighted in the black box in Exhibit 3 as opposed to the unfavorable result highlighted in the red box. Dr. Jackson deviates from his normally used criteria and practice in selecting the result more favorable to his argument. This is an obvious instance of cherry picking.

Exhibit 3 – Results from Kogan, et al. (2017)

Table 7. Impact of Tax Levy Failure on Student Achievement (Student-Lev

	State "Value Added"	
	(1)	(2)
2 years prior	–0.126 (0.129)	–0.0239 (0.0902)
1 year prior	—	—
Election year	–0.109 (0.115)	–0.0333 (0.0844)
1 year after	–0.0999 (0.119)	–0.0400 (0.0895)
2 years after	–0.222* (0.106)	–0.136^ (0.0800)
3 years after	–0.201^ (0.111)	–0.140 (0.0849)
4 years after	–0.0307 (0.110)	0.0750 (0.0888)
5 years after	–0.135 (0.120)	–0.0776 (0.0983)
6 years after	–0.163 (0.127)	0.00584 (0.101)
N	24,796	24,796
District count	571	571
Levy count	4,324	4,324
Mean dependent variable	0.01	0.01
Model	RD	RD
Specification	Quad.	Linear
Restricted bandwidth	No	No
Levy type	Op. & Cap.	Op. & Cap.

Another example of how Dr. Jackson's inconsistent criteria for selecting results within studies to feature in his meta-analysis tilts the evidence in favor of his argument can be seen in Cellini, et al. (2010). While Dr. Jackson focuses on results after four years for non-capital projects, in his Exhibit C he chooses to "measure outcomes six years after the increase in capital spending." (p. 13) In his Delaware report, submitted less than a year ago, Dr. Jackson chose to "relate the year-five change in outcomes to the average spending levels…" (p. 26) Yet, in the worksheet Dr. Jackson provided in this case, "SchoolSpendingPapers," column D lists "Years_Expose" as "4" for Cellini, et al. Whether Dr. Jackson chooses to feature the results after three, four, five, or six years is a moving target.

If Dr. Jackson had focused on the Cellini result after four years, he would have had a negative result, as can be seen in Exhibit 4. If Dr. Jackson had focused on results after five years, the result would have been positive, but less so than after six years, which is the finding he chooses to feature.

Exhibit 4 – Results from Cellini, et al. (2010)

TABLE VII
THE EFFECT OF BOND PASSAGE ON ACADEMIC ACHIEVEMENT, HOUSING MARKET TRANSACTIONS, AND HOMEBUYER AND SCHOOL DISTRICT CHARACTERISTICS: ONE-STEP ESTIMATES OF TOT EFFECTS

	1 yr later (1)	2 yrs later (2)	3 yrs later (3)	4 yrs later (4)	5 yrs later (5)	6 yrs later (6)	*N* (7)
A. Academic achievement							
Reading, grade 3	−0.010	−0.023	0.058	−0.026	0.039	0.103	6,660
	(0.054)	(0.051)	(0.053)	(0.058)	(0.061)	(0.064)	
Math, grade 3	−0.012	−0.034	0.030	0.026	0.058	**0.160**	6,660
	(0.057)	(0.054)	(0.062)	(0.062)	(0.069)	(0.075)	

In his Exhibit C for the New York report, Dr. Jackson makes a case for focusing on results six years after the initiation of capital projects, saying that “capital projects often entail some temporary disruption to everyday operations during the renovation/construction period that may be deleterious to student outcomes.” (p. 13) But this argument does not justify why it is best to focus on results four, five, or six years after capital projects begin, only that the initial results may be different from later ones. He also only makes the case for focusing on results in the sixth year after knowing that results were generally more positive in that year than using his previous standard of focusing on results after five years. And the fifth year standard is more positive than using the same standard as non-capital interventions and focusing on results after four years.

It is also unclear why it would be appropriate only to consider positive results in out-years while ignoring the harms inflicted in the initial years. The cumulative negative impact on student achievement seen in several studies in the early years may equal or exceed the gains observed in a single, sixth year. Dr. Jackson could have calculated an average of effects over multiple years, just as he calculates an average of spending changes over multiple years. His choice to focus only on the sixth year after capital projects begin is difficult to explain other than that it produces more favorable results for Dr. Jackson’s argument.

C) In the New York report, Dr. Jackson says that his systematic review of the appropriate research “provides overwhelming evidence that policies that increase school spending improve student test scores, educational attainment, and wages.” (p. 10) But Dr. Jackson and other researchers on whom Dr. Jackson relies for his conclusion, describe the research literature very differently in their scholarly work. In the studies included in Dr. Jackson’s meta-analysis, these scholars describe the relationship between additional spending and student outcomes as “mixed,” “contradictory,” and “inconclusive.” Listed below are quotations from those researchers, including Dr. Jackson, himself:

Jackson, Johnson, Persico (2016): "Overall, the evidence on the effects of SFRs [school finance reforms] on academic outcomes is mixed, and the effects on long-run economic outcomes is unknown." (p. 160)

Lafortune, Rothstein, and Schanzenbach (2018): "SFRs are arguably the most substantial national policy effort aimed at promoting equality of educational opportunity since the turn away from school desegregation in the 1980s. But there is little evidence about their effects on student achievement…. The literature regarding whether 'money matters' in education (Hanushek 1986, 2003, 2006; Card and Krueger 1992a; Burtless 1996) is contentious and does not offer clear guidance." (pp. 2-3)

Cellini, Ferreira, and Rothstein (2010): "Despite the importance of capital spending, little is known about the overall impact of public infrastructure investment on economic output, and even less is known about the effects of school facilities investments…. Also closely related is the long literature on the effects of school spending more generally. Hanushek (1996) reviews more than ninety studies and concludes that '[s]imple resource policies hold little hope for improving student outcomes,' but Card and Krueger (1996) dispute Hanushek's interpretation of the literature…. Angrist and Lavy (2002) and Goolsbee and Guryan (2006) exploit credibly exogenous variation in school technology investments. Neither study finds shortrun effects on student achievement." (p. 216)

Lafortune and Schonholzer (2018): "The empirical literature on capital expenditures offers little guidance with regard to these questions. Several studies find no or imprecise effects of capital expenditures on student achievement (see Cellini et al., 2010; Bowers and Urick, 2011; Goncalves, 2015; Martorell et al., 2016), while others find some evidence of positive impacts on student achievement, often only in reading and English-language arts (Welsh et al., 2012; Neilson and Zimmerman, 2014; Hong and Zimmer, 2016; Conlin and Thompson, 2017; Hashim et al., 2018). Other studies have looked at longer-run impacts of school construction programs in other countries that expand access to education (e.g. Duflo, 2001, 2004), measuring the effects of more general increases in human capital accumulation. Despite inconclusive evidence in the literature and general skepticism among economists, resource-based capital expenditure programs continue to be used by policymakers at the state and local level as tools to improve schools and reduce achievement gaps." (pp. 1-2)

Lee and Polachek (2018): "Current analyses find contradictory evidence of the effect of school expenditures on dropout and graduation rates." (p. 131)

Rauscher (2019): "Debates about the efficiency of education funding for student achievement have continued at least since the 1966 Coleman Report (e.g., Hanushek

1989, 1996; Burtless 1996; Greenwald et al. 1996; see Biddle and Berliner 2002 and Baker 2016 for reviews), including contemporary evidence of no relationship between funding and achievement (Morgan and Jung 2016)…. Existing research provides contradictory evidence about the effects of education funding on student achievement (e.g., Jackson, Johnson and Persico 2016; Morgan and Jung 2016).” (pp. 1, 27)

Johnson (2015): “Despite its fiscal importance, evidence on the effectiveness of Title I is mixed (Matsudaira, Hosek, and Walsh 2012; Cascio, Gordon, and Reber 2013; Van der Klaauw 2008).” (p. 50)

Papke (2008): “Yinger (2004) discusses education finance litigation and resulting reforms to state finance systems. He concludes that, while some of the evidence indicates that state aid can boost student performance, none of the findings is definitive and some are quite ambiguous.” (p. 466)

Hyman (2017): “However, it is less clear whether the changes in spending affected student achievement, with some studies finding positive effects and others finding no effects.”

Conlin and Thompson (2017): “Recent literature has focused on using quasi-experimental designs to identify the causal effect of capital investment on student outcomes and housing prices. A set of quasi-experimental papers (Cellini et al., 2010; Hong & Zimmer, 2016; Kogan, Lavertu, & Peskowitz, 2017; Martorell, Stange, & McFarlin, 2016) estimate regression discontinuity designs using the majority rule cutoff in school bond referendum elections to compare outcomes (test scores and/or housing prices) for districts that just pass a bond referendum to fund additional capital expenditures to those that just fail to pass a bond referendum and generally find mixed evidence on the role of capital investments on student achievement.” (p. 14)

Dr. Jackson’s claim that three is “overwhelming evidence that policies that increase school spending improve student test scores, educational attainment, and wages” (p. 10) is at odds with his own prior description of the research literature, and the descriptions of the bulk of the researchers in the studies on which Dr. Jackson relies for his claim. The fact that these contradictory descriptions are typically found in peer-reviewed articles further supports the view that Dr. Jackson’s bold claim about “overwhelming evidence” is inconsistent with what the research community believes about the pattern of findings on this issue.

D) School finance reforms (SFRs) are the basis for a large portion of the research in Dr. Jackson’s meta-analyses (presented in his Exhibits C and E) claiming that additional spending improves student outcomes. However, there is considerable disagreement among the researchers who examine SFRs as to what constitutes an SFR and when the event occurred. For example, in the 2018 article Jesse Rothstein co-authored with Lafortune and Schanzenbach, they provide an online appendix (p. ix) documenting all of the differences between how they treated SFRs and how Jackson, Johnson, & Persico (2016) did. In total, there were 23 disagreements in 14 different states over what constituted a school finance reform and when it occurred. (See Exhibit 5) Obviously, the numerous studies relying on SFRs to identify the effects of additional spending depend to a large degree on the subjective judgment of the researchers about how and when events should be classified. Any enterprise that dependent on inconsistent human judgment is lacking in scientific precision. This dependence on subjective judgment makes SFR studies ripe for cherry picking because researchers can consciously or unconsciously tilt their subjective judgments toward decisions that yield favorable results.

Dr. Jackson characterizes SFRs as roughly equivalent to a randomized experiment, describing them as “exogenous policy shocks” (Jackson, Exhibit C, p. 5 and Exhibit E, p. 6) that would isolate the causal effects of additional spending on student outcomes because the exact timing and location of those events cannot be anticipated by policymakers and educators. But the researchers who study these SFRs cannot even agree on where or when those events took place. If researchers differ so much on what constitutes an SFR and when it took place, it is unreasonable to think that these disputed events somehow consistently reveal the causal effects of extra funding. The SFR researchers cannot agree on where or when SFRs took place, so it is unreasonable for Dr. Jackson to be so confident that wherever or whenever these things occurred, they must have improved student outcomes.

Exhibit 5 – Screenshot of Portion of Online Appendix E from Lafortune, Rothstein, & Schanzenbach (2018) Listing Discrepancies in the Identification of School Finance Reforms Between Lafortune, Rothstein, & Schanzenbach (2018) and Jackson, Johnson, & Persico (2016)

The states and years for which the two tabulations disagree are:

- Alabama, 1993
- Arizona, 2007
- Connecticut, 1995 & 2010
- Idaho, 1993 & 1998
- Maryland, 1996 & 2005
- Michigan, 1997
- Montana, 1993 & 2008
- New Hampshire, 2006
- New Jersey, 1991, 1998 & 2000
- New Mexico, 1998 & 1999
- Oregon, 2009
- South Carolina, 2005
- Texas, 2004
- Washington, 1991, 2007 & 2010

This includes only cases in scope for both lists but coded differently. This in particular means that we do not discuss our tabulation of legislative school finance reforms, as these are out of JJP's scope. For each state, we discuss only the events where the two tabulations disagree; see Online Appendix Table A1 for a full listing of events in each state.

E) Dr. Jackson's claim that "increased spending leads to higher test scores, higher educational attainment, and higher wages, particularly for low-income students" (p. 4) is not consistently supported by his own evidence. Dr. Jackson reports different findings about whether increased spending has significantly greater benefits for low-income students in the two different versions of the meta-analysis he has appended to his expert report. In the November 2020 version (Exhibit C) he writes that "For test scores, we find little evidence that the average reported effect for low-income children differs from that of high-income children." (p. 23) Dr. Jackson comments on how regression estimates for low income students are positive but not statistically significant, and he emphasizes "It is important to note that these estimated effects are reported with error so that the difference between high- and low-income children is not precisely estimated." (p. 23) None of the coefficients for low-income students he reports in Table 4 of Exhibit C are statistically significant for test scores or educational attainment.

In the March 2021 version of the meta-analysis Dr. Jackson appends to his report (Exhibit E), however, Dr. Jackson claims to find a statistically significant difference between effects for low and high income students: "The formal test of whether the marginal impacts are the same for the low- and non-low-income groups and no different from zero yields a p-value of 0.0497 – indicating that we can reject the null (at the 5% level) that the coefficients on the income-level indicators are equal to zero." (p. 29) The result that was statistically insignificant in the November 2020 version became statistically significant in the March 2021 version. Apparently Dr. Jackson made changes over those five months to get the p-value three ten-thousands below the threshold needed to be described as statistically significant. Trying different analyses until the result becomes statistically significant is precisely what p-hacking looks like.

F) Dr. Jackson's analysis of the effects of school spending cuts during the Great Recession depends on the claim that states with the greatest reliance on state revenue for education spending made the greatest cuts during the downturn. As Dr. Jackson puts it, "The share of revenue coming from state appropriations is central to our empirical strategy." (Jackson, Exhibit D, p. 6) To determine the extent to which education spending in each state comes from state sources of revenue, Dr. Jackson relies upon data from the U.S. Census, saying, "School finance data come from the Annual Survey of School System Finances at the US Census Bureau." (Jackson, Exhibit D, p. 6) Claiming to rely on this data source, Dr. Jackson reports that "The share of funding that comes from state sources varies between 0 percent (Washington, DC) and 87 percent (Hawaii)." (Jackson, Exhibit D, p. 6)

Contrary to Dr. Jackson's claim, the U.S. Census does not report that Washington, D.C. has "0 percent" reliance on state sources of revenue for its education spending. Instead, the Census clearly describes D.C.'s revenue from state sources for education spending as "Not applicable." (See Exhibit 6) Even if Dr. Jackson were to have taken his data on the state share of education spending from the U.S. Department of Education's National Center for Education Statistics (NCES), which is another reputable source for this kind of information, he would have seen that D.C.'s education revenue from state sources is also described as "Not applicable." (See Exhibit 7)

Assigning "0 percent" to D.C. for its state share of education spending when data sources describe it as "Not applicable" is improperly assigning a value that is not drawn from the source and is highly inappropriate. Instead, Dr. Jackson should not have altered the information from his data source and should have listed D.C. as missing, which would exclude D.C. from the analyses. While Dr. Jackson does occasionally drop D.C. from his analysis as a robustness check, improperly assigning 0% for D.C. and including D.C. in most analyses masks the extent to which the inclusion of D.C. makes his analyses sensitive to other data and research design choices. As will be seen in Key Claim 2, his claims about the harms of cutting education spending are highly dependent on inaccurately assigning the value of 0 to D.C. for its state share of revenue along with other very specific data choices.

Exhibit 6 – U.S. Census Table on State Sources of Education Revenue

Table 3.
Revenue From State Sources for Public Elementary-Secondary School Systems by State: Fiscal Year 2013

(In thousands of dollars. Detail may not add to total because of rounding. For meaning of abbreviations and symbols, see footnotes)

Geographic area	Total	General formula assistance	Compensatory programs	Special education	Vocational programs	Transportation programs	Other and nonspecified state aid[1]	State payments on behalf of LEA
United States	272,916,892	184,361,517	5,550,350	18,169,015	986,126	4,183,883	45,599,077	14,066,924
Alabama	3,898,347	3,280,844	79,066	1,580	0	263,308	273,549	0
Alaska	1,707,449	1,152,263	0	0	0	73,541	154,810	326,835
Arizona	2,934,165	2,849,321	18,066	0	9,474	0	57,304	0
Arkansas	3,847,045	1,905,348	196,768	265,381	18,290	0	1,315,755	145,503
California	35,141,208	19,141,429	1,171,055	2,905,093	4,114	482,131	10,183,454	1,253,932
Colorado	3,693,829	3,319,183	125	133,441	28,906	52,834	159,340	0
Connecticut	3,870,444	1,518,871	107,055	520,392	6,750	70,095	829,753	817,528
Delaware	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
District of Columbia	X	X	X	X	X	X	X	X

X Not applicable.
[1] Includes staff improvement programs, bilingual education programs, gifted and talented programs, school lunch programs, capital outlay and debt service programs, nonspecified, and all other revenue from state sources.
Note: See Appendix B for a description of state-specific reporting anomalies. Due to the varying content of individual state aid programs, this information should not be used to compare the fiscal commitments of the states to the objectives of the specific programs shown in this table. Annual Survey of School System Finances statistics include the finances of charter schools whose charters are held directly by a government or a government agency. Charter schools whose charters are held by nongovernmental entities are deemed to be out of scope for the Annual Survey of School System Finances.
Source: U.S. Census Bureau, 2013 Annual Survey of School System Finances. Data are not subject to sampling error, but for information on nonsampling error and definitions, see introductory text. Data users who create their own estimates from these tables should cite the U.S. Census Bureau as the source of the original data only.

Exhibit 7 – NCES Table on State Sources of Education Revenue

Table 235.20. Revenues for public elementary and secondary schools, by source of funds and state or jurisdiction: 2012-13

[In current dollars]

State or jurisdiction	Total (in thousands)	Federal			State		Local (including intermediate sources below the state level)					
									Property taxes		Private[2]	
		Amount (in thousands)	Per pupil	Percent of total	Amount (in thousands)	Percent of total	Amount (in thousands)[1]	Percent of total	Amount (in thousands)	Percent of total	Amount (in thousands)	Percent of total
1	2	3	4	5	6	7	8	9	10	11	12	13
United States	**$603,686,987**	**$55,862,552**	**$1,128**	**9.3**	**$273,101,724**	**45.2**	**$274,722,710**	**45.5**	**$221,961,238**	**36.8**	**$11,620,194**	**1.9**
Alabama	7,188,210	850,523	1,142	11.8	3,936,486	54.8	2,401,201	33.4	1,099,433	15.3	321,517	4.5
Alaska	2,670,758	324,045	2,464	12.1	1,830,051	68.5	516,661	19.3	299,136	11.2	19,536	0.7
Arizona	9,385,733	1,278,835	1,174	13.6	3,965,426	42.2	4,141,471	44.1	3,013,964	32.1	228,247	2.4
Arkansas	5,051,804	612,256	1,259	12.1	2,624,126	51.9	1,815,421	35.9	1,568,431	31.0	148,938	2.9
California	66,026,445	7,388,302	1,173	11.2	35,878,654	54.3	22,759,489	34.5	18,324,216	27.8	402,845	0.6
Colorado	8,905,156	702,772	814	7.9	3,765,335	42.3	4,437,048	49.8	3,599,102	40.4	340,019	3.8
Connecticut	10,549,973	461,506	838	4.4	4,163,960	39.5	5,924,506	56.2	5,757,213	54.6	109,174	1.0
Delaware	1,909,503	192,422	1,491	10.1	1,123,567	58.8	593,514	31.1	506,397	26.5	15,084	0.8
District of Columbia	2,094,445	200,097	2,628	9.6	†	†	1,894,347	90.4	611,079	29.2	10,909	0.5

—Not available.
†Not applicable.
#Rounds to zero.
[1] Includes other categories of revenue not separately shown.
[2] Includes revenues from gifts, and tuition and fees from patrons.
NOTE: Excludes revenues for state education agencies. Detail may not sum to totals because of rounding.
SOURCE: U.S. Department of Education, National Center for Education Statistics, Common Core of Data (CCD), "National Public Education Financial Survey," 2012-13. (This table was prepared August 2015.)

G) In addition to improperly assigning a result for D.C., Dr. Jackson's Great Recession analysis (Jackson, Exhibit D) is plagued by other discrepancies between the numbers in his data set and the Census data that he claims to use as his source. Dr. Jackson's numbers for state share of education spending do not correspond to those reported by the Census. Dr. Jackson's numbers also differ dramatically from the state share data reported by the U.S. Department of Education's National Center for Education Statistics. These discrepancies can be seen in Exhibit 8. Differences larger than 1 percentage point or where the value has been assigned (i.e., D.C.) are highlighted in yellow. There are 13 discrepancies larger than 1 percentage point between the numbers in Dr. Jackson's data set and the numbers reported by the U.S. Census, which is his source. There are 26 discrepancies this large between Dr. Jackson's numbers and those reported by NCES.

In addition to his reliance on apparently erroneous data, the fact that the Census and NCES numbers differ significantly from each other raises questions about which data source Dr. Jackson should rely upon for his analyses. Both the Census and NCES are highly reputable sources, but they apply different standards to how revenues are classified when calculating what funds are considered as coming from state or local revenue. As will be seen below in Key Claim 2, Dr. Jackson's results in the Great Recession analysis are highly sensitive to which data source he selects. The fact that Dr. Jackson selects the data source that yields more favorable results for his argument while not presenting the alternative results relying on another reputable data source severely undermines the reliability of his findings.

Exhibit 8 – Discrepancies in State Share Data, by Source

State	Jackson	Census	NCES	Jackson minus Census	Jackson minus NCES
Alabama	60.2%	60.2%	60.6%	0.0%	-0.4%
Alaska	64.9%	64.9%	66.3%	0.0%	-1.4%
Arizona	48.1%	48.5%	51.7%	-0.4%	-3.6%
Arkansas	75.7%	76.0%	56.7%	-0.3%	19.0%
California	57.9%	59.9%	61.3%	-2.0%	-3.4%
Colorado	42.1%	42.4%	42.2%	-0.3%	-0.1%
Connecticut	37.8%	38.5%	39.6%	-0.7%	-1.8%
Delaware	61.2%	63.0%	62.0%	-1.7%	-0.8%
District of Columbia	0.0%	N/A	N/A	N/A	N/A
Florida	39.4%	39.4%	38.8%	0.0%	0.6%
Georgia	45.1%	45.2%	45.4%	0.0%	-0.3%
Hawaii	84.8%	84.8%	84.8%	0.0%	0.0%
Idaho	65.5%	65.5%	67.1%	0.0%	-1.6%
Illinois	32.9%	33.8%	31.2%	-0.9%	1.7%
Indiana	47.3%	48.5%	53.5%	-1.2%	-6.2%
Iowa	44.8%	46.5%	46.5%	-1.7%	-1.7%
Kansas	58.4%	58.4%	57.5%	0.0%	0.9%
Kentucky	57.9%	57.9%	57.3%	0.0%	0.6%
Louisiana	43.6%	43.9%	44.8%	-0.3%	-1.3%
Maine	43.2%	44.5%	44.9%	-1.2%	-1.7%
Maryland	42.0%	42.0%	42.1%	0.0%	-0.1%
Massachusetts	41.8%	42.1%	41.9%	-0.3%	-0.1%
Michigan	54.6%	57.3%	57.5%	-2.6%	-2.9%
Minnesota	64.4%	65.8%	65.9%	-1.4%	-1.5%
Mississippi	53.7%	53.8%	54.5%	-0.1%	-0.7%
Missouri	40.8%	41.1%	33.3%	-0.3%	7.5%
Montana	49.0%	49.4%	49.7%	-0.3%	-0.7%
Nebraska	32.3%	33.0%	33.1%	-0.7%	-0.9%
Nevada	57.5%	57.5%	30.8%	0.0%	26.7%
New Hampshire	37.1%	38.6%	38.6%	-1.4%	-1.5%
New Jersey	40.0%	41.3%	42.1%	-1.3%	-2.1%
New Mexico	71.2%	71.2%	70.8%	0.0%	0.4%
New York	45.2%	45.4%	44.8%	-0.2%	0.4%
North Carolina	58.8%	58.8%	65.7%	0.0%	-6.8%
North Dakota	34.7%	36.1%	36.3%	-1.4%	-1.6%
Ohio	43.0%	44.1%	45.6%	-1.1%	-2.6%
Oklahoma	51.2%	51.2%	54.2%	0.0%	-3.1%
Oregon	52.4%	52.8%	52.3%	-0.4%	0.1%
Pennsylvania	34.3%	35.8%	36.5%	-1.5%	-2.3%
Rhode Island	38.5%	38.7%	39.9%	-0.2%	-1.4%
South Carolina	50.6%	50.7%	50.8%	-0.1%	-0.2%
South Dakota	33.1%	33.2%	33.9%	-0.1%	-0.8%
Tennessee	45.9%	46.1%	45.6%	-0.2%	0.3%
Texas	43.1%	43.2%	44.8%	-0.1%	-1.7%
Utah	56.3%	56.3%	56.7%	0.0%	-0.4%
Vermont	68.3%	88.5%	85.9%	-20.2%	-17.5%
Virginia	40.3%	41.0%	41.0%	-0.6%	-0.6%
Washington	61.9%	62.4%	62.5%	-0.5%	-0.6%
West Virginia	58.1%	58.1%	59.1%	0.0%	-1.1%
Wisconsin	49.2%	50.1%	50.0%	-0.9%	-0.9%
Wyoming	52.8%	52.9%	52.8%	-0.1%	0.0%

H) There are also multiple potential sources for data on per pupil spending for each state. The U.S. Department of Education's National Center for Education Statistics and the U.S. Census report this information aggregated at the state level from surveys they administer to education officials. Dr. Jackson does not use either of these sources. Instead, to compute state-level per pupil spending Dr. Jackson aggregates that information himself from the district-level surveys that the Census administered. The numbers Dr. Jackson uses for state-level per pupil spending do not correspond to those reported by NCES or by the Census. The discrepancies between the per pupil spending in Dr. Jackson's data set and the other potential sources can be seen in Exhibit 9, which shows the change in per pupil spending between 2007 and 2017 for each data source. Discrepancies between Dr. Jackson's numbers and those reported by other sources that exceed $250 per pupil are highlighted in yellow.

In addition to his reliance on apparently erroneous data, the fact that these multiple, reputable data sources differ significantly from each other raises questions about which data source Dr. Jackson should rely upon for his analyses. The different sources apply different standards to how per pupil spending is calculated. As will be seen below in Key Claim 2, Dr. Jackson's results in the Great Recession analysis are highly sensitive to which data source he selects. The fact that Dr. Jackson selects the data source that yields more favorable results for his argument, while not presenting the alternative results relying on other reputable sources, severely undermines the reliability of his findings.

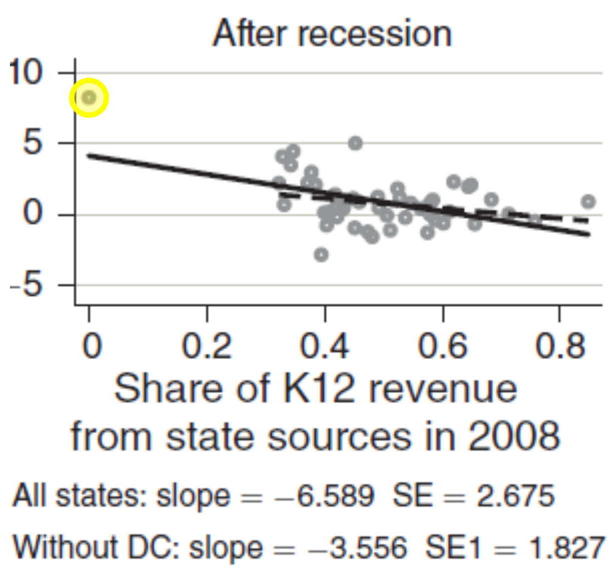
Exhibit 9 – 2007-2017 Change in Per Pupil Expenditure, by Source

	Jackson	Census Total PPE	NCES Total PPE
Alabama	($642)	($646)	($672)
Alaska	$2,083	$2,081	$2,156
Arizona	($1,599)	($1,604)	($1,038)
Arkansas	($504)	($44)	($117)
California	$749	$621	$822
Colorado	($198)	($127)	($246)
Connecticut	$2,985	$2,980	$2,518
Delaware	$177	$125	($159)
District of Columbia	$8,290	$8,346	$6,197
Florida	($2,872)	($2,873)	($2,525)
Georgia	($971)	($964)	($998)
Hawaii	$888	$887	$881
Idaho	($690)	($690)	($832)
Illinois	$4,109	$3,911	$3,870
Indiana	($1,252)	($1,184)	($728)
Iowa	$1,126	$1,079	$1,030
Kansas	$1,014	$1,029	$309
Kentucky	$11	$9	$411
Louisiana	$970	$513	$596
Maine	$475	$586	$956
Maryland	$521	$568	$472
Massachusetts	$1,404	$1,413	$2,314
Michigan	$780	($56)	($739)
Minnesota	$1,957	$2,278	$2,247
Mississippi	($239)	($240)	($16)
Missouri	$166	$241	($19)
Montana	$1,234	$1,228	$1,037
Nebraska	$2,204	$2,320	$1,445
Nevada	($1,303)	($1,318)	($1,107)
New Hampshire	$2,224	$2,218	$2,418
New Jersey	$119	($27)	($212)
New Mexico	$16	$27	($66)
New York	$5,045	$5,141	$4,142
North Carolina	($430)	($473)	($630)
North Dakota	$4,469	$4,649	$4,745
Ohio	$944	$957	$258
Oklahoma	($1,132)	($1,124)	($675)
Oregon	$1,822	$1,849	$1,441
Pennsylvania	$3,495	$3,898	$2,365
Rhode Island	$2,102	$2,140	$1,270
South Carolina	($91)	($112)	($110)
South Dakota	$674	$755	$673
Tennessee	$864	$927	$967
Texas	$380	$371	$335
Utah	$404	$397	$216
Vermont	$1,028	$2,245	$3,038
Virginia	($789)	($805)	($622)
Washington	$2,300	$2,289	$2,049
West Virginia	($30)	($35)	$774
Wisconsin	$472	$215	$240
Wyoming	$1,154	$1,092	$1,012

Key Claim 2 – Dr. Jackson's Great Recession analysis is not robust to reasonable changes in model specification, data source, or outcome selection. The sensitivity of his results to numerous, very specific choices renders the work unreliable for predicting the effects of changes in school spending in Schenectady and New York City.

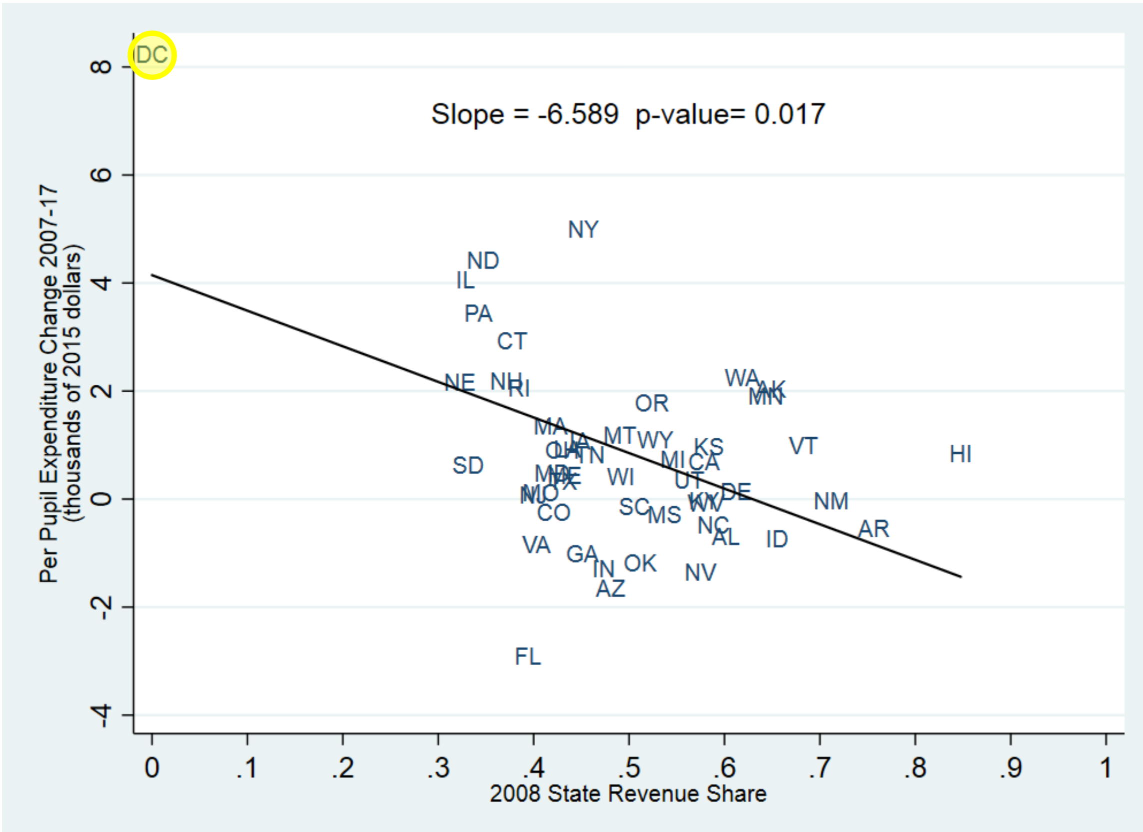
A) Dr. Jackson's claims about the harm done by education spending cuts rely upon the study he co-authored examining the effects of the decline in certain states in education spending on student outcomes, appended to his report in Exhibit D.[1] That analysis hinges on the claim that states that derive a greater share of their education funding from state revenue made, on average, significantly greater cuts in per pupil spending. As Dr. Jackson writes, "The pattern of larger spending cuts in states that were more reliant on state-appropriated revenues to fund public education motivates our instrumental variables approach." (Jackson, Exhibit D, p. 11) And as Dr. Jackson also emphasizes, "The share of revenue coming from state appropriations is central to our empirical strategy." (Jackson, Exhibit D, p. 6) To demonstrate the validity of this empirical strategy, Dr. Jackson presents a scatter plot in Figure 3 of Exhibit D that shows a significantly greater decline in per pupil spending between 2007 and 2017 for states that had a higher share of education spending coming from state sources of revenue. (See Exhibit 10) We replicate that figure in Exhibit 11, adding state abbreviation labels.

Exhibit 10 – Figure 3 Reproduced from Dr. Jackson's Exhibit D, with D.C. Highlighted



[1] The analyses presented in Key Claim 2 draw upon the work in Goldstein and McGee (2020).

Exhibit 11 – Reproduction of Figure 3 from Dr. Jackson's Exhibit D, with state abbreviations



That figure uses Dr. Jackson's data for state share and per pupil spending, which, as we have already demonstrated in Key Claim 1, differs from his cited sources and other reputable data sources. The slope of the line in Figure 3 is also clearly influenced by Dr. Jackson's decision to replace the "not applicable" designation for D.C.'s state share found in all reputable data sources with "0." Improperly assigning D.C. the value of 0 for state share of education spending places a data point in the far upper left corner of Figure 3 (highlighted in yellow), which drives the significant downward slope in the line through all of the state data points. Figure 3 also provides the slope estimate without D.C. Fixing this apparent data error cuts the estimated slope nearly in half and results in a statistically insignificant relationship between state revenue share and spending changes.

To more formally test whether the key empirical claim on which Dr. Jackson's Great Recession analyses rest is sensitive to corrections of his apparent data errors and source choices, I have constructed Exhibit 12. It shows what the slope of the line in his Figure 3 would be if we excluded D.C. from the analysis, corrected his apparent data errors, and drew data from different reputable sources for state share and per pupil spending information.

The rows represent different sources for data on the extent to which education spending comes from state revenue. The "Jackson Share" row contains results using the numbers in Dr. Jackson's data set, while the "Census Share" row contains results using the actual Census data Dr. Jackson claims as his source, and "NCES Share" contains results from the U.S. Department of Education's National Center for Education Statistics. The columns represent different sources for data on per pupil spending changes between 2007 and 2017. The "Jackson PPE" column contains results using the state spending numbers in Dr. Jackson's data set that he compiled by aggregating district-level responses to the Census survey. Dr. Jackson uses total expenditures which includes both operating and capital expenditures. The columns labeled "Census" provide results using the state-level data that the Census aggregated itself. And the columns labeled "NCES" contains results using the state-level per pupil spending numbers collected by the U.S. Department of Education's National Center for Education Statistics. For both Census and NCES, I provide results using both total expenditures and current expenditures (i.e., only operating expenditures).

Exhibit 12 – Sensitivity of the Slope in Dr. Jackson's Figure 3 to Data Source

Including Washington D.C.					
	Jackson PPE	**Census Total PPE**	**NCES Total PPE**	**Census Current PPE**	**NCES Current PPE**
Jackson Share	-6.5890**	-6.3444**	-4.6433**	-3.7716**	-3.4885*
	(1.7102)	(1.7733)	(1.5698)	(1.3124)	(1.3830)
Census Share	-5.8027**	-5.3714**	-3.6876*	-3.0495*	-2.6099
	(1.6555)	(1.7259)	(1.5281)	(1.2731)	(1.3434)
NCES Share	-5.3498**	-5.0271**	-3.3779*	-2.9502*	-2.5377
	(1.6939)	(1.7546)	(1.5489)	(1.2831)	(1.3516)
Excluding Washington D.C.					
	Jackson PPE	**Census Total PPE**	**NCES Total PPE**	**Census Current PPE**	**NCES Current PPE**
Jackson Share	-3.5563	-3.1987	-2.3802	-2.2393	-2.4153
	(1.8274)	(1.8947)	(1.7379)	(1.4859)	(1.6019)
Census Share	-2.8697	-2.2743	-1.3760	-1.4635	-1.3816
	(1.7302)	(1.7982)	(1.6485)	(1.4096)	(1.5236)
NCES Share	-2.2759	-1.8273	-0.9737	-1.3352	-1.2874
	(1.7507)	(1.8110)	(1.6565)	(1.4139)	(1.5272)

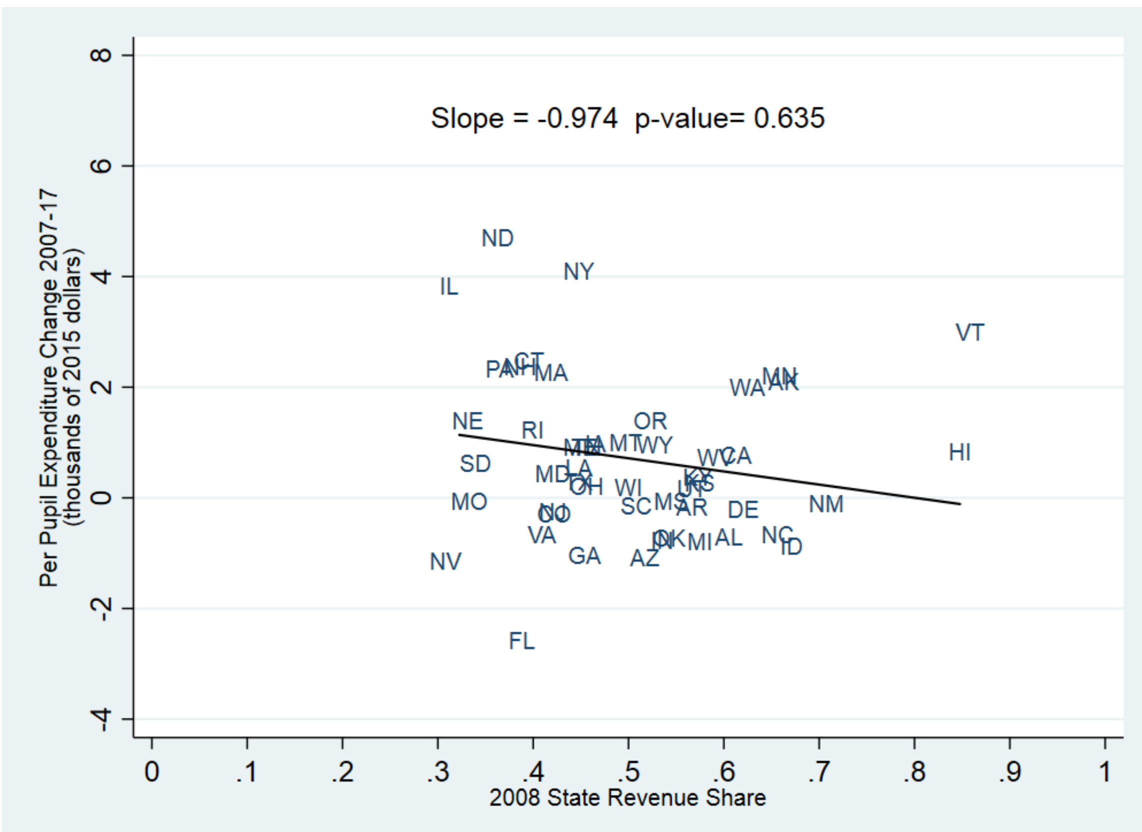
Standard errors in brackets
** p<0.01, * p<0.05

As Exhibit 12 clearly shows, Dr. Jackson's pivotal claim that per pupil spending significantly declines between 2007 and 2017 in states with greater reliance on state revenue is highly sensitive to his data choices. In fact, if D.C. is excluded from the analysis because its state share information is "not applicable," as both the Census and NCES indicate, the relationship between state share and spending is not statistically significant in any of the 15 possible combinations of data sources. Even if we consider all of the 30 permutations of data choices, the relationship between the share of education spending from state revenue and the change in per pupil spending between

2007 and 2017 is not statistically significant in the majority (17) of analyses, with the statistically insignificant relationships highlighted in yellow.

To illustrate how much Dr. Jackson's results change when using different reputable data sources, see Exhibit 13. It produces the comparison found in Dr. Jackson's Figure 3, but uses data from the U.S. Department of Education's National Center for Education Statistics (NCES). And since NCES lists Washington, D.C. as "not applicable," Exhibit 13 excludes D.C. Rather than showing a statistically significant decline of more than $650 in per pupil spending between 2007 and 2017 for every additional 10 percentage points in reliance on state revenue, as found in Dr. Jackson's Figure 3, Exhibit 13 shows no statistically significant relationship between changes in per pupil spending and state share.

Exhibit 13 – Replication of Dr. Jackson's Figure 3 with NCES Total PPE Data



If there were a real relationship between state share and spending cuts during the Great Recession, it should be manifest regardless of which reputable data source one chooses. The relationship Dr. Jackson claims to find, which is the premise of his entire study, only appears if he makes very specific and irregular data choices. Most importantly, Dr. Jackson has to assign the value of "0" to D.C. for its share of education spending coming from state revenue, despite his and other data sources stating that the information is "not applicable." The fact that there ceases to be a relationship between state share and spending cuts during the Great Recession if Dr. Jackson's apparent errors are corrected and if alternative data sources are used renders his analyses unreliable for the purposes of determining whether potential spending cuts in Schenectady and New York City would affect student outcomes.

B) The very specific and irregular data choices Dr. Jackson makes do not only affect the premise of the empirical strategy in his Great Recession study, they also determine the findings of that study. Exhibits 14 and 15 show the estimated effect of additional spending during the Great Recession on student outcomes given Dr. Jackson's analytical model but altering his data sources and treatment of D.C. Exhibit 14 shows effects on student test scores as measured by the National Assessment of Educational Progress (NAEP), while Exhibit 15 shows effects on college enrollment rates. Results that are not statistically significant are highlighted in yellow.

   If D.C. is excluded, as both the Census and NCES indicate it should be, most (23 out of 30) combinations of data sources would yield results that are not statistically significant. Even if we consider all permutations of data sources, including those with D.C., the majority (38 out of 60) of all analyses produce results that are not statistically significant. That is, Dr. Jackson's finding that spending cuts during the Great Recession harmed student outcomes is entirely dependent on what appear to be errors in his data and improperly assigning the value "0" for D.C.'s state share. Any independent researcher attempting to replicate Dr. Jackson's analysis, accepting his analytical approach, would be unlikely to find a statistically significant relationship between cuts in school spending and student outcomes during the Great Recession using publicly available and reputable data sources. The non-replicability of Dr. Jackson's findings render them uninformative when considering the effects of potential spending cuts in Schenectady and New York City on student outcomes.

Exhibit 14 – Replication of Dr. Jackson's Great Recession Test Score Results Using Alternative Data Sources

NAEP Including Washington D.C.					
	Jackson PPE	Census Total PPE	NCES Total PPE	Census Current PPE	NCES Current PPE
Jackson Share	0.0387**	0.0396**	0.0938	0.0524**	0.0768
	[0.0111]	[0.0113]	[0.0950]	[0.0150]	[0.0750]
Census Share	0.0470**	0.0484**	-0.0313	0.0629**	-0.0369
	[0.0096]	[0.0095]	[0.0579]	[0.0143]	[0.0568]
NCES Share	0.0338**	0.0342**	0.1484	0.0489**	0.1439
	[0.0118]	[0.0116]	[0.2737]	[0.0180]	[0.2828]
Excluding Washington D.C.					
	Jackson PPE	Census Total PPE	NCES Total PPE	Census Current PPE	NCES Current PPE
Jackson Share	0.0302*	0.0317*	0.0449	0.0431*	0.0553
	[0.0129]	[0.0139]	[0.0275]	[0.0213]	[0.0402]
Census Share	0.0362*	0.0389*	0.0544	0.0508	0.0414
	[0.0174]	[0.0191]	[0.0419]	[0.0308]	[0.0449]
NCES Share	0.0208	0.0224	0.0204	0.0315	0.0320
	[0.0147]	[0.0165]	[0.0157]	[0.0252]	[0.0286]

Robust standard errors in brackets
** p<0.01, * p<0.05

Exhibit 15 – Replication of Dr. Jackson's Great Recession College Enrollment Results Using Alternative Data Sources

College Enrollment Rate Including Washington D.C.					
	Jackson PPE	Census Total PPE	NCES Total PPE	Census Current PPE	NCES Current PPE
Jackson Share	0.0124**	0.0125**	0.0317	0.0187**	0.0307
	[0.0039]	[0.0037]	[0.0248]	[0.0062]	[0.0215]
Census Share	0.0138*	0.0138*	0.0189	0.0211*	0.0187
	[0.0055]	[0.0053]	[0.0180]	[0.0084]	[0.0161]
NCES Share	0.0263	0.0261	0.0892	0.0463	0.1186
	[0.0155]	[0.0152]	[0.1168]	[0.0315]	[0.1857]
Excluding Washington D.C.					
	Jackson PPE	Census Total PPE	NCES Total PPE	Census Current PPE	NCES Current PPE
Jackson Share	0.0100	0.0102*	0.0142	0.0135*	0.0180
	[0.0050]	[0.0050]	[0.0073]	[0.0060]	[0.0100]
Census Share	0.0111	0.0114	0.0199	0.0151	0.0213
	[0.0084]	[0.0085]	[0.0156]	[0.0095]	[0.0162]
NCES Share	0.0418	0.0409	0.0461	0.0523	0.0665
	[0.0400]	[0.0372]	[0.0435]	[0.0486]	[0.0780]

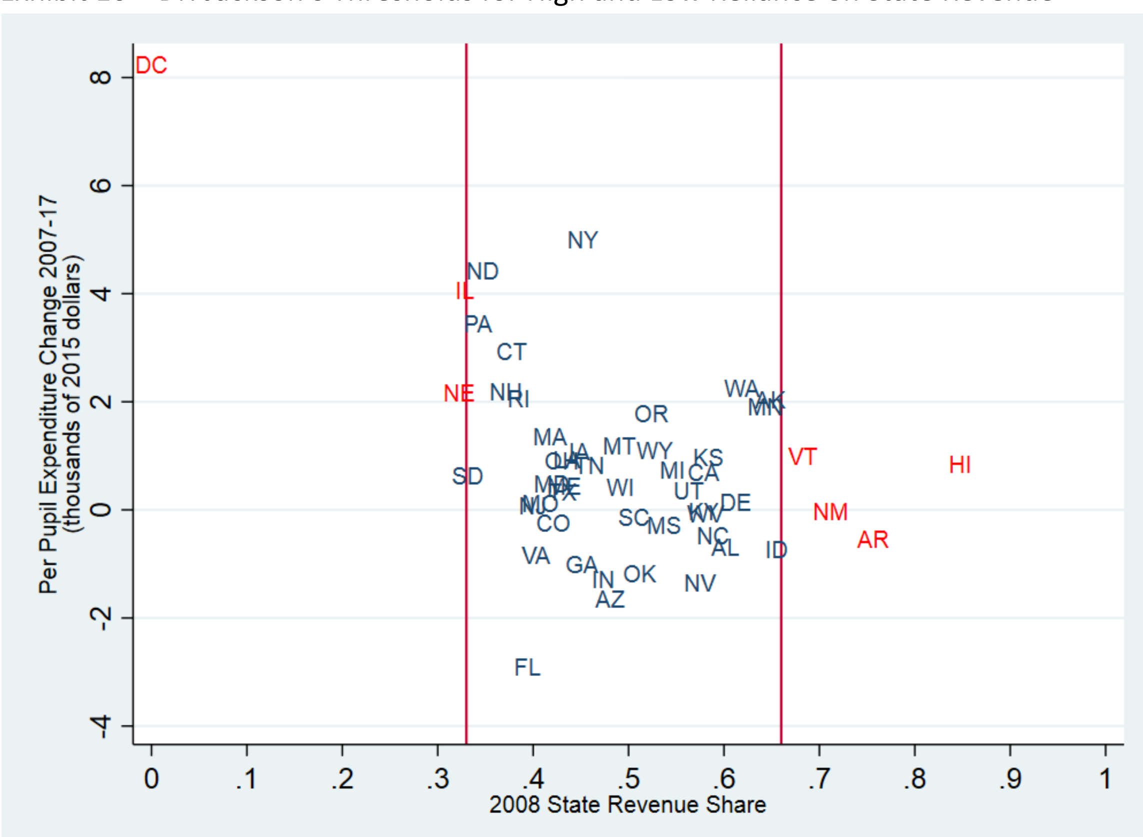
Robust standard errors in brackets
** p<0.01, * p<0.05

C) Even if one were to accept Dr. Jackson's data in the Great Recession study, including its apparent errors and inclusion of D.C., his results are still highly sensitive to how he constructs the statistical model for those analyses. In particular, Dr. Jackson makes the highly unusual choice of basing his analyses on a comparison of outlier cases. That is, Dr. Jackson's Great Recession study focuses on comparing states with low reliance for education spending on state revenue relative to states with high reliance on state revenue. Dr. Jackson arbitrarily defines lower reliance as those with less than 33% of their education spending coming from state revenue and those with high reliance as those with more than 66% of their education spending coming from state revenue.

Dividing states into low- and high-reliance groups in this way means that Dr. Jackson is effectively basing his conclusion on comparing only seven states, three in the low category and four in the high category. Exhibit 16 illustrates where Dr. Jackson sets the thresholds for low- and high-reliance and marks the state abbreviations in red that are included in those categories. The thresholds he uses fall right in the middle of clusters of states, artificially dividing them into the pivotal categories or not.

Dr. Jackson inaccurately describes his classification scheme as being in thirds of state share: "Schools that have one-third or less of their revenues from state sources are in the low group (g = 1), those with between one- and two-thirds are in the middle group (g = 2), and those that have more than two-thirds of their revenues from state sources are in the high group (g = 3)." (Exhibit D, p. 14) If he had done this, the thresholds would have been at 33.33% and 66.67% and South Dakota would have been included in the low reliance category. Moving the thresholds for classifying states as low- and high-reliance, even slightly, substantially changes which states form the basis of the comparison in his Great Recession study.

Exhibit 16 – Dr. Jackson's Thresholds for High and Low Reliance on State Revenue



To test the sensitivity of Dr. Jackson's conclusions to how he classifies states as having high or low reliance on state revenue, I examined what his results would be if he used slightly different thresholds for dividing states into high and low reliance categories. As can be seen in Exhibits 17 and 18, almost any other reasonable way of dividing states into low and high reliance categories would yield null results. Dr. Jackson is only able to obtain his finding that spending cuts during the Great Recession harmed student outcomes by drawing the lines separating states into categories in a very particular way.

I include results in Exhibits 17 and 18 that use the thresholds of one-thirds and two-thirds to categorize states as described in Dr. Jackson's text. This change only alters one state's categorization, adding South Dakota to the low-reliance group. However, this very minor change results in statistically insignificant results in most analyses (15 out of 20).

The alternative ways of categorizing states in Exhibits 17 and 18 are based on rank ordering all of the states in terms of their reliance on state revenue for education spending, even if we accept Dr. Jackson's apparently faulty data. We could then define high- and low-reliance as the bottom 10% of states versus the top 10%, the bottom 15% against the top 15%, and so on. None of the alternative ways of classifying states that I present would generate a statistically significant relationship between changes in school spending and changes in either NAEP test scores or college enrollment. The statistically insignificant results are highlighted in yellow. For a list of which states would be placed in the high and low reliance on state revenue categories for each of these different comparisons, see Appendix Table 1.

Picking a particular model for comparison that yields positive results while slightly different comparisons would yield null results is known as specification shopping. Just like choosing data sources, inclusion criteria, or outcomes that are favorable to one's argument, selecting a favorable way to analyze those data is characteristic of advocacy efforts rather than scientific research. Cherry-picking or specification shopping is highly inappropriate and fundamentally inconsistent with generally accepted scientific principles.

Exhibit 17 – Test Score Results of Dr. Jackson's Great Recession Study with Alternative Thresholds for Comparing States with High and Low Reliance

NAEP Including Washington D.C.					
	Jackson PPE	**Census Total PPE**	**NCES Total PPE**	**Census Current PPE**	**NCES Current PPE**
[0.33, 0.66]	0.0387** [0.0111]	0.0396** [0.0113]	0.0938 [0.0950]	0.0524** [0.0150]	0.0768 [0.0750]
[1/3, 2/3]	0.0357** [0.0111]	0.0365** [0.0113]	0.0890 [0.0875]	0.0491** [0.0149]	0.0775 [0.0750]
[10pctl, 90pctl]	0.0254 [0.0190]	0.0229 [0.0196]	0.0787 [0.0990]	0.0339 [0.0258]	0.1042 [0.1466]
[15pctl, 85pctl]	0.0106 [0.0242]	0.0087 [0.0227]	0.0198 [0.0678]	0.0170 [0.0320]	0.0529 [0.1165]
[20pctl, 80pctl]	0.0275 [0.0322]	0.0222 [0.0303]	0.1353 [0.3023]	0.0311 [0.0429]	0.1964 [0.5276]
[25pctl, 75pctl]	0.0426 [0.0419]	0.0334 [0.0378]	0.3111 [0.8720]	0.0427 [0.0556]	0.2785 [0.6590]
Excluding Washington D.C.					
	Jackson PPE	**Census Total PPE**	**NCES Total PPE**	**Census Current PPE**	**NCES Current PPE**
[0.33, 0.66]	0.0302* [0.0129]	0.0317* [0.0139]	0.0449 [0.0275]	0.0431* [0.0213]	0.0553 [0.0402]
[1/3, 2/3]	0.0265* [0.0124]	0.0278* [0.0133]	0.0394 [0.0247]	0.0389 [0.0205]	0.0494 [0.0368]
[10pctl, 90pctl]	-0.0030 [0.0219]	-0.0045 [0.0210]	-0.0066 [0.0298]	-0.0034 [0.0314]	-0.0034 [0.0362]
[15pctl, 85pctl]	-0.0005 [0.0286]	-0.0023 [0.0250]	-0.0048 [0.0365]	0.0023 [0.0360]	0.0018 [0.0461]
[20pctl, 80pctl]	0.0441 [0.0661]	0.0240 [0.0583]	0.1173 [0.1757]	0.0462 [0.1044]	0.1894 [0.3525]
[25pctl, 75pctl]	0.0501 [0.0748]	0.0288 [0.0594]	0.1500 [0.3325]	0.0385 [0.0999]	0.2250 [0.4985]

Robust standard errors in brackets
** p<0.01, * p<0.05

Exhibit 18 – College Enrollment Results of Dr. Jackson's Great Recession Study with Alternative Thresholds for Comparing States with High and Low Reliance

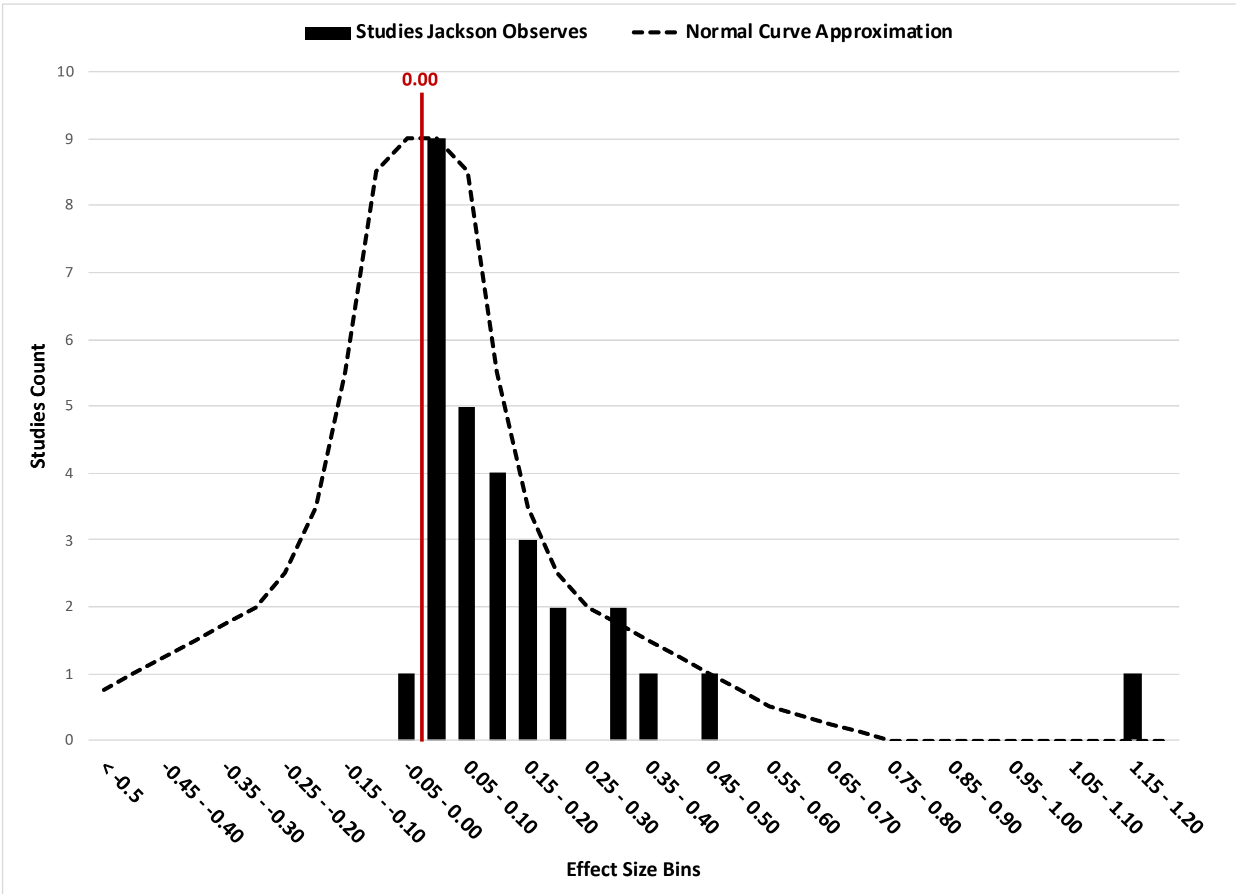
College Going Including Washington D.C.					
	Jackson PPE	Census Total PPE	NCES Total PPE	Census Current PPE	NCES Current PPE
[0.33, 0.66]	0.0124** [0.0039]	0.0125** [0.0037]	0.0317 [0.0248]	0.0187** [0.0062]	0.0307 [0.0215]
[1/3, 2/3]	0.0088 [0.0059]	0.0087 [0.0059]	0.0226 [0.0206]	0.0142 [0.0082]	0.0256 [0.0196]
[10pctl, 90pctl]	0.0112 [0.0066]	0.0098 [0.0057]	0.0252 [0.0210]	0.0171 [0.0090]	0.0312 [0.0247]
[15pctl, 85pctl]	0.0104 [0.0101]	0.0090 [0.0091]	0.0177 [0.0202]	0.0162 [0.0156]	0.0296 [0.0306]
[20pctl, 80pctl]	0.0064 [0.0158]	0.0053 [0.0144]	0.0125 [0.0341]	0.0066 [0.0236]	0.0113 [0.0524]
[25pctl, 75pctl]	0.0021 [0.0205]	0.0006 [0.0187]	0.0018 [0.0654]	-0.0046 [0.0351]	0.0233 [0.1535]
Excluding Washington D.C.					
	Jackson PPE	Census Total PPE	NCES Total PPE	Census Current PPE	NCES Current PPE
[0.33, 0.66]	0.0100 [0.0050]	0.0102* [0.0050]	0.0142 [0.0073]	0.0135* [0.0060]	0.0180 [0.0100]
[1/3, 2/3]	0.0052 [0.0070]	0.0051 [0.0072]	0.0073 [0.0096]	0.0085 [0.0088]	0.0124 [0.0113]
[10pctl, 90pctl]	0.0140 [0.0080]	0.0128 [0.0071]	0.0170 [0.0097]	0.0200 [0.0104]	0.0219 [0.0119]
[15pctl, 85pctl]	0.0004 [0.0153]	-0.0003 [0.0141]	-0.0010 [0.0186]	0.0004 [0.0202]	-0.0007 [0.0249]
[20pctl, 80pctl]	0.0051 [0.0238]	0.0015 [0.0234]	0.0168 [0.0364]	-0.0026 [0.0469]	0.0500 [0.1972]
[25pctl, 75pctl]	-0.0035 [0.0311]	-0.0063 [0.0282]	-0.0058 [0.0670]	-0.0185 [0.0530]	-0.0118 [0.1601]

Robust standard errors in brackets
** p<0.01, * p<0.05

Key Claim 3 – Dr. Jackson's reviews of the research, or meta-analyses, on the relationship between additional school spending and student outcomes are not based on a complete and unbiased set of studies. They are therefore uninformative regarding how possible education spending changes might affect student outcomes in Schenectady and New York City.

A) A complete and unbiased set of studies would tend to have results that were symmetrically distributed around the mean estimated effect. The set of studies Dr. Jackson includes has a distribution of results that is starkly asymmetrical, with a highly suspicious cliff of findings on the positive side of 0. (See Exhibit 19) The rising number of positive studies as effects approach zero and the near-total absence of negative findings strongly indicates that studies are missing from Dr. Jackson's meta-analysis and that the results are biased in a positive direction. As Brodeur, et al. (2016) wrote in the leading economics journal, *American Economic Review*, "If the underlying distribution of test statistics (for any method) is continuous and infinitely differentiable, any surplus of outcomes just above a threshold is taken as evidence of publication bias or p-hacking." (p. 2)

Exhibit 19 – The Asymmetrical Distribution of Results in Dr. Jackson's Meta-Analysis



To compile the histogram in Exhibit 19, the effect size for each study Dr. Jackson included in his Table 1 (Jackson, Exhibit E, p. 9) summary of the research was placed into bins in intervals of .05 effect sizes. If Dr. Jackson reported more than one result for a study, the average of results for those studies was used. Among the set of studies Dr. Jackson includes in his meta-analysis, the greatest number (nine) had effect sizes between 0 and .05, which is very small. These studies have positive results, but they are barely positive. The next highest number of studies (five) had results between .05 and .10. These 15 studies with barely positive results constitute the majority of all studies Dr. Jackson considers for his meta-analysis. If the bulk of Dr. Jackson's studies have results barely above zero, one would expect that even by chance as many studies should have results lower than this modal point as above it. But the set of studies Dr. Jackson considers has only 1 study below the modal point, while 19 are above it. If the results were symmetrical, we would expect the number above and below the modal point to be roughly equal.

Dr. Jackson is aware of this concern. To test for the possibility that his set of studies is distorted by publication bias, he uses Egger's test, which measures the extent to which study results are symmetrically distributed, with the variance in results decreasing as the estimates become more precise. Based on Egger's test, Dr. Jackson admits, "we do find evidence of some asymmetry." (Jackson, Exhibit C, p. 24) But Dr. Jackson then makes two different kinds of adjustments to conclude that "the impacts of any potential publication bias on our estimates are small." But as Dr. Jackson acknowledges, these adjustments "assum[e] that the source of the asymmetry is publication bias," (Jackson, Exhibit C, p. 24) which is only one type of bias that could produce such an asymmetrical distribution. Publication bias is the lower likelihood that studies with null results would be published and readily available for a meta-analysis than studies with statistically significant findings.

A different, and more likely, cause of the stark asymmetry in Dr. Jackson's meta-analysis is that he failed to consistently define the intervention his meta-analysis is examining and then conducted an inadequate search for studies that should have been included. Dr. Jackson claims that "we collected and classified all known credible causal studies of the impact of public school spending on student outcomes in the United States" (Jackson, Exhibit C, p. 25, and Exhibit E, p. 36) when in fact he only collected a limited and biased number of those studies. Studies that examine "the impact of public school spending" would include all rigorous evaluations of education policy interventions that require additional funds. There are likely many such studies, but Dr. Jackson only considers the set of studies that self-consciously include themselves as part of a literature on school spending.

That is, a group of researchers and activists have organized a body of research to advocate for increasing school spending. Other researchers, who are not part of this movement, also examine policy interventions that involve additional resources but they tend not to describe their research as being about school spending. Instead, they see themselves as studying the effects of things like schools purchasing technology, offering merit pay for teachers, implementing new curricula or pedagogies, or adopting certain school turnaround strategies -- among many possible education policy interventions that require additional funds. But there is no reason for Dr. Jackson to believe that a study only examines "the impact of public school spending" if the money is used to renovate school buildings instead of buying new technology. Nor is there any reason for him to include studies of school turnaround efforts if they are part of a literature advocates for additional school spending cite, while excluding other school turnaround studies that advocates tend not to mention. Dr. Jackson's limited and inconsistent inclusion of studies in his meta-analysis produced the highly skewed distribution of results we see in Exhibit 19.

The inadequate procedures Dr. Jackson used to search for studies to include in his meta-analysis enabled this problem. As he describes it, he began with a set of studies with which he was familiar and "then supplemented this with Google Scholar searches on relevant terms (school spending and causal)." (Jackson, Exhibit C, p. 6, and Exhibit E p. 7) Searching for "school spending" would miss studies that involve school spending but instead describe themselves as being about school technology, merit pay, curricular or pedagogical innovations, school turnaround strategies, etc. Dr. Jackson adds that he "consulted the cited papers and all papers that cite those on our list to identify other possible papers to include." (Exhibit C, p. 6) But given that researchers tend only to cite studies relevant to their specific issues, searching citations in school construction studies would not yield studies on merit pay.

Lastly, to find more studies, Dr. Jackson "asked active researchers in the field to locate any additional papers beyond the list we compiled" by "using a broad appeal on Twitter." (Jackson, Exhibit C, p. 6) But asking your followers on Twitter for study suggestions is not actually a "broad appeal," even if you have thousands of followers including some people with differing perspectives. Only people who are regularly active on Twitter would have a chance to see that request, even if they were followers. And the group of people who are likely to respond to such a request would be further skewed toward one's friends and others who share one's perspectives.

Doing a Google search on an incomplete set of terms and then asking one's friends on Twitter for suggestions is an improper method for identifying studies to include in a meta-analysis. According to Kugley, et al. (2017), accepted methods for identifying studies to include in meta-analyses "require a thorough, objective and reproducible search of a range of sources to identify as many relevant studies as possible," adding that "a search of one database alone is typically not considered adequate." When searching these multiple data sets, Kugley, et al. stress that researchers should "use a wide variety of search terms for each concept." Dr. Jackson's approach in his meta-analysis does not resemble these best practices. His process was not thorough, objective, or reproducible. He searched only one database, Google. And he only reports searching a limited set of terms, "school spending" and "causal." Dr. Jackson's method for identifying studies to include in his meta-analysis amounted to a closed loop of starting with a set of studies identified by advocates as being about school spending, searching citations within that limited literature, and then asking his like-minded friends for additional suggestions. This was a biased process that generated the skewed results we see in Exhibit 19.

B) The fact that Dr. Jackson's meta-analysis is incomplete and biased can also be illustrated by providing examples of studies that meet his criteria but were not included. In February of 2021, Eric Brunner, Ben Hoen, and Joshua Hyman released a study estimating the causal effects of additional school revenue generated by the exogenous event of wind farms being built in certain locations. Based on a large national sample with over 200,000 observations, they conclude: "The effects 5 years post are -0.037 and -0.026 with and without Texas, respectively, but again neither is statistically significant. Given the standard errors of 0.026 and 0.027, we can rule out positive impacts of approximately 2–3 percent of a SD. While these estimates do suggest a negative effect, we hesitate to interpret them as such given the imprecision of the estimates, and prefer to conservatively infer a lack of positive impacts. It is worth noting, however, that a negative impact on achievement is not entirely implausible." (p. 19) Their results show negative estimated effects of additional spending on student outcomes as long as 12 years after the extra revenue is provided. (See Appendix Figure VIII: Effects of Turbine Installation on Student Outcomes, 12 Years Out, p. 58)

Dr. Jackson was made aware of the existence of this study on February 19, 2021 on Twitter in response to his release of a working paper version of his meta-analysis. Dr. Jackson replied, "Will definitely update with this study." (Jackson, Twitter, 2021) But the Brunner, Hoen, and Hyman (2021) study is not included in either version of Dr. Jackson's meta-analysis appended to his expert report, including the one dated March 3, 2021. The fact that the Brunner et al. (2021) study has negative estimated effects would certainly change the results of Dr. Jackson's coin flip analysis so that it is no longer accurate for him to claim that there is a 1 in 4,320,893 chance that he is mistaken in his conclusions. Dr. Jackson knew that this figure was mistaken in the meta-analysis dated March 3, 2021 but kept it in the paper and submitted it with his expert report nonetheless.

Another example of a study missing from Dr. Jackson's meta-analysis can be found when Roland Fryer conducted an evaluation of a randomized experiment in which schools in New York City were provided with additional funds for a merit pay program. In 2013, Fryer found that this additional spending tended to have a negative effect on student achievement: "Providing incentives to teachers based on school's performance on metrics involving student achievement, improvement, and the learning environment did not increase student achievement in any statistically meaningful way. If anything, student achievement declined." (p. 377)

There is no good reason why Fryer's study is not included in Dr. Jackson's meta-analysis. It examines the effects of "policy variation in school spending," as Dr. Jackson requires. (Jackson, Exhibit C, p. 5) It evaluates those effects with an actual experiment, which more convincingly demonstrates causal effects than any of the quasi-experimental studies in Dr. Jackson's meta-analysis. The merit bonus Fryer examines was worth an average of $3,000 per school staff member, which is a significant amount of additional spending. And the study actually took place recently in New York, making it highly relevant to the current court proceedings. If New York schools were provided with additional funds as a result of this litigation, they might decide to use those funds to incentivize improved teaching with merit bonuses. According to Fryer's research, this would be more likely to result in harming student outcomes than improving them.

Another example of a study missing from Dr. Jackson's meta-analysis is Pham, et al.'s (2020) evaluation of school turnaround efforts in Tennessee. The policy intervention examined by Pham, et al. is part of the same national effort that Carlson and Lavertu (2018) studied in Ohio and that Dr. Jackson included in his meta-analysis. Unlike the positive results produced by Carlson and Lavertu, Pham, et al. find that after four years school turnaround efforts in Tennessee have negative but statistically insignificant results. (Pham, et al., Table 7, p. 13) And unlike Carlson and Lavertu, Pham, et al. is missing from Dr. Jackson's meta-analysis. Pham, et al. used a difference-in-difference research design, which is one of the methodologies Dr. Jackson considers causal, so it is unclear why Dr. Jackson would not include it in his meta-analysis.

Yet another example of a study missing from Dr. Jackson's meta-analysis is Balu, et al. (2015). It is an evaluation of a federally funded reading program called Response to Intervention. The program provided extra, small-group reading instruction to struggling students. Since there was a test score cut-off for eligibility for the program, the evaluation could use a regression discontinuity research design, which is one of the methodologies Dr. Jackson believes yields causal findings. As Balu, et al. describe their results: "Assignment to Tier 2 or Tier 3 intervention services in impact sample schools had a negative effect on performance on a comprehensive reading measure for first-graders just below the Tier 1 cut point on a screening test. The estimated effects on reading outcomes in Grades 2 and 3 are not statistically significant." (p. ES-13) According to this study, which was commissioned by the U.S. Department of Education, devoting additional resources to small group reading instruction, as schools in New York might do if provided with additional funds, may not improve student achievement.

Lastly, we could consider Lipsey, et al.'s (2018) evaluation of pre-kindergarten programs operated by public schools in Tennessee. Tennessee provided local school districts with additional funds to operate voluntary pre-kindergarten (VPK) programs, but because there was more demand than spots available, a lottery was used to determine which students would enroll. Using this randomized experiment, Lipsey, et al. found that by the end of 3rd grade, "the control children outperformed the VPK treatment children in all three subject areas." (p. 168) Despite the relevance of Lipsey, et al. and despite its rigorous examination of the effects of a likely use of additional school funding, this study is not included in Dr. Jackson's meta-analysis.

In addition to these several examples of studies missing from Dr. Jackson's meta-analysis, there are many more we could list if we were to conduct a proper and exhaustive search for research on the effects of additional spending on student outcomes. As these examples suggest, the effects of additional spending on student outcomes is mixed and complicated. Certain uses of additional funds might help student outcomes, while others might hurt or have no effect. It is further complicated by the fact that the same intervention might yield positive results in certain contexts, but produce null or negative results in others.

Dr. Jackson asserts in his report that his findings are "incompatible with the notion that 'money does not matter.'" (p. 10) Given the mixed and complicated literature on the effects of policy interventions that require additional spending, the question of whether "money matters" is a straw man that misses the point. (Jackson, Exhibit C, p. 21) It is nearly impossible to draw useful conclusions about whether additional spending, in general, causes favorable outcomes or not. A proper answer is almost always contingent on how the money is likely to be used, whether that use is appropriate for the given context, and how well implemented that use is. Even if Dr. Jackson's meta-analysis had been properly conducted, it would be inappropriate to extrapolate the findings to the current context in New York. But as we have also clearly demonstrated, Dr. Jackson's meta-analysis was not properly conducted and relies upon an incomplete and biased set of studies. Lastly, it is not useful to discuss whether additional money matters without also considering whether the existing allocation and uses of resources could be improved.

Key Claim 4 – Dr. Jackson's reviews of the research, or meta-analyses, focus on a set of studies that he claims "employ quasi-experimental methods to isolate causal impacts." (p. 7) While the research designs Dr. Jackson prioritizes can approximate the causal estimates derived from actual experiments if certain assumptions are strictly met, in the set of studies Dr. Jackson considers, those assumptions are routinely violated and their results should not be considered causal. In addition, Dr. Jackson's own analyses of spending and student outcomes in Schenectady and New York City do not even attempt to use credibly causal research designs.

In his meta-analysis, Dr. Jackson repeatedly emphasizes the importance of focusing on studies that he considers to be "credibly causal." (Jackson, Exhibit C, p. 2) According to Dr. Jackson, credible studies have to use one of a handful of research designs. By dismissing all studies that do not use these methods, Dr. Jackson is able to ignore a large and mixed research literature when drawing conclusions about what research says about the relationship between additional spending and student outcomes. Just because a study uses a particular research design that Dr. Jackson prefers, however, does not mean that its results are credibly causal. As will be discussed below, many of the studies Dr. Jackson deems credible and includes in his review fail to satisfy the assumptions required for their findings truly to be considered causal. If we should focus only on credibly causal studies when considering the effects of potential spending cuts in Schenectady and New York City, as Dr. Jackson argues, then we would need to dismiss many of the studies that inform his conclusions.

A) Studies Using Difference in Difference Event Study (Diff in Diff) or Instrumental Variable (IV) Research Designs

A large portion of the studies Dr. Jackson claims to be causal examine the effects of school finance reforms (SFRs). These studies employ one of two research designs that can produce results that approximate being causal: Difference in Difference Event Study (Diff in Diff) or Instrumental Variable (IV).

Essentially, the Diff in Diff studies compare the trend in an outcome before and after adoption of an SFR relative to the trend before and after that date in places where that event did not occur. As Dr. Jackson rightly noted in his Delaware report, to isolate something approximating the causal effect of additional spending on outcomes, this estimation strategy requires that "there were no other coincident policies or changes that occurred at the time of the event." (p. 34)

Those studies that employ the IV approach use the event of the SFR as an "instrument" to predict the resulting change in spending. They then examine the relationship between this predicted level of spending and student outcomes. This technique can produce results that approximate being causal as long as the SFR has no effect on student outcomes other than through how it alters school spending. Like with the Difference in Difference approach, this requires assuming that the SFR was not coincident with policy changes that could affect student outcomes in ways other than spending.

However, school finance reforms typically occur as part of reform packages that include multiple changes in how schools are governed and operated in addition to changing their funding. The contemporaneous changes in other policies affecting student outcomes violate the assumptions required for both Diff in Diff and IV analyses to be considered causal estimates of the relationship between spending and student outcomes.

Clark (2003) and Candelaria & Shores (2019), which are included in Dr. Jackson's meta-analyses, describe in detail how SFRs are often coincident with policy changes that extend beyond additional spending. Clark acknowledges that the reform she is studying is not simply about changes in overall spending levels: "The Kentucky Education Reform Act (KERA), implemented in 1990, is one of the most ambitious and influential education reform policies ever attempted by any state. KERA's main components include a new funding system to correct large financial disparities between school districts, curriculum revision and standardization, and increased school and district accountability." (p. 1) And Candelaria & Shores (2019) emphasize that this undermines the ability to draw causal conclusions: "the primary threat to validity is whether changes in graduation rates following court-ordered finance reform can be wholly attributable to changes in school spending. For the exclusion restriction to hold, we must assume that the court reforms affect graduation rates only through their effect on spending. This assumption is violated in cases where court-ordered finance affects other unobserved policy changes that also affect graduation rates." (pp. 52-53)

Lafortune, Rothstein, & Schanzenbach (2018), which is also included in Dr. Jackson's meta-analyses, also acknowledges this threat to drawing causal conclusions, but the authors dismiss it, "Some of the reforms were accompanied by governance, curriculum, or accountability changes, though our assessment is that these additional changes were typically not very important or impactful." (p. 6) These authors may think that these other policy changes were unimportant, but the inability to rule out that student outcomes were caused by other policy changes means that their analysis is not causal and is no different from the observational studies that Dr. Jackson declares are uninformative. In both cases, researchers assert that other unexamined factors probably do not matter, but they cannot conclude that with scientific confidence.

Another study featured in Dr. Jackson's meta-analyses and that he co-authored, Jackson, Johnson, & Persico (2016), considers the possibility that "recent policy reforms that started in the late 1980s (such as charter schools and test-based accountability)" (p. 205) confound their results. The authors then divide their sample to see if more recent cohorts of students, who may have also been exposed to these recent reforms, show different effects from SFRs than previous ones. The problem with this approach is that it assumes that coincident changes in school spending and other policy changes have not been an issue in both time periods. The multi-state studies of SFRs simply cannot be considered causal because the effects of school spending they are studying are inseparable from other school reforms adopted around the same time.

The three studies in Dr. Jackson's meta-analysis focused on the effect of an SFR in Michigan, called Proposal A, (Hyman (2017), Papke (2008), and Roy (2011)) face the same intractable difficulty with drawing causal conclusions. While Proposal A did change spending for certain school districts, it was contemporaneous with the adoption of charter schools and accountability measures that may have also affected those same districts. As researchers Julie Berry Cullen and Susanna Loeb put it, “In addition to the changes in raising and delivering funds to school districts, Proposal A also included new school choice measures and led to a new accountability system. It is likely inevitable that such dramatic increases in the state role in education are accompanied by increased state oversight and involvement.” (2004, p. 13) The inability of these studies to separate the effect of funding changes from the effect of these other manifestations of state oversight and involvement means that they are unable to generate causal estimates on the relationship between spending and student outcomes.

The benefits of accountability and other reforms that have occurred at the same time as SFRs are not merely hypothetical. There is a large body of evidence showing improvements in student outcomes from these reforms that are as large, or larger, than those produced by SFRs. For example, Dee & Jacob (2011) examine the effects of national accountability reforms adopted in No Child Left Behind (NCLB) and find: “Our results indicate that NCLB generated statistically significant increases in the average math performance of fourth graders (effect size = 0.23 by 2007) as well as improvements at the lower and top percentiles. There is also evidence of improvements in eighth-grade math achievement, particularly among traditionally low-achieving groups and at the lower percentiles.” (p. 418) Carnoy & Loeb (2002) conduct a study of the effects of state accountability reforms and find that “students in high-accountability states averaged significantly greater gains on the NAEP 8th-grade math test than students in states with little or no state measures to improve student performance.” (p. 305)

Research by Sass, Zimmer, Gill, & Booker (2016) examine charter schools in Florida and conclude that “students attending charter high schools are more likely to persist in college, and that in their mid-20s they experience higher earnings.” (p. 683) Using a randomized experiment, Angrist, Cohodes, Dynarski, Pathak, & Walters (2016) find that charter schools in Boston improve student performance on state, AP, and SAT tests, although the attainment benefits in college are less clear.

The fact that SFRs are often accompanied by other reforms and those other reforms have been demonstrated to significantly improve student outcomes means that it is quite plausible that some or all of the benefits Dr. Jackson attributes to additional spending from SFRs could actually be attributable to other contemporaneous policy changes. This distorts Dr. Jackson’s assessment of the relationship between additional spending and student outcomes, making his conclusions about the effects of spending changes in Schenectady and New York City unreliable.

B) Studies of School Tax Elections Using Regression Discontinuity (RD) Research Designs

Another common type of study that Dr. Jackson considers causal and includes in his meta-analyses involves the use of Regression Discontinuity (RD) in school tax elections. (Abott, Korgan, Lavertu, & Peskowitz (2019); Baron (2019); Cellini, Ferreira, & Rothstein (2010); Hong and Zimmer (2016); Kogan, Lavertu, & Peskowitz (2017); Lee & Polachek (2018); Lee & Polachek (2014); Martorell, Stange, & McFarlin (2016); and Rauscher (2019))   These studies compare student outcomes for districts where voters barely pass measures to increase school spending to districts where those measures barely fail. The argument for why this approximates causal effects is that whether a measure barely passes or barely fails can be thought of as essentially random. If the districts that barely pass measures later have better student outcomes, these researchers believe, it would have to be because they won the election and not any other pre-existing differences.

Whether school tax measures pass or not could affect student outcomes through paths other than increased spending by altering district leadership or by changing the mix of students in the district, as demonstrated by changing housing prices. If school tax elections alter outcomes in ways other than by changing spending, then these studies are unable to separate the effect of school spending from the effects of these other consequences of election outcomes.

In addition to these difficulties with drawing causal conclusions about the effects of increased spending on student outcomes, it is worth noting that these analyses routinely violate another assumption required to think of RD as approximating causal results. To believe that school districts with election outcomes barely above a passing threshold and those barely below are effectively randomly assigned, we would have to believe that school districts are unaware of how close they likely are to the cut-off and be unable to do anything to alter that outcome. But districts typically monitor the progress of their election campaigns through polling or by their informal sense of the community. And if they detect that measures are struggling, they can exert more effort and devote additional resources toward passage.

Losing a school tax election, even by a modest margin, is therefore likely to be associated with administrative incompetence, which is also very likely to be negatively associated with future student outcomes. As a Chamber of Commerce official commented following the defeat of a school tax increase in California, “I think the school district is mismanaging how they spend their money and mismanaging how they create a quality education for all their kids. [Before asking for more money] the district needs to get its house in order both fiscally and academically.” (Blume, 2019) The bias introduced by the fact that districts that lose elections tend to be less administratively competent is exacerbated by the fact that many of these RD studies do not restrict their samples only to elections that are very close to the threshold for winning.

C) Changing school funding is likely to change student composition in schools, which could also alter outcomes independently of the additional spending.

The student composition of school districts is likely to change as a result of school tax increases or other significant policy changes that raise school spending. Additional funding is likely to attract families that value higher educational spending. As Bayer et al. (2020) observe, "The sharp increase in house prices that accompanies an exogenous increase in school spending naturally affects who can afford and who is willing to pay to live in a school district…. That exogenous increases in school spending decrease the fraction of children in poverty within a district suggests that the house price effects documented above likely combine a direct effect of school spending and an indirect effect that results from the changing socioeconomic composition of the school district." (p. 27, 29) This means that in many of the studies in Dr. Jackson's meta-analyses, it is impossible to fully distinguish changes in student outcomes caused by additional spending from the changes that would occur from different student composition. It is true that many of these studies run analyses to see if schools saw changes in student demographics and free lunch status of their students following the influx of additional spending. But these analyses cannot observe or control for all dimensions on which student composition might change and therefore cannot rule out the confounding influences of changing student composition. This also means that these "causal" studies ultimately rely on observational correlations of a handful of student characteristics to draw their conclusions, just like the observational research literature that Dr. Jackson dismisses as uninformative.

D) In general, changes in school spending are almost never exogenous, making it very difficult to draw any causal conclusions with confidence.

SFR studies that use Diff in Diff or IV research designs and school tax election studies that use RD constitute the bulk of the studies in Dr. Jackson's meta-analyses. For the reasons described above, we should be dubious that these studies actually generate causal estimates of the relationship between school spending and student outcomes. The remaining studies in Dr. Jackson's meta-analysis are no more likely to be considered causal. The general problem is that in the real world we rarely have "natural experiments" in which school spending varies for reasons that are effectively random. That is, changes in school spending are almost never exogenous.

In his Delaware report, Dr. Jackson talks about studying SFRs as if they approximate experiments in which "the timing and location of the money drop is random." (p. 46) The term "money drop" is just rhetorical flourish, not a metaphor for any actual school spending process. In reality, the timing and location of money allocated to schools almost never approximate randomness. Even SFRs are political events that unfold over many years, and are shaped by the characteristics and academic trajectory of the affected schools. These processes are so slow-moving and complicated that even the researchers who study SFRs cannot agree on where or when SFRs have occurred. There is nothing magical about additional money generated by SFRs or other policy changes that makes studying those dollars causal while studying all other dollars allocated to schools uninformative.

E) Dr. Jackson invokes the authority of the What Works Clearinghouse (WWC) to support his selection of studies to examine for his meta-analysis, but the WWC standards are not consistent with those Dr. Jackson employs. None of the IV or Diff-in-Diff studies in Dr. Jackson's meta-analysis fully meet WWC standards. Those studies have quasi-experimental designs, or QEDs. As the WWC "Standard Handbook 4.0" that Dr. Jackson cites clearly states: "QEDs cannot receive the highest WWC rating, but can receive the rating Meets WWC Group Design Standards With Reservations." (p. 9)

In his report, Dr. Jackson wrongly claims that WWC draws a distinction in causal credibility between QED and observational studies, writing, "This practice of privileging quasi-experimental methods over observational studies is consistent with the Standards Handbook of the What Works Clearinghouse (WCC)…. Studies using randomized controlled trials, quasi-experimental design, regression discontinuity design, and single-case design satisfy the eligibility screens. Observational studies do not." (p. 8)

In fact, the term "observational studies" never appears in the WWC Handbook. Instead, the WWC considers QED and observational studies to be the same thing: "A study is eligible to be reviewed as a QED if it compares outcomes for subjects in an intervention group with outcomes for subjects in a comparison group, but does not rely on random assignment to determine membership in the two groups…. Assignment to the intervention may depend on both observed and unobserved characteristics…. The two groups may differ on characteristics researchers were able to measure, such as test scores, or on characteristics that researchers were not able to measure, such as motivation. Even with equivalence on measured characteristics, there may be differences in unmeasured characteristics that could introduce bias into an estimate of the effect of the intervention. Bias is a systematic difference between the true impact of the intervention and the estimated impact, which can lead to incorrect conclusions about the effect of the intervention." (p. 9)

WWC allows some QEDs to meet its standards with reservations if they meet certain quality criteria, such as similarity of comparison groups at baseline. WWC lists "acceptable approaches" for ensuring comparability of comparison groups, including "regression covariate adjustments in ordinary least squares models," "regression covariate adjustments in hierarchical linear models," and "simple gain scores… when the baseline and outcome measures are the same and have a strong relationship." (p. 16)

WWC never identifies IV or Diff-in-Diff as the only acceptable type of QEDs. Rather than there being a bright-line difference between IV, Diff-in-Diff, and other QED methods that Dr. Jackson calls "observational," WWC recognizes the reality that some IV and Diff-in-Diff studies are not credibly causal while other studies that Dr. Jackson calls "observational" might be informative, at least with reservations.

F) Dr. Jackson is also inconsistent in adhering to his methodological standard that "observational" studies are "unlikely to reflect causal impacts that are informative for policy." (Exhibit E, p. 2) To support the claim that students in Schenectady would be harmed by potential spending cuts, Dr. Jackson alleges that the district received less funding than it should have and this has contributed to weaker outcomes: "In short, in 2016-17, Schenectady was paid nearly $66M less in 2016 than the State itself calculated for an adequate education. That is approximately a $6,773 per pupil shortfall. As a result of this deficit, Schenectady's students performed poorly by any academic measure." (p. 18) Dr. Jackson's causal attribution of sub-par outcomes to alleged funding shortfalls is not based on any type of quasi-experimental analysis that would meet either WWC standards or those for inclusion in his meta-analysis.

   A few sentences after blaming sub-par outcomes in Schenectady on an alleged "shortfall" in funding without employing a credibly causal analysis, Dr. Jackson then credits improvements in Schenectady to its rising funding: "from 2015 to 2019, Schenectady received a higher percentage of the calculated Foundation aid… This increase led to a significant increase in educational attainment, test scores, and graduation rates for Schenectady." (p. 19) This attribution of improvements in student outcomes to changes in school spending were also made without a credibly causal analysis. As Dr. Jackson said in his Delaware report, "Observed correlations that are not driven by policy changes or policy variation, no matter how sophisticated, are unlikely to inform our understanding of the effect of increasing school spending on outcomes." (p. 36) Once again, Dr. Jackson is cherry-picking by selectively applying the standard that credibly causal research designs are required for a comparison to be informative.

Key Claim 5 – The particular context, including initial spending levels and political circumstances, differs dramatically between most of the studies Dr. Jackson includes in his review of the literature and the current context in Schenectady and New York City.

A) Studies showing the effects of additional funding when school spending is very low are unlikely to be applicable to Schenectady and New York City given the districts' already high level of spending. Both districts currently spend far more per pupil than most of the states and districts did when they were examined by studies in Dr. Jackson's meta-analysis. According to the Fiscal Accountability Summary (2018-19), total expenditures per pupil in Schenectady were $20,151 and in New York City were $27,732.

By contrast, most of the multi-state studies of the effects of SFRs include examining the effect of Tennessee's reforms in the early 1990s. According to the Digest of Education Statistics, in 1990 Tennessee had current school expenditures of $6,791, adjusted into 2018-19 dollars. Even if students in places like Tennessee benefited from increasing such low levels of spending, there is no reason to expect that New York students would receive comparable benefits by adding to their much higher level of funding.

B) The current political, social, and educational context in Schenectady and New York City often differs dramatically from the situations examined in Dr. Jackson's list of studies, making their results unlikely to be applicable to the district's circumstances.

In addition to examining the effects of spending at times and in places where the level of spending was much lower than in Schenectady and New York City, many of the studies in Dr. Jackson's meta-analysis examine the effects of additional spending in circumstances dramatically different from those currently found in the district. For example, Cascio, Gordon, & Reber (2013) look at the effect of Title I funding in the southern states in the 1960s. They found that racial politics were so fraught in the South at that time that Title I funding only improved outcomes for white students but made no difference for black students. The effects of school spending in the South in the 1960s is not applicable to what Schenectady and New York City could expect from increased spending today.

Martorell, Stange, & McFarlin (2016) examine the effects of school bond passage in mostly rural school districts in Texas. Schenectady and New York City are not rural and sparsely populated districts, so it is highly doubtful that what they learned in Texas would apply to these districts.

Neilson & Zimmerman (2014) study the effects of a school construction program in New Haven, where "schools reported problems with more than half of basic service systems, such as heating, air conditioning, plumbing, and lighting." If those are not similarly extensive problems in Schenectady and New York City, additional spending to repair basic service systems that are not broken are unlikely to yield the same effects as Neilson & Zimmerman claim from New Haven. Similarly, Lafortune & Schonholzer (2018) study the effects of a school construction program in Los Angeles where over-crowding was a serious issue. Again, if over-crowding is not a problem to the same extent in Schenectady and New York City, there is no reason to expect that additional spending would produce the same effects.

The context examined in most of the studies in Dr. Jackson's list differs so dramatically from the current context in Schenectady and New York City that extrapolating from those results to confidently predict what we should expect from additional spending in the district seems imprudent.

C) Dr. Jackson's Great Recession study in his Exhibit D is also unlikely to be applicable to circumstances in New York. According to Dr. Jackson's own data, New York actually increased per pupil spending by $2,142 per pupil between 2007 and 2013. So the negative effects Dr. Jackson claims to have found from spending cuts during the Great Recession were not generated by the situation in New York during that period. New York would be an outlier – an exception – to the pattern Dr. Jackson claims to have observed during the Great Recession. If New York's experience was exceptional during the Great Recession, it might be exceptional again during a future economic downturn.

D) In his report, Dr. Jackson does reference what he describes as "a credible study of New York State specifically" (p. 4) to support the assertion that his review of studies mostly about other places in the U.S. does apply to the New York context. That study by Gigliotti and Sorenson (2018), however, is not credibly causal. It compares changes in student outcomes in New York districts that experienced declines in student enrollment relative to those that experienced increases in enrollment. The logic of their comparison is that a "hold harmless" provision in the school funding formula means that districts with declining enrollment did not receive less state funding as a result, thereby increasing the per pupil expenditures for the students who remained. Districts with rising enrollment would not have experienced this net increase in per pupil funding.

The obvious difficulty with Gigliotti and Sorenson's comparison is that districts with declining enrollment may differ substantially in their demographic and economic changes, in both observed and unobserved ways, from districts with growing populations and enrollment. The authors acknowledge this: "One concern with our results is that large population losses in a local area might lead to shifts in the local macroeconomy, which could have resulting impacts on student achievement." (p. 176) To address this concern they add a "control for local unemployment rate" to their analysis and find that doing so has a "minimal effect on the estimated coefficients in our main model." (p. 176) Of course, controlling for the local unemployment rate does not fully adjust for all of the ways in which districts with rising and falling populations may differ. In addition, depending on adding a control for observed unemployment to draw their conclusion violates Dr. Jackson's standard that "observed correlations… are unlikely to inform our understanding of the effect of increasing school spending on outcomes." (Jackson's Delaware Report, p. 36)

While Dr. Jackson references Gigliotti and Sorenson (2018) as an example of research set in New York to support his argument, he omits discussion of other studies that were also about New York spending changes that produced negative results. In particular, another study found in his meta-analysis, Weinstein, et al. (2009), examined a Title I program in New York. That study, titled, "Does Title I Increase Spending and Improve Performance? Evidence from New York City" finds that increased funding actually lowered student test scores. A similar study by Van der Klaauw (2008) also examined Title I in New York and reached a similar conclusion. Dr. Jackson included the Van der Klaauw study in his 2018 literature review but it is not in his current meta-analysis. And Dr. Jackson never considered Fryer (2013), which also found negative results from a New York program to spend additional resources on merit pay. Once again, Dr. Jackson has multiple studies set in New York but cherry-picks one with a favorable result for his argument to mention in his report.

Lastly, if Dr. Jackson is going to invoke individual studies, rather than rely on the collective results of his meta-analysis, to support his argument about the likely effects of school spending changes in Schenectady and New York City, it is worth noting that over half of the studies for which he reports estimated effects do not have statistically significant results. In Table 1 of Dr. Jackson's Exhibit C, he provides the estimated effects and standard errors for 29 studies. According to his description of those results, only 14 of those studies have statistically significant results. That is, even most of the studies Dr. Jackson selects for his meta-analysis do not find a statistically significant relationship between changing school spending and student outcomes.

Appendix Table 1 – Different Groupings of States into High and Low State Share Categories

	Including Washington D.C.					
	Jackson	1/3 and 2/3	10th and 90th Percentiles	15th and 85th Percentiles	20th and 80th Percentiles	25th and 75th Percentiles
Low	(3 states)	(4 states)	(5 states)	(7 states)	(10 states)	(13 states)
	DC	DC	DC	DC	CT	CT
	IL	IL	IL	IL	DC	DC
	NE	NE	NE	ND	FL	FL
		SD	PA	NE	IL	IL
			SD	NH	ND	MO
				PA	NE	ND
				SD	NH	NE
					PA	NH
					RI	NJ
					SD	PA
						RI
						SD
						VA
High	(4 states)	(4 states)	(5 states)	(7 states)	(10 states)	(12 states)
	AR	AR	AR	AK	AK	AK
	HI	HI	HI	AR	AL	AL
	NM	NM	ID	HI	AR	AR
	VT	VT	NM	ID	DE	DE
			VT	MN	HI	HI
				NM	ID	ID
				VT	MN	KS
					NM	MN
					VT	NC
					WA	NM
						VT
						WA

	Excluding Washington D.C.					
	Jackson	1/3 and 2/3	10th and 90th Percentiles	15th and 85th Percentiles	20th and 80th Percentiles	25th and 75th Percentiles
Low	(2 states)	(3 states)	(5 states)	(7 states)	(11 states)	(12 states)
	IL	IL	IL	CT	CT	CT
	NE	NE	ND	IL	FL	FL
		SD	NE	ND	IL	IL
			PA	NE	ND	MO
			SD	NH	NE	ND
				PA	NH	NE
				SD	NJ	NH
					PA	NJ
					RI	PA
					SD	RI
					VA	SD
						VA
High	(4 states)	(4 states)	(4 states)	(7 states)	(10 states)	(12 states)
	AR	AR	AR	AK	AK	AK
	HI	HI	HI	AR	AL	AL
	NM	NM	NM	HI	AR	AR
	VT	VT	VT	ID	DE	DE
				MN	HI	HI
				NM	ID	ID
				VT	MN	KS
					NM	MN
					VT	NC
					WA	NM
						VT
						WA

Bibliography

Abott, C., Kogan, V., Lavertu, S., & Peskowitz, Z. (2020). School district operational spending and student outcomes: Evidence from tax elections in seven states. *Journal of Public Economics*, *183*, 104142.

Angrist, J.D., Cohodes, S.R., Dynarski, S.M., Pathak, P.A., & Walters, C.R. (2016) Stand and Deliver: Effects of Boston's Charter High Schools on College Preparation, Entry, and Choice. *Journal of Labor Economics,* 34(2) 275-318.

Balu, R., Zhu, P., Doolittle, F., Schiller, E., Jenkins, J., & Gersten, R. (2015) Evaluation of Response to Intervention Practices for Elementary School Reading. U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/pubs/20164000/pdf/20164000_es.pdf

Baron, E. J. (2019). School Spending and Student Outcomes: Evidence from Revenue Limit Elections in Wisconsin. *Available at SSRN 3430766*.

Bayer, P., Blair, P.Q., Whaley, K. (2020) A National Study of School Spending and House Prices. *NBER Working Paper*.

Biasi, B. (2019). *School Finance Equalization Increases Intergenerational Mobility: Evidence from a Simulated-Instruments Approach* (No. w25600). National Bureau of Economic Research.

Blume, H. (2019). L.A. school tax measure defeated, in blow to Garcetti and Beutner. *Los Angeles Times*, June 4. Retrieved from https://www.latimes.com/local/lanow/la-me-edu-lausd-parcel-tax-election-20190604-story.html

Brodeur, Abel, Mathias Lé, Marc Sangnier, and Yanos Zylberberg. 2016. "Star Wars: The Empirics Strike Back." American Economic Journal: Applied Economics, 8 (1): 1-32.

Brunner, Eric, Ben Hoen, and Joshua Hyman. (2021). School District Revenue Shocks, Resource Allocations, and Student Achievement: Evidence from the Universe of U.S. Wind Energy Installations. (EdWorkingPaper: 21-352). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/ssze-jq26

Brunner, E., Hyman, J., & Ju, A. (2019). School Finance Reforms, Teachers' Unions, and the Allocation of School Resources. *The Review of Economics and Statistics*, 1–47.

Candelaria, C. A., & Shores, K. A. (2019). Court-Ordered Finance Reforms in the Adequacy Era: Heterogeneous Causal Effects and Sensitivity. *Education Finance and Policy*, *14*(1), 31–60.

Card, D., & Payne, A. A. (2002). School finance reform, the distribution of school spending, and the distribution of student test scores. *Journal of Public Economics*, *83*(1), 49-82.

Carlson, D., & Lavertu, S. (2018). School improvement grants in Ohio: Effects on student achievement and school administration. *Educational Evaluation and Policy Analysis*, *40*(3), 287-315.

Carnoy, M., & Loeb, S. (2002). Does External Accountability Affect Student Outcomes? A Cross-State Analysis. *Educational Evaluation and Policy Analysis*, 24(4), 305–331.

Carrell, S., & Carrell, S. (2006). Do Lower Student to Counselor Ratios Reduce School Disciplinary Problems?. *Contributions to Economic Analysis & Policy*, 5(1), 1463-1463.

Cascio, E. U., Gordon, N., & Reber, S. (2013). Local Responses to Federal Grants: Evidence from the Introduction of Title I in the South. *American Economic Journal: Economic Policy*, *5*(3), 126–159.

Cellini, S. R., Ferreira, F., & Rothstein, J. (2010). The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design. *Quarterly Journal of Economics*, *125*(1), 215–261.

Clark, M. A. (2003). *Education Reform, Redistribution, and Student Achievement: Evidence from the Kentucky Education Reform Act*. Unpublished Manuscript.

Conlin, M., & Thompson, P. N. (2017). Impacts of new school facility construction: An analysis of a state-financed capital subsidy program in Ohio. *Economics of Education Review*, *59*, 13–28.

Cullen, J., & Loeb, S. (2004). "School Finance Reform in Michigan: Evaluating Proposal A." in John Yinger, editor *Helping Children Left Behind: State Aid and the Pursuit of Educational Equity*. Cambridge, MA: MIT Press.

Dee, T.S., & Jacob, B. (2011). The Impact of No Child Left Behind on Student Achievement. *Journal of Policy Analysis and Management*, 30(3), 418-446.

Downes, T. & Figlio, D. (1998). School Finance Reforms, Tax Limits, and Student Performance: Do Reforms Level Up or Dumb Down?. Institute for Research on Poverty Discussion Paper no. 1142-97.

Egger, M., Smith, G.D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ [British Medical Journal]*, 315(7109): 629–634.

Fiscal Accountability Summary (2018-19). State of New York. Retrieved from https://data.nysed.gov/fiscal.php?year=2019&instid=800000038389 and https://data.nysed.gov/fiscal.php?year=2019&instid=800000048663

Fryer, R. (2013). Teacher Incentives and Student Achievement: Evidence from New York City Public Schools. *Journal of Labor Economics*. 31(2), 373-403.

Gigliotti, P., & Sorensen, L. C. (2018). Educational resources and student achievement: Evidence from the Save Harmless provision in New York State. *Economics of Education Review*, *66*, 167–182.

Goldstein, Jessica, and Josh B. McGee. (2020). Did Spending Cuts During the Great Recession Really Cause Student Outcomes to Decline?. (EdWorkingPaper: 20-303). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/qzrd-0323

Goncalves, F. (2015). The Effects of School Construction on Student and District Outcomes: Evidence from a State-Funded Program in Ohio. *SSRN Electronic Journal*.

Gray, L. (2014). Katy superintendent fights for bond issue. *Houston Chronicle*, July 23. Retrieved from https://www.houstonchronicle.com/news/education/article/Katy-superintendent-fights-for-bond-issue-5300834.php

Guryan, J. (2001). Does Money Matter? Regression-Discontinuity Estimates from Education Finance Reform in Massachusetts. *NBER Working Paper*.

Holden, K. L. (2016). Buy the Book? Evidence on the Effect of Textbook Funding on School-Level Achievement. *American Economic Journal: Applied Economics*, *8*(4), 100–127.

Hong, K., & Zimmer, R. (2016). Does Investing in School Capital Infrastructure Improve Student Achievement. *Economics of Education Review*, 44.

Hoxby, C.M. (2001). All School Finance Equalizations are Not Created Equal. *The Quarterly Journal of Economics,* 116(4), 1189–1231.

Husted, T.A., & Kenny, L.W. (2000). Evidence on the Impact of State Government on Primary and Secondary Education and the Equity-Efficiency Trade-Off. *The Journal of Law & Economics*, 43(1), 285-308.

Hyman, J. (2017). Does Money Matter in the Long Run? Effects of School Spending on Educational Attainment. *American Economic Journal: Economic Policy*, *9*(4), 256–280.

Jackson, C. K. (2018) Does School Spending Matter? The New Literature on an Old Question (No. 25368). National Bureau of Economic Research.

Jackson, C.K. (2020) Expert Report of Clement Kirabo Jackson (Revised). In Re Delaware Public Schools Litigation, Court of Chancery of the State of Delaware, C.A. No. 2018-0029-VCL.

Jackson, C.K. (2021) Twitter. February 19. Accessed at https://twitter.com/KiraboJackson/status/1362947005476204551?s=20

Jackson, C. K., Johnson, R. C., & Persico, C. (2016). The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, *131*(1), 157-218.

Jackson, C. K., Wigger, C., & Xiong, H. (2018). *Do School Spending Cuts Matter? Evidence from the Great Recession* (No. w24203). National Bureau of Economic Research.

Johnson, Rucker C. (2015). Follow the Money: School Spending from Title I to Adult Earnings. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, *1*(3), 50.

Kogan, V., Lavertu, S., & Peskowitz, Z. (2017). Direct Democracy and Administrative Disruption. *Journal of Public Administration Research and Theory*, *27*(3), 381–399.

Kreisman, D., & Steinberg, M. P. (2019). *The Effect of Increased Funding on Student Achievement: Evidence From Texas's Small District Adjustment*. Unpublished Manuscript.

Kugley, S., Wade, A., Thomas, J., Mahood, Q., Jørgensen, A.-M.K., Hammerstrøm, K. and Sathe, N. (2017). Searching for studies: a guide to information retrieval for Campbell systematic reviews. *Campbell Systematic Reviews*, 13: 1-73. https://doi.org/10.4073/cmg.2016.1

Lafortune, J., Rothstein, J., & Schanzenbach, D. W. (2018). School Finance Reform and the Distribution of Student Achievement. *American Economic Journal: Applied Economics*, *10*(2), 1–26.

Lafortune, J., & Schonholzer, D. (2018). Do School Facilities Matter? Measuring the Effects of Capital Expenditures on Student and Neighborhood Outcomes. *Working Paper*, 85.

Lee, K.-G., & Polachek, S. W. (2018). Do school budgets matter? The effect of budget referenda on student dropout rates. *Education Economics*, *26*(2), 129–144.

Lipsey, M., Farran, D., Durkin, K., (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly*, 45 (4), 155-176.

Martorell, P., Stange, K., & McFarlin, I. (2016). Investing in Schools: Capital Spending, Facility Conditions, and Student Achievement (Revised and Edited). *Journal of Public Economics*.

Matsudaira, J. D., Hosek, A., & Walsh, E. (2012). An integrated assessment of the effects of Title I on school behavior, resources, and student achievement. *Economics of Education Review*, *31*(3), 1–14.

Miller, C. L. (2018). *The Effect of Education Spending on Student Achievement: Evidence from Property Values and School Finance Rules*. Unpublished Manuscript.

Neilson, C. A., & Zimmerman, S. D. (2014). The effect of school construction on test scores, school enrollment, and home prices. *Journal of Public Economics*, *120*, 18–31.

Papke, L. E. (2008). The Effects of Changes in Michigan's School Finance System. *Public Finance Review*, *36*(4), 456–474.

Pham LD, Henry GT, Kho A, Zimmer R. (2020) Sustainability and Maturation of School Turnaround: A Multiyear Evaluation of Tennessee's Achievement School District and Local Innovation Zones. *AERA Open*. doi:10.1177/2332858420922841

Rauscher, E. (2019). Delayed Benefits: Effects of California School District Bond Elections on Achievement by Socioeconomic Status. *EdWorkingPaper*, *No. 19-18*.

Roy, J. (2011). Impact of School Finance Reform on Resource Equalization and Academic Performance: Evidence from Michigan. *Education Finance and Policy*, *6*(2), 137–167.

Sass, T.R., Zimmer, R.W., Gill, B.P., & Booker, T.K. (2016). Charter High Schools' Effects on Long-Term Attainment and Earnings. *Journal of Policy Analysis and Management*, 35(3), 683–706.

U.S. Census Bureau, Public Education Finances: 2013 (2015). Table 3. Revenue From State Sources for Public Elementary-Secondary School Systems by State: Fiscal Year 2013. Retrieved from https://www.census.gov/content/dam/Census/library/publications/2015/econ/g13-aspef.pdf

U.S. Department of Education, Digest of Education Statistics. (2015). Table 235.20. Revenues for public elementary and secondary schools, by source of funds and state or jurisdiction: 2012-13. Retrieved from https://nces.ed.gov/programs/digest/d15/tables/dt15_235.20.asp

U.S. Department of Education, Digest of Education Statistics. (2019). Table 236.65. Current expenditure per pupil in fall enrollment in public elementary and secondary schools, by state or jurisdiction: Selected years, 1969-70 through 2016-17. Retrieved from https://nces.ed.gov/programs/digest/d19/tables/dt19_236.65.asp

van der Klaauw, W.(2008). Breaking the link between poverty and low student achievement: An evaluation of Title I. *Journal of Econometrics*, 142(2), 731–756.

Weinstein, M. G., Stiefel, L., Schwartz, A. E., & Chalico, L. (2009). Doesn't Title I Increase Spending and Improve Performance? Evidence from New York City. *IESP Working Paper Series*.

What Works Clearinghouse. Standards Handbook Version 4.0. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf