NEPC Review: Everything You Know About State Education Rankings is Wrong (Reason Foundation, November 2018) and Fixing the Currently Biased State K-12 Education Rankings (University of Texas, September 2018)

Reviewed by:

Bruce D. Baker
Rutgers University

November 2018

National Education Policy Center

School of Education, University of Colorado Boulder
Boulder, CO 80309-0249
(802) 383-0058
nepc.colorado.edu

Acknowledgements

NEPC Staff

Kevin Welner
Project Director

William Mathis
Managing Director

Alex Molnar
Publications Director

Suggested Citation: Baker. B.D. (2018). *NEPC Review: "Everything You Know About State Education Rankings Is Wrong" and "Fixing the Currently Biased State K-12 Education Rankings."* Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/thinktank/review-rankings.

Funding: This review was made possible in part by funding from the Great Lakes Center for Educational Research and Practice.

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: Everything You Know About State Education Rankings Is Wrong (Reason Foundation, November 2018) and Fixing the Currently Biased State K-12 Education Rankings (University of Texas, September 2018)

Reviewed by:

Bruce D. Baker
University of Colorado Boulder

November 2018

Executive Summary

The libertarian Reason Foundation recently published a policy brief that offers an alternative ranking of states' education systems. The brief is based on a working paper from the Department of Finance and Managerial Economics at the University of Texas at Dallas (UTD). These two reports begin with the presumption that high average test scores combined with lower school spending should be the basis for state rankings, which are reasonable premises, depending upon how the analyses are approached. But the reports then head off the rails. Offering a 'corrected' representation of student outcomes and a crude analysis asserting that spending has no relation to those outcomes, the reports declare states such as New Jersey and Vermont to be poor-performing, highly inefficient systems by comparison to states like Texas. The reports then estimate a regression model to confidently assert that the higher performing states are those with a) weaker teachers' unions and b) more children in charter schools. However, the reports' corrected outcome measures, weighting significantly unbalanced racial groups as equal and treating racial groups as equated across states without regard for economic status, are specious at best. Regressing multiple, highly related, interdependent measures against a specious outcome measure leads to even more suspect findings, and would only mislead policymakers.

NEPC
NATIONAL EDUCATION POLICY CENTER

NEPC Review: Everything You Know About State Education Rankings Is Wrong (Reason Foundation, November 2018) and Fixing the Currently Biased State K-12 Education Rankings (University of Texas, September 2018)

Reviewed by:

Bruce D. Baker
University of Colorado Boulder

November 2018

I. Introduction

In this review, I critique a web-based policy brief produced by the libertarian Reason Foundation which was based on a working paper from the Department of Finance and Managerial Economics at the University of Texas at Dallas (UTD).[1] The working paper and subsequent Reason brief had three apparent goals:

1. To reveal that common rankings (*U.S. News*, *Education Week*, WalletHub) of best states for education are biased, in that they often merely reflect how whiter and richer Northeastern states tend to have high test scores on the National Assessment of Educational Progress;

2. To reveal that state expenditures on education are not linked to student performance, or to the quality of state education systems, and that some state education systems are far more efficient than others; and

3. To show that states with stronger teachers' unions, and with smaller shares of children in charter schools, tend to have lower student outcomes when controlling for differences in spending and differences in pupil-to-teacher ratios.

The first point is generally without dispute. Major U.S. news outlets for years have produced half-baked comparisons of state education systems based on summaries raw data on average test scores, per-pupil spending, pupil-to-teacher ratios and so on. Clearly, as the authors explain, all of these factors are conflated with racial and economic composition of states. However, the authors' methodological solutions to this problem range from illogical to merely

wanting of methodological rigor or precedent.

It makes sense to view state (or local) education systems through alternative lenses of a) quality, b) effectiveness and c) efficiency. I delineate between quality, effectiveness and efficiency as follows:

> **Quality** might be characterized in terms of raw outcomes, along with the programs and services that coincide with those outcomes. Parents of school-aged children might wish to live and work in a state that does spend sufficiently on schools and have high-quality services. Parents might also knowingly or unknowingly be choosing a state with advantageous demographics. But their interest often lies in simultaneously accessing good schools AND advantaged peers.
>
> **Effectiveness** is different from raw quality in that effectiveness attempts to parse institutional contributions to student outcomes. That is, good effectiveness measures capture the extent to which institutions (or states) contribute to changes in student outcomes, controlling for who those students are and where they started. Policymakers should surely consider effectiveness and how to improve the effectiveness of schools. Parents/consumers might consider this as well, but they still may have greater interest in unconditional, raw outcomes.
>
> **Efficiency** adds yet another dimension, considering how much was spent or how many resources were used toward achieving a certain level of effectiveness. Policymakers should certainly consider, and hope to minimize, the expenditure of public resources toward achieving specific effectiveness and overall quality levels. Parents/consumers may also have an interest in finding the school district or state that offers that best combination of lower personal cost (tax burden) and sufficiently high quality. But caution is warranted in making efficiency the singular focus in choosing among systems. The school system that spends almost nothing and gets almost nothing for it, may be an efficient system, but not a good one.

The reports' delineations are less precisely articulated. The reports' approach in pursuing their second goal—showing that spending is not linked to performance and quality—involves two steps. First, the report offers a new quality rating for states, breaking out average NAEP scores by race, normalizing (z-scores) each state's race subgroups against the national average for each subgroup, and averaging those scores weighting each racial group equally for all states. Second, the report evaluates the relationship between these new state outcome measures and a "cost of living" adjusted calculation of per-pupil spending for each state.

Yet, as discussed below, the authors' choice to "correct" state outcome quality measures by treating the average outcomes of racial subgroups as equally weighted, and ignoring economic status entirely (using insufficient excuses for doing so), creates equal or greater biases than those it purports to resolve. In particular, two problems arise.

- The approach inappropriately places equal weight in states like Vermont or Wyoming on students comprising 1 to 2% of the population as the other 98 to 99%.
- The approach assumes highly aggregated racial classifications to represent similar

> populations across states (Hispanics in Texas are compared to Hispanics in Florida) and thus to be a more useful proxy for advantage/disadvantage than measures of economic status (because poverty thresholds do not consistently represent low-income populations across states).

The methodological flaws run even deeper in the report's regression model, which purportedly reveals that the strength of state teachers' unions is a (or even THE) primary contributor to lower student outcomes. This is compounded by the report's logical flaws. If, for example, stronger teachers' unions are associated with greater investment in public schooling (perhaps through a mechanism of political action), and if greater investment is associated with better programs, services and student outcomes, then stronger teachers' unions exert an overall positive effect on the quality of the education system. It makes little sense to analyze whether strong teachers' unions affect student outcomes while controlling for those resource differences; this simply negates the causal mechanism behind the improved outcomes. Such an approach could, by better methods and models, help parse whether union strength, while leading to higher spending and higher outcomes, does so less efficiently. This remains an open question, awaiting more rigorous empirical analyses than those provided in this study.

II. Findings and Conclusions of the Report

As noted above, the brief and working paper offer three major conclusions:

- "Traditional rankings are riddled with methodological flaws." (Reason brief)
- "expenditures are not linked to student performance." (Reason brief)
- "union strength has a powerful negative effect on student performance." (Reason brief)

The authors elaborate on these findings in their working paper, explaining:

> Many states in New England and the upper Midwest, with mostly white populations, fall in [our new] rankings, whereas many states in the South and Southwest, with large minority populations, score much higher than they do in conventional rankings. (working paper, p. 1)

And further, regarding their regression analyses of predictors of state school system quality, the authors note:

> We also find no evidence of a relationship between student performance and teacher-pupil ratios and private school enrollment, but some evidence that charter school enrollment has a positive impact. (working paper, p. 1)

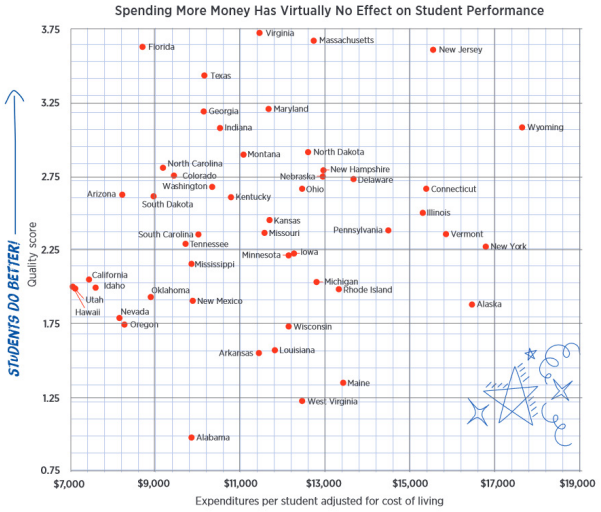
III. Rationale for Findings and Conclusions

The Reason report and the working paper (collectively, "the reports") appear to have baked an analysis to support the claim that northeastern (specifically citing New Jersey, but also Vermont), higher spending states are not as good as national reports claim and that southern, often lower spending states (specifically Texas), are actually much better. Further, if that's the case, money has no bearing on school quality.

The first step in setting up this claim is to create a new "quality" measure by weighting as equal the normalized NAEP scores for each race group by state, despite their vastly different proportions and disregarding economic status or subgroup composition differences between race groups across states. For example, while Vermont's scores for white children generally exceed national averages, the state's very small black population (<2%) underperforms, even compared to black children nationally.[2] Weighting them the same (rather than 98/2) substantially lowers Vermont's "quality" measure, but may merely reflect the instability of the scores for such a small sample or some other unique attribute of the small, tested sample of black children in Vermont. The same is true for Wyoming. Both are also high spending states (largely because most schools and districts in these white rural states are very low in enrollment, lacking economies of scale). So, the reports kill two birds with one stone, reducing quality measures for states that also happen to be high spending, thus mitigating the pattern across states between spending and quality.

That brings us to the evidence used by the Reason policy brief to drive home the major finding that money has no relation to quality. By this point, the report has achieved its goal of showing that some high-spending states like Wyoming and Vermont really don't perform all that well (once the analysis has inflated the influence of very small minority populations). But even on the adjusted outcome measures, states like New Jersey (for which the authors seem to hold a special degree of contempt) and Massachusetts continue to perform much better than most southern and southwestern states. The next step, therefore, is show that New Jersey in particular does so at exorbitant expense, whereas Texas performs nearly as well (or better) at much lower expense.

Figure 1



Combining these effects across all states, and using a "cost of living" adjustment for per-pupil spending, the authors construct the graph above to conclude:

> **"Spending more money has virtually no effect on student performance."**

It is tempting here to offer thousands of words kicking around the problems with using this anemic evidentiary basis to draw this bold conclusion. But hopefully a brief summary will make the point.

- To begin with, a vast body of far more rigorous, vetted and credible research—all of which is completely ignored in both reports—reveals the opposite;[3]
- To reiterate, the quality measure on the vertical axis is a deeply problematic treatment of NAEP data because small minority samples may dramatically alter a state's "score" and because racial classifications do not represent similar groups across states; and
- The per-pupil spending measure, adjusted only for regional cost of living, is equally insufficient.
    - First, even when it comes to adjusting for regional cost variation, cost of living is not the appropriate or generally accepted measure. Rather, the National Center for Education Statistics has adopted a Comparable Wage Index approach developed by Professor Lori Taylor at Texas A&M, which remains publicly available at the district or state level.[4]
    - Second, a plethora of additional factors, including student needs (poverty in particular), economies of scale, and population sparsity also affect the value of the education dollar (for example, significantly raising the per-pupil cost of providing schooling in states like Wyoming).

In other words, both the measure on the horizontal dimension (cost-of-living adjusted spending) and vertical dimension are rather meaningless if not outright deceptive.

The next step in these reports is to take one (score) and divide it by the other (spending) and call it an efficiency rating.[5] Importantly, if both the numerator and denominator are corrupt or meaningless, then so too is the quotient. As discussed in the next section, there exists a large literature on efficiency analysis in education, none of which is cited (or used, or perhaps ever read) by the authors of the working paper or Reason brief. The reports simply do not qualify as efficiency or cost-efficiency or cost-effectiveness analyses, applying any reasonable standard.

Finally, the reports draw bold conclusions about the lack of relationship between resources and quality and about the strong (negative) relationship between strong unions and quality, based on a deeply problematic regression analysis. The authors explain:

> In order to examine the relationship between expenditures and quality more precisely, we ran multiple regression analyses on our data, which included several other variables. The regression results support the view that expenditures are not linked to student performance. It turns out that throwing more money at something isn't guaranteed to yield improvement—as Kansas City demonstrated

> when, under court control from 1985 to 1997, it became the highest-spending school district in the country. It also failed to increase its performance. (Reason brief)

An unsurprising aside here is that the authors in the Reason brief (but not in the working paper) lean on a common urban legend in libertarian and conservative lore (without citation), which has its origins in a comparably ham-fisted 1998 Cato Institute paper on Kansas City.[6] I, along with Kevin Welner and previously with Preston Green, have rebutted the claim that Kansas City proved that money doesn't matter.[7] (In short, Kansas City was not the nation's highest-spending, and much of the increased spending during the court-ordered desegregation period was focused on infrastructure plus magnet schools designed to attract white students to the district.) I will further dissect the reports' regression model in Section V of this review.

IV. Use of Research Literature

The only body of literature reasonably explored in the reports pertains to the influence of unions on the quality of state school systems.[8] While the reports cite this literature, including footnotes explaining that the literature is indecisive in its conclusions, the authors take no methodological guidance from the more rigorous, recent studies of unionization and student achievement. As discussed below, the reports instead opt for an overly simplified, point-in-time, cross-state regression analysis.

The reports include no citations to research on measuring school performance and efficiency. This relatively large body of literature, summarized in part by Grosskopf, Hayes, and Taylor (2014),[9] could have provided important methodological guidance, related to methods for estimating educational efficiency and to problems with those methods, as well as better understanding cost factors.

The reports similarly ignore outright the large literature on factors affecting education costs and the value of the education dollar,[10] including costs related to a) economies of scale,[11] b) child poverty[12] and c) variation in competitive wages.[13] Further, the authors explain their choice to not use measures of income status (shares of low-income students) in their analysis because income thresholds used for determining child poverty are not sensitive to regional variation in wages and cost of living. However, the same scholars who have written extensively on regional variation in education costs, including competitive wage variation, have provided an elegant solution for adjusting poverty rates for greater comparability across states and regions.[14]

Finally, in drawing their conclusion that money is not related to student outcomes, the reports ignore entirely the vast literature to the contrary, including numerous rigorous recent longitudinal analyses.[15] The authors pay similarly short shrift to the overall body of research when reaching their finding of no relationship between pupil-to-teacher ratios and student outcomes, although they do provide a few choice citations to back up their conclusion.[16]

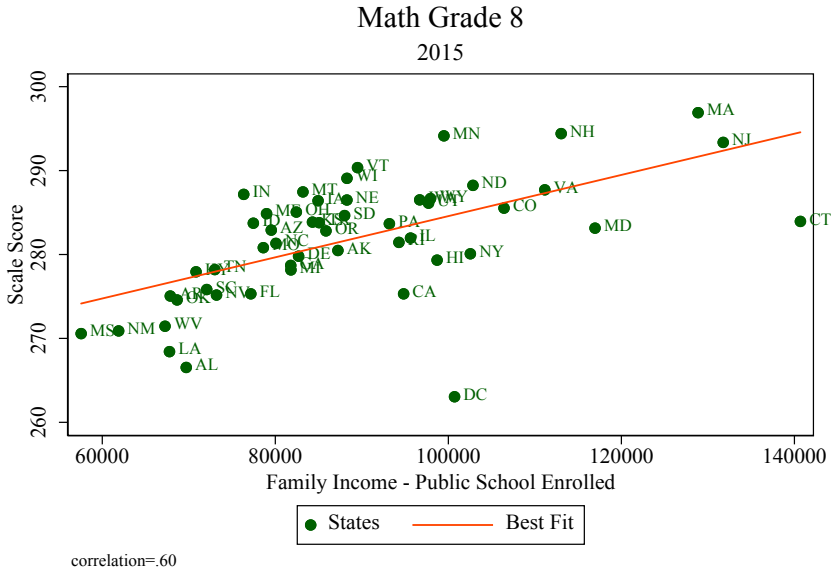
V. Review of Report's Methods

At best, these reports are methodologically inept and intentionally ignorant. I have already provided some critique of the report's methods, but here I will provide more specific illustrations as to why these methods are so problematic and deceptive, if not deceitful. Any basic literature search on Google Scholar would have revealed more rigorous methods and appropriate data for addressing the issue in question.

First, perhaps most troubling is the reports' construction of the "quality" measure. They aggregate, with equal weighting, "three subject matters (Math, Reading, and Science), four major ethnic groups (Whites, Blacks, Hispanics, and Asian/Pacific Islanders) and two grades[17] (fourth and eighth)" (working paper, p. 8). They explain further that: "We give each of the twenty-four tests[18] equal weight and base our ranking on the average of the test scores.[19] This ranking is thus limited to measuring learning, and does so in a way that avoids the aggregation fallacy. We refer to this as the 'quality' rank"[17] (p. 8).

In fact, the authors have created a new and more deceptive aggregation fallacy. Again, their approach creates a bizarre representation of state performance wherein very small, high- or low-performing racial subgroups can substantially alter state ratings. The choice to ignore outright or brush aside income or poverty measures is inexcusable when state-level NAEP scale scores remain so highly related to income variation across states, even without adjusting income for regional costs (correlation = .60). A more logical solution is to construct an income- or poverty-adjusted measure of NAEP performance to correct for the fact that richer, whiter states have higher average scale scores. I have done that here, taking the distance from each point in **Figure 2** to the diagonal line (expected NAEP scale score for a state with Y family income).

Figure 2



The reports assert that income or poverty measures mean different things across states,[18] while assuming (without validating) that racial classifications mean the same across states. This assumption is especially untrue for the second largest group nationally, and largest in some states – "Hispanic" populations. The Texas "Hispanic" population is unlike that of Florida or California, in terms of generational status, average income, education levels and national origin. These differences are so significant as to sometimes render the classification relatively meaningless – and cross-state normalization of their test scores without consideration for income or group composition becomes deeply problematic. "Hispanics" of various types are not geographically distributed in a random or equal way.

The case is similar if not more extreme for "Asian" populations, and while their influence on state average NAEP scores is limited, that's not the case with the analyses used in these reports. Their equal group weighting method dramatically overstates that influence. A population of low-income Cambodian refugees is likely to have lower scores than a population of wealthier, established "Asian" families.[19] Presuming these racial aggregations to be meaningful for cross-state performance comparisons on NAEP is problematic, at best.

Figure 3, 4 and 5 show that a) the Hispanic populations in Florida are substantively different from those of California or Texas, having a much larger share of Cuban and "other" Hispanic children. Further, the families of Cuban children tend to have much higher income than those of Mexican children, and commensurately higher test scores, on average.[20] The reports' method of assuming racial/ethnic group to be of constant meaning across states, and ignoring income differences, in this case bestows a significant advantage on Florida.

Figure 3

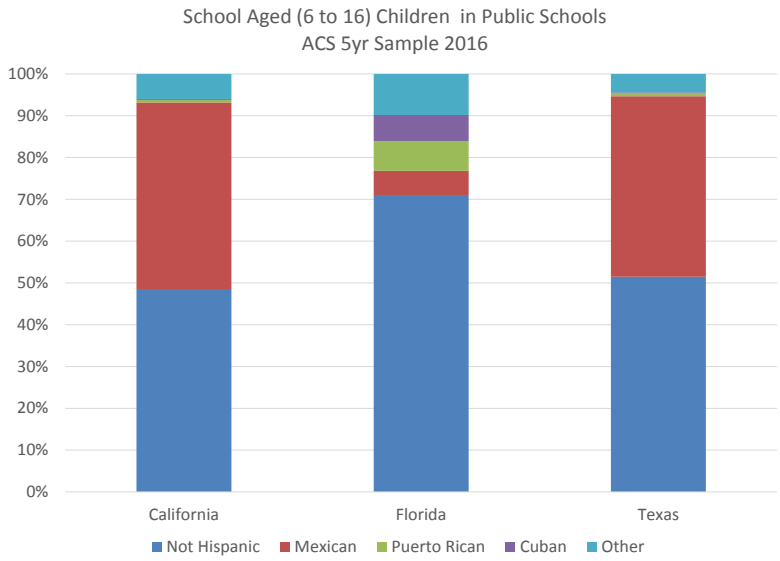


Figure 4

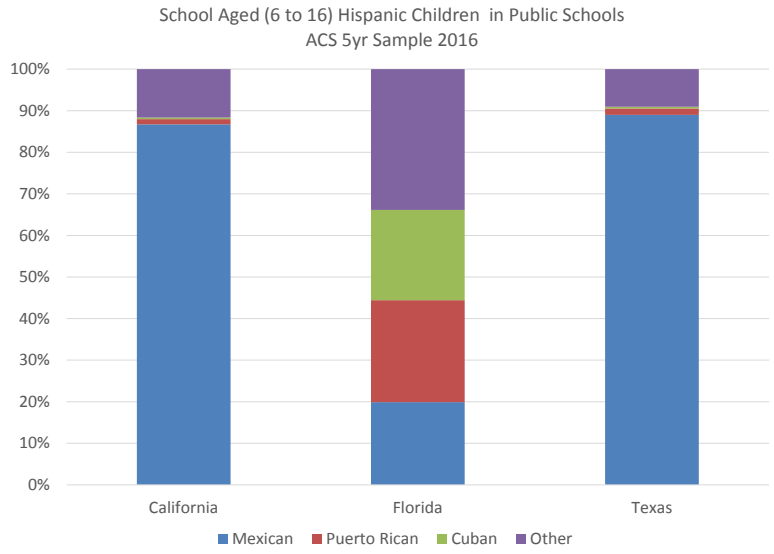
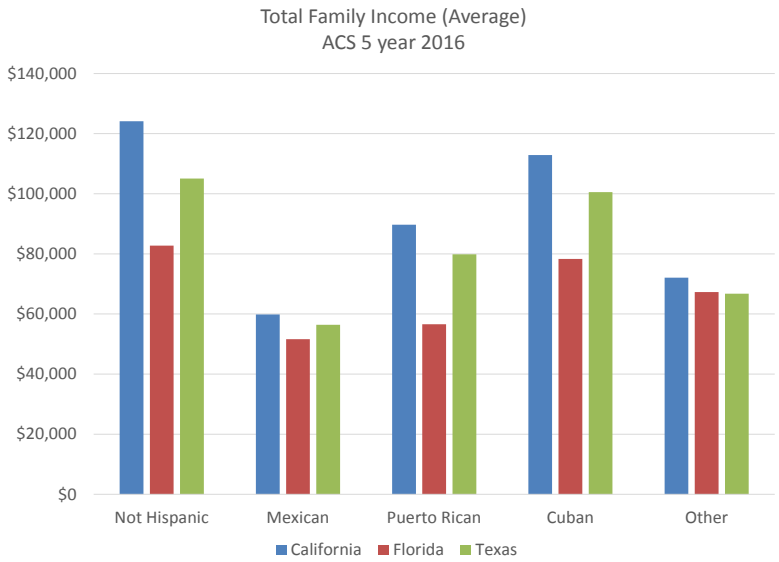


Figure 5



The reports seem to acknowledge this problem, to an extent. But, making matters worse, the reports emphasize as valid bias corrections, downward moves of states like Vermont, but then brush aside as spurious, upward moves of jurisdictions like Washington, DC, noting: "DC's high ranking is driven by the unusually outstanding scores of its white students, which were more than four standard deviations above the mean in each test subject they participated in" (p. 11). That is, when the direction of a state's change suits the authors' preferences,

they report uncritically, but when states (or DC) move against their preferences, they present concerns about their own methods.

The spending adjustment compounds the reports' problems. Ignoring or simply being unaware of the literature on regional cost variation in education, the authors choose to adjust their spending measure using a cost of living measure: "The statewide cost of living adjustments are taken from the Missouri Economic Research and Information Center's "Cost of Living Data Series 2017 Annual Average" (p. 13).[21] The vast body of research on this topic explains that reasonable models attempting to evaluate the efficiency of education dollars spent must consider three factors: a) regional variation in competitive wages, b) regional variation in population density and economies of scale, and c) variations in student needs including child poverty concentration, language proficiency and the distribution of children with mild, moderate, and severe disabilities. Most analyses fall short on the precision and completeness of these measures. These new analyses don't even come close.

The final straw in the working paper is the regression analysis reported below. Here, the authors take their adjusted quality rating as the dependent measure, and they estimate a regression model, including spending (and spending squared), students per teacher, and union strength. It is unclear if more than a single year of data were used in their regression. Regardless, the authors ignore two key truths. First, spending tends to be higher in states with stronger unions (correlation = .522,[22] whether causal or not). Second, pupil-to-teacher ratios tend to be lower in states where spending is higher (correlation = .60 in 2015).[23] That is, half (spending and spending squared, union strength, and students per teacher) of their independent variables are significantly related to one another. Further, as stated previously, from a logical standpoint if unions create political pressure to increase spending, which leads to higher quality outcomes, then the authors have explicitly negated the causal mechanism.

Table 1

Table 4. Multiple Regression Results Explaining Quality of Educating

	Cost of Living Adjusted		Nominal Dollars	
	coef	p	coef	p
Expenditure per Student	3.89E-05	0.871	3.75E-04*	0.062
Expenditure per Student^2	-2.75E-10	0.977	-1.04E-08*	0.089
Union Strength	-0.01125*	0.091	-0.024**	0.026
Students per Teacher	-0.04499	0.219	0.013	0.755
Private School Share of Students	-0.68112	0.823	-1.193	0.691
Charter Share of Students	1.96458**	0.033	1.098	0.342
Vouchers Allowed	-0.18267	0.435	-0.14306	0.538
Constant	0.53484	0.765	-2.44191	0.11
R-squared/Observations	0.150	51	0.217	51

Source: Working paper, p. 18.

I decided to test the robustness of some of these findings with readily available panel data (state level) from the *School Funding Fairness Data System*. Note that this is also not a rigorous analysis of this issue. It's just marginally more rigorous than the working-paper analysis; but I did not include here the "union strength" measure.[24]

Here, the outcome measure is the adjusted poverty rate, adjusted NAEP scale score. That is, the score is adjusted for differences in poverty rates across states,[25] where the poverty measure is separately adjusted for differences in competitive wages across states.[26] Using a biennial, symmetrical panel of data from 2000 to 2015, applying robust standard errors, Table 2 shows that the coefficients for spending (expressed in thousands of dollars per pupil, adjusted for wage variation, scale and population density) are positive in all cases, and statistically significant for Grade 4 Reading. Contrary to the working-paper analysis, the statewide share of children in charter schools displays a strong negative association,[27] consistently across all tests. Teachers per 100 pupils are positively associated with fourth grade outcomes, even in a model including the collinear spending measures. Put bluntly, even this simple robustness check casts significant doubt on the regression estimates offered in these two reports.

Table 2

	(adj.) Poverty Adj. Math 8		(adj.) Poverty Adj. math4		(adj.) Poverty Adj. read8		(adj.) Poverty Adj. read4	
	coef	**R.S.E**	**coef**	**R.S.E**	**coef**	**R.S.E**	**coef**	**R.S.E**
curexp000s	0.288	0.252	0.347	0.250	0.307	0.228	**0.386***	**0.202**
curexp000s_sq	-0.014	0.010	-0.018*	0.010	-0.015*	0.009	-0.019**	0.008
Statewide Share Enrolled in Charter Schools	**-7.476*****	**1.991**	**-7.441*****	**1.681**	**-9.338*****	**1.788**	**-8.030*****	**1.417**
% 6 to 16 Enrolled in Public School	6.346**	2.680	6.016**	2.417	4.446	3.594	2.741	3.308
Predicted Teachers per 100 Pupils at 20% Poverty	0.185	0.113	**0.242****	**0.109**	0.199	0.127	**0.268****	**0.109**
year==2000	-0.282	0.348	-0.330	0.372				
year==2002					-0.335	0.297	-0.323	0.293
year==2003	-0.275	0.240	-0.333	0.261	-0.350	0.253	-0.338	0.261
year==2005	-0.259	0.187	-0.315	0.200	-0.328*	0.198	-0.332	0.215
year==2007	-0.259*	0.141	-0.310**	0.153	-0.322**	0.147	-0.334**	0.165
year==2009	-0.229**	0.097	-0.266**	0.110	-0.270***	0.100	-0.278**	0.108
year==2011	-0.175**	0.081	-0.199*	0.103	-0.193**	0.087	-0.186*	0.104
year==2013	-0.100*	0.053	-0.112	0.070	-0.106*	0.060	-0.099	0.077
year==2015 (base)								
_cons	-7.773***	2.938	-8.029***	2.557	-6.202*	3.714	-5.530*	3.301
*note: *** p<0.01, ** p<0.05, *p<0.1*								

Data Source: Baker, B.D., Srikanth, A., Weber, M.A. (2016). Rutgers Graduate School of Education/ Education Law Center: School Funding Fairness Data System. Retrieved from: http://www.schoolfundingfairness.org/data-download (State data panel)

VI. Review of Validity of Findings and Conclusions

Given that the outcome measure of interest and the input measure (cost-of-living adjusted spending) are each deeply problematic, any conclusions drawn by constructing ratios with these measures or estimating regression models using these measures are suspect, if not rubbish. Yes, current popular ratings are bad and biased. No, the alternative offered in the Reason piece and the underlying working paper does not provide useful corrections to those biases. Nor do the scatterplot or regression analyses provided yield any valid insights into how, why or whether school spending matters in determining the quality of education, either at the school level or state aggregate.

The authors attempt to use their findings to make specific comparisons and draw conclusions about New Jersey and Texas, in response to comments made by New Jersey Governor Murphy:

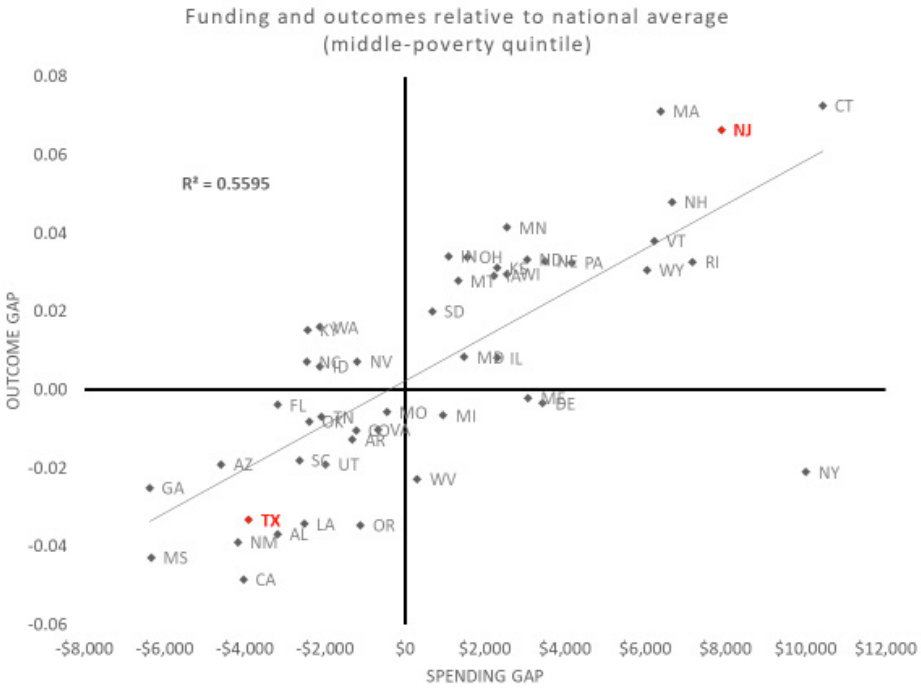
> Just last April, New Jersey Gov. Phil Murphy cited *Education Week* to claim the Garden State is superior in education to Texas. Murphy was responding to an April op-ed by Texas Gov. Greg Abbott inviting Jersey residents fed up with high taxes to move to Texas. Instead, Murphy should consider taking a page from the Lone Star State's playbook. Importing the policies that make Texas second on the list for "efficiency" could help maintain New Jersey's high performance while discarding its current punishing taxes. (Reason brief)

The reports' flawed adjusted quality metrics do indeed raise Texas outcomes to a relatively high level, and Texas does spend less per pupil than New Jersey, even on more thoroughly cost-adjusted measures. So Gov. Murphy can learn from Texas how to spend less. But he might not like the results of the decreased educational resources.

In fact, the reports' findings stand in sharp contrast to our own, from a working paper we released in the summer of 2018 in which we estimate a more thorough education cost model, using publicly accessible school district-level data from 2009 to 2015 to compare spending across states with respect to a target of achieving current average national outcomes.[28]

We estimate, for all districts nationwide, the predicted cost per pupil of achieving national average outcomes, assuming average cost efficiency (based on predictable efficiency factors). We then compare the actual per-pupil spending to the cost prediction and calculate the gap measures shown on the horizontal axis in Figure 6. This figure compares those spending gap measures to outcomes, relative to national average (0), for the median quintile of districts in each state by adjusted poverty rate. So, in many ways this figure is analogous to the scatterplot from the Reason report – comparing relative spending levels to relative outcome levels. But we control for a variety of cost factors, normalize and equate efficiency and address other important statistical concerns with conducting such analyses. In doing so, we find that spending variation is related to outcome variation.

Figure 6



Further, we show that states like New Jersey, Massachusetts and Connecticut spend more than necessary if their goal were to achieve national average outcomes—and, in fact, achieve well above national average outcomes. In contrast, states like Texas spend much less and achieve much less. In other words, you get what you pay for. Finally, making crude efficiency comparisons based on Figure 6, New Jersey and Massachusetts actually have higher outcomes than expected given their spending levels (above the diagonal line), but Texas actually has lower than expected outcomes, even given its relatively low spending levels (falling below the diagonal). Florida, by contrast, performs relatively average despite its low spending levels, and is thus relatively efficient. The position of New York State in the figure does raise efficiency concerns, but might be caused in part by insufficient equating of either the expenditure or outcome measures across states.

VII. Usefulness for Guidance of Policy and Practice

The Reason piece and the underlying working paper, attempting to analyze the performance of state education systems, are about as useful for guiding policy and practice as the popular media school rating systems they set out to replace. That is: not at all. The authors are correct in asserting that raw comparisons of average NAEP scores across states provide little insight into the effectiveness of state education systems. The authors are also correct in suggesting that it might further be useful to compare the relative cost efficiency of state education systems. That is, how much bang do they provide for the buck? Beyond those ba-

sic assertions, the analyses provided do little or nothing to advance the conversation. Even worse, they lead to misguided and unfounded conclusions too numerous to summarize in this short review. The reports' remedy for aggregate comparisons of NAEP scores across states is worse than the disease, eschewing measures of economic status as incomparable across state lines and adopting racial classifications as a rational substitute, even though they are not. Policymakers, whether or not in New Jersey, would be wise to decline any and all invitations to consider the authors' analyses or recommendations.

Notes and References

1 Liebowitz, S.J. & Kelly, M. (2018, May 25). *Fixing the currently biased state K-12 education rankings*. Available at SSRN: https://ssrn.com/abstract=3185152 or http://dx.doi.org/10.2139/ssrn.3185152 . The working paper was subsequently republished, with only small changes, as a CATO policy brief, available at https://object.cato.org/sites/cato.org/files/pubs/pdf/pa_854.pdf

Liebowitz, S.J. & Kelly, M. (2018, November). Everything you know about state education rankings is wrong. *Reason Magazine*. Available at: https://reason.com/archives/2018/10/07/everything-you-know-about-stat

2 National Center for Education Statistics. *Nation's report card: National assessment of educational progress*. Retrieved November 14, 2018, from https://www.nationsreportcard.gov/ndecore/xplore/NDE

3 See research discussed in Baker, B.D. (2016). *Does money matter in education?* Washington, DC: Albert Shanker Institute.

4 Taylor, L. *Education comparable wage index*. College Station, TX: Texas A&M University. Retrieved November 14, 2018, from http://bush.tamu.edu/research/faculty/Taylor_CWI/

5 An interesting aside here is that when the authors calculated their adjusted quality measure, they converted all NAEP scores to z-scores, which have an average of 0 and normal distribution around that average. In order to arbitrarily shift all efficiency values to positive values, the authors add (arbitrarily) 2.5 to each quality rating, calling it "slightly adjusted" (a technical term with which I am not aware). Efficiency calculations are described as follows: "calculated by taking a slightly revised version of the state's Z-score and dividing it by either the nominal dollar amount of educational expenditure or the cost of living adjusted education expenditure made by the state" (p. 15 of the working paper).

The authors explain in their Note #18: It would be a mistake to use straightforward Z-scores from Table 1 when constructing the "Z-Score/$" variable because states with Z-scores near zero and thus near each other would hardly differ from one another even when their expenditures per student were very different. Instead we added 2.50 to each Z-score so that all states have positive Z-scores and the lowest state would have a revised Z-score of 1. We then divided each state's revised Z-score by the expenditure per student to arrive at the values shows in Table 2" (pp. 15-16).

6 Ciotti, P. (1998). *Money and school performance: Lessons from the Kansas City desegregation experiment*. Washington, DC: Cato Institute.

7 Green III, P.C., & Baker, B.D. (2006). Urban legends, desegregation and school finance: Did Kansas City really prove that money doesn't matter. *Mich. J. Race & L.*, *12*, 57;

Baker, B., & Welner, K. (2011). School finance and courts: Does reform matter, and how can we tell. *Teachers College Record*, *113*(11), 2374-2414.

"...as Green and Baker (2006) explained, "in either adjusted or unadjusted dollars, KCMSD ranked first nationally for only one year, 1992. In most years, KCMSD trailed Boston (MA), Rochester (NY), Pittsburgh (PA), Portland (OR) or Newark (NJ) school districts" (p. 86). Further, "much of the short-term funding increase was indeed directed toward such infrastructure improvements and not direct classroom and related operating expenses, an approach designed as part of a magnet system intended to draw white students from the suburbs" (Baker & Welner, p. 2396).

8 Cowen, J.M., & Strunk, K. O. (2015), The impact of teachers' unions on educational outcomes: What we know and what we need to learn." *Economics of Education Review 48*, 208-23;

Hoxby, C.M. (1996). How teachers' unions affect education production," *The Quarterly Journal of Economics* *111*(3), 671-718, https://doi.org/10.2307/2946669; or

Kingdom, G. & Teal, F. (2010). Teacher unions, teacher pay and student performance in India: A pupil fixed effects approach, *Journal of Development Economics 91*(2), 278-88, https://doi.org/10.1016/j.jdeveco.2009.09.001;

For studies that find no effect of unionization, see Lovenheim, M.F. (2009). The effect of teachers' unions on education production: Evidence from union election certifications in three Midwestern states, *Journal of Labor Economics 27*(4), 525-87, https://doi.org/10.1086/605653;

More recently, only very small negative effects on student performance were found in Mariano, B.D., & Strunk, K.O. (2018). Bad end of a bargain?: Revisiting the relationship between collective bargaining agreements and student achievement. *Economics of Education Review*, https://doi.org/10.1016/j.econedurev.2018.04.006

9 Grosskopf, S., Hayes, K.J., & Taylor, L.L. (2014). Efficiency in education: research and implications *Applied Economic Perspectives and Policy 36*(2), 175–210.

For a comprehensive review of literature on efficiency analysis in education, see:

Baker, B.D. (2018). *Educational inequality and school finance: Why money matters for America's students*. Cambridge, MA: Harvard Education Press.

10 Duncombe, W., & Yinger, J. (2008). Measurement of cost differentials. In H.F. Ladd & M.E. Goertz (Eds.), *Handbook of research in education finance and policy*, pp. 238-256. New York City, NY: Routledge.

11 Andrews, M., Duncombe, W., & Yinger, J. (2002). Revisiting economies of size in American education: are we any closer to a consensus?. *Economics of education review*, *21*(3), 245-262.

Duncombe, W., & Yinger, J. (2007). Does school district consolidation cut costs? *Education Finance and Policy*, *2*(4), 341-375.

12 Duncombe, W., & Yinger, J. (2005). How much more does a disadvantaged student cost? *Economics of Education Review*, *24*(5), 513-532.

13 Taylor, L.L., & Fowler Jr, W.J. (2006). *A comparable wage approach to geographic cost adjustment. research and development report*. NCES-2006-321. Washington, DC: National Center for Education Statistics.

14 Baker, B.D., Taylor, L., Levin, J., Chambers, J., & Blankenship, C. (2013). Adjusted poverty measures and the distribution of Title I aid: Does Title I really make the rich states richer? *Education Finance and Policy*, *8*(3), 394-417.

15 Baker, B.D. (2018) Educational inequality and school finance: Why money matters for America's students. Cambridge, MA: Harvard Education Press Group.

16 Schanzenbach, D.W. (2014) *Does class size matter?* Boulder, CO: National Education Policy Center. Retrieved November 14, 2018, from http://nepc.colorado.edu/publication/does-class-size-matter

17 Here are the authors' own explanations regarding the construction of their quality index score, which are embedded in author's notes 18 and 19:

> (18) When states do not report scores for each of the twenty-four NEAP categories, those states have their average scores calculated based on those exams which are reported.
>
> (19) We equate the importance of each (of the 24) tests by forming, for each of the 24 possible exams, a 'Z-score' for each state, under the assumption that these state test scores have a normal distribution. The Z-statistic for each observation is the difference between a particular state's test score and the average score over all the states, divided by the standard deviation of those scores over the states. Our overall ranking is merely the average Z-score for each state. Thus, exams with greater variations or higher (lower) mean scores do not have any greater weight than any

other test in our sample. The Z-score measures how many standard deviations a state is above or below the mean score calculated over all states. One might argue that we should use an average weighted by the share of students, but we choose to give each group equal importance. If we had used population weights the rankings would not have changed very much since the correlations between the two sets of scores is 0.86, and four of the top five and four of the bottom five states remain the same.

18 "Because the distribution of cost of living values across states is not symmetrical, the difference in scores between students with subsidized lunches and students without, across states, is likely to be biased" (working paper, p. 8).

19 Kao, G. (1995, February). Asian Americans as model minorities? A look at their academic performance, *American Journal of Education 103*(2), 121-159.

20 Baker, B.D., Keller-Wolff, C., & Wolf-Wendel, L. (2000). Two steps forward, one step back: Race/ethnicity and student achievement in education policy research. *Educational Policy 14*(4), 511-529.

21 Missouri Economic Research and Information Center. *Cost of living data series*. Retrieved November 14, 2018, from https://www.missourieconomy.org/indicators/cost_of_living

22 Baker, B.D. *Endangering intelligent conversation: Comments on the latest Hanushekian crisis manifesto*. School Finance 101. Retrieved November 14, 2018, from https://schoolfinance101.wordpress.com/2013/09/16/endangering-intelligent-conversation-comments-on-the-latest-hanushekian-crisis-manifesto/

23 Baker, B.D., Srikanth, A., & Weber, M.A. (2016). *Rutgers Graduate School of Education/Education Law Center: School Funding Fairness Data System*. Retrieved November 14, 2018, from http://www.schoolfundingfairness.org/data-download

24 For more relevant and rigorous analyses, see:

Baker, B., & Weber, M. (2016). Beyond the echo-chamber: State investments and student outcomes in US elementary and secondary education. *Journal of Education Finance, 42*(1), 1-27.
(includes fixed, random and mixed effects models)

25 Baker, B.D., Srikanth, A., & Weber, M.A. (2016). *Rutgers Graduate School of Education/Education Law Center: School Funding Fairness Data System*. Codebook. Retrieved November 14, 2018, from https://drive.google.com/file/d/0BxtYmwryVI00NDNjMWNwS3ZaMDg/view

26 Baker, B.D., Taylor, L., Levin, J., Chambers, J., & Blankenship, C. (2013). Adjusted poverty measures and the distribution of Title I Aid: Does Title I really make the rich states richer? *Education Finance and Policy*, *8*(3), 394-417.

27 For those tempted to see this as evidence of charter-school harm, please again note that this type of analysis does not allow for any sensible causal inference. It is offered merely to demonstrate the lack of robustness of the analyses in these two reports.

28 Baker, B.D., Weber, M., Srikanth, A., Atzbi, M., & Kim, R. (2018). *The real shame of the nation: Causes and consequences of interstate inequity in public school investment*. Newark and New Brunswick, NJ: Education Law Center of New Jersey & Rutgers GSE.