NEPC Review: Bigger Bang, Fewer Bucks? (University of Arkansas Department of Education Reform, February 2018)

Reviewer:
Julian Vasquez Heilig
California State University Sacramento
May 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


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

This publication is provided free of cost to NEPC’s readers, who may make non-commercial use of it as long as NEPC and its author(s) are credited as the source. For inquiries about commercial use, please contact NEPC at nepc@colorado.edu.
A recent report released by the University of Arkansas Department of Education Reform contends that charter schools produce more achievement per dollar invested, as compared to public schools. This newest report is focused on city-level analyses in eight US cities (Atlanta, Boston, Denver, Houston, Indianapolis, New York City, San Antonio, and Washington D.C.) and uses cost effectiveness and Return on Investment (ROI) ratios. It concludes that charter schools deliver a weighted average of an additional 4.34 NAEP reading points and 4.73 NAEP math points per $1000 invested. The report also argues that that charter schools offer an advantage of $1.77 in lifetime earnings for each dollar invested, representing a ROI benefit of 38%. However, there are a variety of methodological choices made by the authors that threaten the validity of the results. For example, the report uses revenues rather than actual expenditures – despite well-established critiques of this approach. The report also fails to account for the non-comparability of the student populations in charter and comparison public schools. Three other problems also undercut the report’s claims. First, even though the think tank’s earlier productivity report included a caveat saying that causal claims would not be appropriate, the new report omits that caution. Second, the report’s lack of specificity plagues the accuracy and validity of its calculations; e.g., using state-level data in city-level analyses and completely excluding race and gender. Finally, the authors again fail to reconcile their report with the extensive literature of contrary findings.
I. Introduction

In February 2018, the University of Arkansas Center for Education Reform released a second report aiming to address the relative productivity of charter schools (CS) as compared with traditional public schools (TPS). The newest report focuses on city-level cost effectiveness and lifetime Return on Investment (ROI) calculations in eight US cities (Atlanta, Boston, Denver, Houston, Indianapolis, New York City, San Antonio, and Washington D.C.). The report is entitled *Bigger Bang, Fewer Bucks? The Productivity of Public Charter Schools in Eight U.S. Cities.*

Both reports claim that CS’s are more effective in producing achievement at less “cost” (meaning spending) per pupil than TPS’s. This latest report includes two main sections. In the first, the authors calculate a measure of cost effectiveness using the National Assessment of Educational Progress (NAEP) and per-pupil revenues. In the second section, they calculate a Return on Investment (ROI) between CS and TPS sectors using Center for Research on Education Outcomes (CREDO) achievement data and measures of lifetime earnings. There are two appendices that describe methods and data sources.

The cost effectiveness results are reported in terms of NAEP scores per $1,000 expenditure at the state level. The ROI between the CS and TPS sectors uses CREDO achievement calculations. In both analyses, the authors conclude that CS spend far less public dollars per pupil than TPS, yet generally produce financial outcomes as good as, or superior to, that of TPS. This sequel renews the attempts to demonstrate the cost effectiveness and ROI benefits of charter schools.

This review finds a repetition of many of the shortcomings that Gene Glass found in the first report in this series. Specifically, the report’s comparison of achievement scores between
the CS and TPS sectors suffers from multiple sources of methodological invalidity and questionable data. For example, Miron and Baker criticized the cost effectiveness analyses in the charter funding analyses as incorrectly focusing on revenues instead of actual expenditures. Additionally, the new ROI analyses using CREDO achievement data at the city-level, fails to consider that African American and Latino students’ test scores in the eight cities examined trend differently from the overall performance in CS compared to TPS. In fact, neither African Americans or Latinos are mentioned once in the entire report.

II. Findings and Conclusions of the Report

Comparing NAEP achievement for CS versus TPS in the subject cities, the report concludes that the charter schools outperform TPSs in both math and reading cost effectiveness as well as in ROI.

The cost effectiveness findings show:

- A CS cross-city advantage of 4.34 points in reading and 4.73 points in math per $1,000, representing a cost effectiveness benefit of 32% in reading and 33% in math.
- A student weighted CS advantage of 3.99 points in reading and 4.37 points in math per $1,000, representing a cost effectiveness benefit of 35% in reading and 36% in math.
- A cost effectiveness advantage for CS compared to TPS regarding NAEP reading and math scores ranging across the cities from a low of 2% (Houston) to a high of 67% and 68%, respectively (Washington, D.C.).

The return on investment findings show:

- A CS advantage of $1.77 in lifetime earnings for each dollar invested, representing a ROI benefit of 38%.
- A student weighted CS advantage of $2.09, representing a ROI benefit of 53%.
- Spending only half of the K-12 educational experience in CS results in $5.40 in benefits for each invested dollar, a 16 percent advantage relative to a full-time (13 year) K-12 experience in TPS or 29 percent if student-weighted.
- The CS advantage for an entire K-12 education for charters schools compared to TPS ranges from 4 percent (Houston) to 85 percent (Washington, D.C.).
III. The Report’s Rationale for its Finding and Conclusions

The report’s calculation of the relative cost effectiveness and ROI of charter and traditional schools is based on the examination of NAEP test score averages and CREDO achievement data from the eight metropolitan areas. Estimates of the per-pupil expenditures in the two sectors are provided for each city.

To calculate cost effectiveness, the report simply divides average NAEP scores by the average expenditure at the city level, which is said to produce a cost effectiveness ratio that can be compared between the two sectors.

The ROI calculation includes several components. The report divides the “income returns to investment” by the “cost of investment.” They calculate the numerator by multiplying lifetime earnings in a given state by the average impact (standard deviation difference) in CREDO’s achievement data, Hanushek’s cognitive ability multiplier (.13), and Hanushek’s depreciation of learning (.70). They calculate the denominator—cost of investment—by multiplying Wolf et al.’s prior measurement of per-pupil revenue by years of sector enrollment.

IV. The Report’s Use of Research Literature

Issues such as the relative costs and effectiveness of charter schools and traditional public schools are widely debated in the research literature of the past two decades. In the first report in this series, the authors did briefly attempt to refute claims by Miron and Baker that big funding inequities between the two sectors occur because special populations are more expensive to educate and that TPS schools serve greater proportions of needy children. However, as discussed below, these earlier critiques still hold true.

In this second report, alternative perspectives for comparing the cost effectiveness and ROI of CS and TPS are now entirely absent from the report and research contrary to their perspective is not addressed. Glass, in the first NEPC review of this series, made the case that the research literature CS and TPS are often not directly comparable. He stated, “...the failure to reconcile the reported findings with a large literature of contrary evidence is particularly egregious” (p. 3).

V. Review of the Report’s Methods

The report hinges on the estimation of two things: 1) the relative cost effectiveness based on revenues and performance on NAEP achievement data for students enrolled in CS and TPS, and 2) the ROI of educating the average pupil in CS versus TPS based on CREDO achievement and lifetime earnings data.

First, the authors did not address on the incomparability of using revenues instead of actual
expenditures. A cursory response to this critique can be found in the methodology section of the new report. In the authors’ eyes, using revenue instead of expenditures is not a problem because charter schools purposefully chose not to provide a suite of services. But this evades the real problem. Glass\textsuperscript{13} joined Miron\textsuperscript{14} and Baker\textsuperscript{15} in arguing that expenditures often pass through districts (special education, compensatory education, food, transportation, capital costs etc.) As a result, there are large “apples and oranges” comparisons. While the report mentions that expenditure and pass-through data were “usually” obtained, it does not adequately explain the scale and scope of expenditures and pass-through financial data for metropolitan areas other than New York City. Despite this lack of information, the authors attempt to convince readers that TPS and CS are comparable by simply taking overall revenue spent on everything that schools do, rather than specifically comparing expenditures related to instruction. This comparability error glaringly and unremittingly remains the weak link of their cost effectiveness calculations.

Second, the report uses a single grade to measure cost effectiveness – eighth grade NAEP scores. The NAEP is also administered at grades four and 12. The first report claimed that these data were ignored because fourth grade would underestimate effects and 12th grade scores would overestimate them. The second report still states that fourth grade NAEP “understates” learning. The same paragraph posits, “the results are similar if fourth grade NAEP scores are used in place of eighth grade scores” (p. 11). The authors argue that they ignore 12th grade because NAEP “results likely overstate overall learning levels because they do not include struggling students who dropped out prior to 12th grade” (p. 11). However, the report does not provide documentation to support either of these claims. By not including the fourth grade or 12th grade analyses, the authors forgo the opportunity to publicly explore the robustness of their findings.

Third, the CS and TPS data are not equally representative. In the first productivity report, Wolf et al. presented percentages of Free or Reduced-Price Lunch (FRL) and Special Education pupils in each sector for each state. The authors claimed,

The charter sectors in our study actually tend to enroll a higher percentage of low-income students than the TPS sectors, regardless of whether one uses free lunch or FRL as the poverty measure. The special education enrollment gap of just 3 percentage points is far too small to explain much of the charter school funding gap, even if many of the additional special education students in the TPS sector had the most severe, highest cost, disabilities imaginable. As our revenue study concluded, a far more obvious explanation for the large charter school funding gap is that state and local policies and practices deny public charter schools access to some educational funding streams...(p. 11).\textsuperscript{16}

The current productivity report, however, acknowledges special population disparities between CS and TPS at the city level. It is notable that these disparities occur in a majority of the cities in their report. TPS are serving more English language learners and special education students.

We found that three of our cities – Denver, Houston, and New York – enrolled higher or similar rates of low-income students in their charter sectors compared to their TPS sectors in 2014. The other five cities – Atlanta, Boston,
Indianapolis, San Antonio, and Washington – enrolled a higher rate of low-income students in their TPS than their charter sectors but the differences were only large in the case of Indianapolis. The TPS sectors more consistently enrolled higher percentages of students labeled as English learners or in special education...(p. 10-11).

However, the report asserts that these disparities “failed to explain much of the revenue differences,” but it does not provide any basis to demonstrate, much less prove, that claim. The authors also fail to specifically report special populations data for the eight urban cities for CS and TPS. This omission is a step backwards from the prior productivity report. The authors acknowledge that “different levels of student disadvantage across the public school sectors in our cities explain some but not all of the productivity advantage for public charter schools” (p. 11). Yet, the report does not specify how much “some” is in their report.

Fourth, even though the report focuses on cities with large communities of color, the report fails to take into account or even mention African Americans, Latinos or any other race/ethnicity. Considering the political framing about CS – that they are an education reform that should benefit urban African American and Latino students and families – the reader would expect that the report would disaggregate the analyses by race/ethnicity. Furthermore, the NAACP’s Task Force on High Quality Education report found that one in eight African American students now attends CS in the United States – more than any other race/ethnicity. As a result, it is important to consider whether the ROI analyses would look different if race/ethnicity was taken into account.

Table 1. Comparing CREDO Achievement Results for African Americans in CS and TPS Sectors

<table>
<thead>
<tr>
<th>Sector</th>
<th>TPS Math</th>
<th>TPS Read</th>
<th>CS Math</th>
<th>CS Read</th>
<th>TPS vs. CS Math</th>
<th>TPS vs. CS Read</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta (GA)</td>
<td>-0.25</td>
<td>-0.37</td>
<td>-0.25</td>
<td>-0.36</td>
<td>0</td>
<td>0.01</td>
</tr>
<tr>
<td>Boston (MA)</td>
<td>-0.16</td>
<td>-0.11</td>
<td>0.12</td>
<td>0.03</td>
<td>0.28</td>
<td>0.14</td>
</tr>
<tr>
<td>Denver (CO)</td>
<td>-0.16</td>
<td>-0.19</td>
<td>-0.2</td>
<td>-0.21</td>
<td>-0.04</td>
<td>-0.02</td>
</tr>
<tr>
<td>Houston (TX)</td>
<td>-0.29</td>
<td>-0.28</td>
<td>-0.32</td>
<td>-0.31</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>Indianapolis (IN)</td>
<td>-0.11</td>
<td>-0.06</td>
<td>-0.03</td>
<td>0</td>
<td>0.08</td>
<td>0.06</td>
</tr>
<tr>
<td>New York (NY)</td>
<td>-0.29</td>
<td>-0.29</td>
<td>-0.15</td>
<td>-0.29</td>
<td>0.14</td>
<td>0</td>
</tr>
<tr>
<td>San Antonio (TX)</td>
<td>-0.13</td>
<td>-0.15</td>
<td>-0.24</td>
<td>-0.28</td>
<td>-0.11</td>
<td>-0.13</td>
</tr>
<tr>
<td>Washington DC</td>
<td>-0.38</td>
<td>-0.39</td>
<td>-0.3</td>
<td>-0.34</td>
<td>0.08</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Note in Table 1 that African Americans show negative achievement effects (in standard deviation differences) in both TPS and CS in seven of the eight cities. Only in Boston do Black CS students modestly outperform TPS in reading and math. Notably, Boston CS are a special case considering that a prior NEPC review found that the majority of the CS do not enroll an equivalent proportion of ELL students and many are enrolling larger populations of students with low-cost disabilities. In three cities (New York City, Indianapolis, Washington D.C.), CS have a negative impact on Black students but perform slightly better than TPS.

http://nepc.colorado.edu/thinktank/review-roi

8 of 15
There are also three cities (San Antonio, Houston, Denver) where TPS have a negative impact on Black students but perform slightly better than CS. In the case of Atlanta, the impact is negative for both sectors and they are in a virtual test-score-achievement tie. Considering the varying success for African Americans in TPS at the city-level, it is readily apparent that achievement differentials in the CREDO data are not properly represented by overall averages (that do not account for difference by race/ethnicity) used in the ROI calculations. In essence, the calculations lack accuracy and validity because they underestimate the TPS ROI for African Americans in several cities.

Fifth, not only do the analyses fail to consider differential results by race/ethnicity, but a consideration of gender and the interaction between race/ethnicity and gender is also absent. Ongoing sexism and racism clearly impact lifetime earnings in the United States, and the ROI results are not discounted in any way to account for the persisting inequity in society. For example, the National Women’s Law Center (NWLC) has examined the race and gender wage gap nationally. The NWLC addressed these disparities by state using American Community Survey data. They found,

Based on today’s wage gap, women would lose $418,800 over the course of a 40-year career. For Latinas the career losses mount to $1,043,800, and for Black women the losses are $840,040... This “lifetime wage gap” exists across the country and in every state...

Sixth, the report’s lack of proper comparisons also plagues the accuracy and validity of lifetime earnings in the ROI formula. To calculate the numerator (income return to investment), Wolf et al. multiplied lifetime earnings in a given state by the average impact (standard deviation difference) in city-level CREDO achievement data. The footnotes explain that the ROI analyses use Bureau of Labor Statistics (BLS) state-level data for the calculation of lifetime earnings. However, more specific and relevant wage data for all occupations is available by metropolitan area. Table 2 shows that in some cities, their purported ROI is likely overestimated, and in other cities it is likely underestimated.

### Table 2. Comparing Metropolitan and State-Level Yearly Lifetime Earnings

<table>
<thead>
<tr>
<th>Annual Mean Wage</th>
<th>Metropolitan Area(^{21})</th>
<th>Annual Mean Wage</th>
<th>State(^{22})</th>
<th>Proportional Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>$50,720</td>
<td>Atlanta</td>
<td>$46,540</td>
<td>Georgia</td>
<td>91.8%</td>
</tr>
<tr>
<td>$64,080</td>
<td>Boston</td>
<td>$60,840</td>
<td>Massachusetts</td>
<td>94.9%</td>
</tr>
<tr>
<td>$55,910</td>
<td>Denver</td>
<td>$52,710</td>
<td>Colorado</td>
<td>94.3%</td>
</tr>
<tr>
<td>$52,870</td>
<td>Houston</td>
<td>$47,770</td>
<td>Texas</td>
<td>90.4%</td>
</tr>
<tr>
<td>$46,840</td>
<td>Indianapolis</td>
<td>$42,940</td>
<td>Indiana</td>
<td>91.7%</td>
</tr>
<tr>
<td>$61,790</td>
<td>New York City</td>
<td>$58,910</td>
<td>New York</td>
<td>95.3%</td>
</tr>
<tr>
<td>$45,210</td>
<td>San Antonio</td>
<td>$47,770</td>
<td>Texas</td>
<td>105.7%</td>
</tr>
<tr>
<td>$68,000</td>
<td>Washington D.C</td>
<td>$82,950</td>
<td>District of Colombia</td>
<td>122.0%</td>
</tr>
</tbody>
</table>

Finally, the report’s findings are merely descriptive and not causal. In the first Department of Education reform productivity report, the prior authors inserted a significant caveat. The
findings, readers are told, are:

...merely descriptive, not causal, because charter schools might be reporting higher NAEP scores per $1000 invested than traditional public schools because of the characteristics of students attracted to the charter school sector and not because they actually do a better job educating similar students and at a lower cost. (p. 21).23

Glass24 explained that this caveat “basically undercuts the conclusions and recommendations of the report” and was “missing from the press releases and media coverage.” In the new report, the authors attempted to address this limitation and subsequent critique by claiming that the ROI analyses are “rigorous” because they:

...use CREDO results based on a quasi-experimental methodology that eliminates many observable differences in student background characteristics across the public charter and TPS sectors. (p. 21).25

Yet an NEPC review of CREDO’s methodology and calculations directly undercuts their assumption.26 First, CREDO studies do not compare schools, but are conducted at the “virtual” student level. As a result, their student-level analyses are not considered within the context of classrooms or schools. Second, the NEPC review of CREDO’s approach provided five relevant critiques that suggest that findings from any analyses using the CREDO data as “quasi-experimental” are highly problematic:

• The nature of the comparison between CS and TPS in the CREDO studies is not clear.
• The matching variables used in CREDO’s studies may not be sufficient to support causal conclusions.
• Some lower-performing CS students are systematically excluded from the CREDO studies.
• CREDO’s reasons for the systematic exclusion of lower-scoring CS students does not address the bias arising from the exclusion.
• The CREDO studies lack an appropriate correction for multiple significance tests.

In sum, considering these methodological critiques – a lack of clarity and validity appears to be a trend in this series. Baker27 had previously argued that that the first productivity report suffered from “alarmingly vague documentation regarding data sources and methodologies, and many of the values reported cannot be verified by publicly available [information] or adequately documented...”
VI. Review of the Validity of the Findings and Conclusions

The title of the report, *Bigger Bang, Fewer Bucks? The Productivity of Public Charter Schools in Eight U.S. Cities*, poses a question that the report does not answer. Glass\textsuperscript{28} put it best, “to argue that a simple arithmetic ratio of NAEP points and revenues describes a school’s ‘productivity’ is little more than a weak metaphor” (p. 8).

The lack of validity of these estimates and conclusions can be summarized:

- First, the report continues to use revenues rather than expenditures directly related to student achievement, despite extensive prior critique of the inaccuracy of using this approach.
- Second, even though additional comparison grades are available, the authors used only a single grade from the eight cities.
- Third, unlike the first report, the authors acknowledge differences in student populations between CS and TPS. However, the authors do not address this fundamental comparability problem. TPS serve more special populations and these differences are important.\textsuperscript{29} As a result, the cost effectiveness of CS and TPS in the NAEP analyses is likely biased in favor of CS due to the non-comparability of the student populations.
- Fourth, even though the report focuses on cities that serve majority minority student populations, the report fails to take into account or even mention African Americans, Latinos or any other race/ethnicity.
- Fifth, not only do the analyses fail to consider differential results by race/ethnicity, but a consideration of gender and the interaction between race/ethnicity and gender is also absent.
- Sixth, the report’s use of state-level data instead of available metropolitan region data suggests they used the wrong comparison group.
- Finally, the report’s findings are merely descriptive and not causal. Notably the second report excludes their previous caveat about causal claims.

Considering these methodological issues, claims made in the new report regarding the cost effectiveness and ROI analyses are precarious instead of “rigorous” (p. 21).\textsuperscript{30} By any reasonable interpretation, it is clear that the authors have again produced a product that is still “little more than political arithmetic” (p. 9).\textsuperscript{31}
VII. Usefulness of the Report for Guidance of Policy and Practice

This second report in the CS versus TPS productivity series from the University of Arkansas Department of Education reform will likely be cited by supporters of the CS movement when they are lobbying for increased funding and favorable legislative treatment. Unfortunately, the evidence in this report is so flawed that it provides no valid guidance to educators or policymakers who aim to evaluate cost effectiveness or return on investment for either charter or traditional schools.
Notes and Resources


2 This report refers back to the Glass NEPC review of the first report in this series for two purposes: (1) to explain problems that are carried forward into the new report, and (2) to explain differences in this new report. I will also refer back to the earlier Arkansas report, to explain how the new report fits within the larger arguments that the researchers are putting forward about the purported productivity benefits of charter schools.


19 Also a general critique of CS and TPS comparisons levied by Miron, Baker and Glass earlier in the review.


21 Data calculated from https://www.bls.gov/oes/current/oessrcma.htm

22 Data calculated from https://www.bls.gov/oes/current/oessrcst.htm


http://nepc.colorado.edu/thinktank/review-roi