NEPC

NATIONAL EDUCATION POLICY CENTER

REVIEW OF *LESSONS FROM STATE PERFORMANCE ON NAEP: WHY SOME HIGH-POVERTY STUDENTS SCORE BETTER THAN OTHERS*

Reviewed By

Sharon L. Nichols
University of Texas at San Antonio

February 2016

Summary of Review

This Center for American Progress report examines whether states' adoption of standards-based policies predicts low-income students' NAEP achievement trends in fourth and eighth grade math and reading throughout the 2003-2013 decade. The report claims to analyze changes across five separate two-year intervals, but it only reports findings for 2009-2011, with no explanation of why or any documentation of the representativeness of that single interval. The reported finding for the selected interval is that state adoption of standards policies positively predicts fourth (but not eighth) grade math NAEP and eighth (but not fourth) grade reading NAEP. Even these selected positive results are statistically significant only at the generally unacceptable 0.10 level of significance. The report includes effect sizes but nothing about the percentage of the variance explained in their model. In short, the report does not adequately describe variables or analytic methods or completely report findings, and the data and methods used do not allow for any causal findings. They use state standards adoption across grade levels and subject areas as well as selected accountability mandates as predictors but fail to assess their quality or fidelity of implementation. Yet based on these very problematic and limited analyses, the authors conclude that their analysis "strongly supports the potential of the Common Core to drive improvements in educational outcomes." The study simply does not support this conclusion or the set of recommendations that follow.

Kevin Welner
Project Director

William Mathis
Managing Director

Alex Molnar
Publishing Director

National Education Policy Center

Review of *Lessons From State Performance on NAEP: Why Some High-Poverty Students Score Better Than Others*

Sharon L. Nichols, University of Texas at San Antonio

I. Introduction

The title of the Center for American Progress (CAP) report, "*Lessons From State Performance on NAEP: Why Some High-Poverty Students Score Better than Others*"[1] conveys a goal that is important and informative. After all, decades of educational policy efforts have focused on improving the educational outcomes of our poorest students.[2] The Elementary and Secondary Education Act of 1965 followed by No Child Left Behind Act of 2001 and more recently the Every Student Succeeds Act all have a core goal to improve educational conditions and outcomes for all of our students, especially our nation's most vulnerable including minority, poor, ELL, and special education students.[3] The title seems to promise some answers that are immediately telegraphed in the report's table of contents: The difference between high- and low-scoring states is the adoption of high standards. Specifically recommended is the Common Core.

Standards-based policies have dominated the national landscape since the early 1990s. Although the conversation about standards dates back to the nineteenth century, the more recent impetus was sparked by the recommendations of the 1983 *Nation at Risk* report[4] that called for an overhaul to our nation's educational system, advocating for "higher" standards and expectations for all of our students. Since that time, standards have been part of every Presidential administration's education agenda. President Bill Clinton was an advocate for standards-based reform in Goals 2000, George W. Bush signed NCLB, formalizing the role of standards in education, and President Obama continued that legacy with his support of federally mandated, standards-based reform as mandated in NCLB and reinforced with School Improvement Grants and the Race to the Top initiative.[5] The goal of CAP's report is to examine the impact of these diverse policy efforts by asking whether increases in states' adoption of standards-based reform policies under NCLB are related to low-income students' achievement.

II. Findings and Conclusions of the Report

The description of the analysis is incomplete and unclear rendering the findings confusing. The report claims that across the 2003-2013 decade, they "regressed the change in NAEP scores over a given two-year period onto policy changes from the same two-year period, con-

trolling for where they started." This description implies they conducted 20 regression analyses to examine whether changes in states' standards-based policies (the predictor variables) correlated with changes in low-income students' NAEP achievement in fourth and eighth grade math and reading (the dependent variable) over five two-year time spans (2003-2005, 2005-2007, 2007-2009, 2009-2011, 2011-2013). In the appendix, they "illustrate our model" with results from only the 2009-2011 time period. The results from other time spans are not reported. Results from this one two-year period are interpreted causally as increases in standards-based reform policies cause increases (over the same time period) low-income students' NAEP achievement in fourth math and eighth grade reading to increase -- but not fourth grade reading or eighth grade math. Even then the results are significant at only the generally unacceptable 0.10 level. Furthermore, the results suggesting an effect of only two-tenths of one NAEP point gain in fourth grade math and four-tenths of one NAEP point gain in eighth grade reading is basically meaningless in light of a 26-point Black-White achievement gap[6] and a 24-point high and low income student achievement gap.[7]

It is conventional to report the multiple R when using regression approaches which tells the reader the magnitude of the relationships and the percent of the variance accounted for by their measure of policy adoption. This was not reported for any of the apparently 20 analyses. They report using a "fixed-effects" regression model that would control for within state changes to some degree but that analysis is not presented nor explained. The authors report on four occasions that they supplemented their analysis with "anecdotal" or "impressionistic" interpretations of their data. On this foundation, the report concludes that standards-based reform is an effective policy strategy for increasing low-income student achievement. As a result, the report recommends that states should continue to embrace standards-based reform and should widely adopt the Common Core State Standards to guide these efforts.

III. The Report's Rationale for Its Findings and Conclusions

The report's primary rationale is that if the scores they assigned to the "quality" of each state's standards implementation have a positive relationship with test scores, then the efficacy of standards-based reform is established. There are at least three problems with this rationale. First, as the authors admit, there is no measure of fidelity or intensity of implementation. Counting the quantity of standards adopted says little about the quality, extensiveness or fidelity of implementation. It is exceedingly difficult to objectively (and accurately) measure state-level policies *and* their implementation. Not only are policies in constant flux, but their implementation is widely uneven within and between states.[8] Including an implementation fidelity measure is even more critical when results are used to make wide sweeping policy recommendations as is the case here. A second problem is that the authors contend a causal effect based on an unreported correlation. This is inappropriate. Third, the report fails to account for numerous other variables that are known to influence student achievement such as student poverty, teacher turnover rates, and teacher

quality to name a few.[9] The absence of these considerations, the lack of valid predictors and the opaqueness of the analytic approach fundamentally invalidates the utility of the study's rationale and conclusions.

IV. The Report's Use of Research Literature

The research used to support the study relies heavily on a single study authored by Christopher Swanson and published by *Education Week*.[10] Importantly, Swanson's study is only one of a multitude of studies that have examined the connection between accountability-based policies and test results. This robust and growing literature base is notably absent from the report's review.[11]

In fact, references to peer-reviewed empirical work that support or refute claims made throughout the report are sparse, outdated, or are descriptive rather than analytical. In the two pages dedicated to reviewing what we know about the effectiveness of standards-based reform, the report provides only 11 endnotes. Only one peer-reviewed empirical article was cited and this was a 2002[12] report. Three newspaper articles, three think tank reports (two of which are from the Center for American Progress), one book, and one blog round out the literature review. The citations are cherry-picked in that only two sentences are donated to one study that came to a contrary conclusion. This scant review ignores a rich literature that underscores much of what we know about the uneven standards-based policy implementation throughout the NCLB era, as well as the inherent challenges and cautions associated with the literature's mixed conclusions.[13]

V. Review of the Report's Methods

As noted earlier, the methods suffer from (1) a lack of precision in describing key variables in the analysis, (2) the absence of key control variables in the analysis, and (3) selective reporting of results.

Standards-Based Policy Variable

For every year in the study (2003-2013), a state was given a single policy "score" that allegedly reflected the relative strength of each state's standards-based policies. This score was based on three factors: the extent of standards adoption (subject matter and grade levels), the degree to which state standards were aligned with tests, and the number of test-based consequences mandated by state policy. A higher score meant "stronger" standards-based policy efforts.

These scores were the result of a combination of indicators compiled by the Education Counts Research Center (and sponsored by *Education Week.)*[14] In the standards category, states could earn anywhere from 0-10 points across five different indicators, each of which have to do with whether standards are "clear, specific, and grounded in content." States earned higher points if standards met this criterion in a greater number of subject areas and grade levels. In the assessment category, states could earn 0-6 points, earning more points for having tests aligned in more subject areas and for having tests that included short answer/extended response opportunities. Finally, the accountability category could earn a state from zero to four points with higher scores equating to a higher number of test-based sanctions (e.g., state sanctions low performing schools, requires exit exams, has report cards and/or rewards high-performing schools). Theoretically, then, a state could earn anywhere from 0-20 points based on how their policies were judged in any given year.

Importantly, the varied distributions, relative importance, and weights of these three categories mean that simply adding scores together is problematic. The report acknowledges this but used it anyway. The report fails to clearly explain their approach or rationale for combining the measures. All that is provided is that they "normed the results for each category to a final 10-point scale and averaged them."[15] There are various ways of weighting multiple measures into a single score (normalizing, factor analysis, professional judgment, etc.) but the reader is left adrift on this important point.

In addition to not knowing how the policy predictor scores were computed, another perhaps even more vital problem has to do with meaningfulness of those scores. Assuming the transformation calculations were reliable, *the report's failure to account for fidelity of policy implementation creates a fundamental validity problem.* Apparently, gains were ascribed and causal factors were confirmed based on eyeballing the data.

Regression Analysis

The report conducted regression analyses to see if changes in standards-based policy over time resulted in changes in low-income students' NAEP performance in fourth and eighth grade math and reading. Using a regression model with a total n of only 50 is not recommended practice. The variables entered into the analyses are not well described nor even listed with weights in the conventional manner. The number of variables appears to be perplexingly small in number consisting of three standards aggregate scores, an undefined state NAEP score from two years earlier, and a finance measure that weighed in at 0.0000.

Somewhat puzzling was the use of a school funding equity measure (restricted range) that "expresses the difference between expenditures in the high- and lowest- spending districts" as a control.[16] In such a model, the predictor and control variables should have some logical relationship to the question at hand, however, no rationale is provided for the inclusion of this measure as a control except its availability from *Education Week*. The restricted range is commonly used and measures finance inequities. The authors fail to include additional controls.

Lastly, the report says a fixed-effects regression model was used that would statistically control for within state variation. Although this logic is reasonable, there is no further description or reporting of data describing or presenting such an approach.

Missing Information

As noted earlier, the regression results are not presented, the report provides only results from one set of analyses (2009-2011) out of a possible set of five, no correlation matrix is provided, no Multiple R is reported and variables and their weights are not provided. What makes this state of affairs even more troubling is that the report acknowledges the paucity of their findings but uses it as justification for substituting an "anecdotal" and "impressionistic" analysis that is selective and incomplete. For example, in a review of the data patterns over time, the report notes that states such as "Kansas, Iowa, Idaho, Montana, and North and South Dakota showed some of the lowest gains on NAEP for low-income students since the late 1990s, and broadly speaking, these states have been less committed to standards-based reform."[17] These summary conclusions were not accompanied by data, making it impossible to evaluate these interpretations. The authors do note that confirmation bias may be at play.

VI. Review of the Validity of the Findings and Conclusions

There are significant and fundamental flaws in this report. The authors employ inappropriate research methods, fail to adequately define their approach, and do not report the findings from their analyses. Even if the limited data was taken at full value, the effects are trivial from a policy perspective. As a result, the authors fail to substantiate their conclusions. In short, the analysis does not provide evidence that "stronger" standards-based reform *causes* increases (or is even related to) low-income students' achievement. The analysis is simply too incomplete to warrant such a finding.

Their conclusion that the report justifies widespread adoption of the Common Core State Standards is not substantiated even on its own terms. The Common Core State Standards were not cycling through state adoption until 2010, are yet to be externally validated, have been adopted quite unevenly over time, and implementation has been haphazard.[18] The report's presentation that its results justify support for the CCSS in general is simply not sustained.

VII. Usefulness of the Report for Guidance of Policy and Practice

This report does not add to discussions of policy or practice. Even if the omissions and shortcomings of this report were remedied, the analysis only provides a very narrow snapshot of

how policy might connect to practice. In this report, we only see one set of results from one point in time (and mostly before Common Core Standards were adopted, i.e., 2009-2011), but we are asked to rely on their unsubstantiated claim that standards-based policies work. The strident call for Common Core at the end of this report is misplaced given the mismatched goals, questionable analysis and selected findings from the report.

Notes and References

1   Boser, U., & Brown, C (2016, January 14). *Lessons from state performance on NAEP: Why some high-poverty students score better than others*. Washington, D.C.: Center for American Progress. Retrieved January 20, 2016 from https://cdn.americanprogress.org/wp-content/uploads/2015/12/23090515/NAEPandCommonCore.pdf

2   Herman, J.L. & Haertel, E.H. (Eds.) (2005). Uses and misuses of data for educational accountability and improvement. *The 104th Yearbook of the National Society for the Study of Education (part 2)*. Malden, MA: Blackwell.

3   Elementary and Secondary Education Act of 1965 (20 U.S.C. 6301 et seq.).

     No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 *et seq.* (West 2003)

     Every Student Succeeds Act (ESSA). PL No 114-95. Retrieved February 1, 2016 from https://www.congress.gov/bill/114th-congress/senate-bill/1177/text

4   National Commission for Excellence in Education (1983, April). *A nation at risk: The imperatives for educational reform*. Washington, D.C.: U.S. Department of Education, National Commission for Excellence in Education. Retrieved February 10, 2016 from https://www2.ed.gov/pubs/NatAtRisk/risk.html

5   Goals 2000, Educate America Act. 1994. Public Law 103-227.

     No Child Left Behind (NCLB) Act of 2001, 20 U.S.C.A. § 6301 *et seq.* (West 2003)

     U. S. Department of Education (November 2009). *Race to the Top: Executive Summary*. Washington, D.C.: U.S. Department of Education. Retrieved February 1, 2016 from http://ed.gov/programs/racetothetop/executive-summary.pdf

6   NCES (July 2009). *Achievement Gaps: How Black and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress*. Retrieved February 9, 2016 from http://nces.ed.gov/nationsreportcard/pubs/studies/2009455.aspx

7   Based on data retrieved February 9, 2016 from http://nces.ed.gov/nationsreportcard/naepdata/report.aspx. These data come from 2011 NAEP administration to coincide with the report's study years. The high- and low-income student achievement gap in 2011 is 23 points in fourth grade math (which stayed the same 2009-2011) and 24 points in eighth grade reading (also staying relatively stable across 2000-2011).

8   Dee, T. & Jacob, B. (2009, November). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446.

     Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Educational Policy Analysis Archives, 12*(1), 1-40. Retrieved February 3, 2016 from http://epaa.asu.edu/epaa/v12n1/

     Nichols, S.L., Glass, G.V, Berliner, D.C. (2012). High-stakes testing and student achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives, 20*(20) Retrieved September 16, 2012, from http://epaa.asu.edu/ojs/article/view/1048

9   Berliner, D.C. (2013). Effects of inequality and poverty vs. teachers and schooling on America's youth. *Teachers College Record, 116*(1). Retrieved August 28, 2014 from http://www.tcrecord.org

Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives, 8*(1). Retrieved February 3, 2016 from http://dx.doi.org/10.14507/epaa.v8n1.2000

Ronfeldt, M., Loeb, S., & Wyckoff, J. (2013). How teacher turnover harms student achievement. *American Educational Research Journal, 50*(1), 4-36.

10 Swanson, C. (2006). *Making the connection: A decade of standards-based reform and achievement.* Bethesda, MD: Editorial projects in Education Research Center. Retrieved January 20, 2016 from https://www.edweek.org/media/ew/qz/2006/Making theConnection.pdf

11 Amrein, A.L. & Berliner, D.C. (2002b). High-Stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18). Retrieved December 1, 2014 from http://epaa.asu.edu/epaa/v10n18/

Amrein-Beardsley, A. & Berliner, D. (2003, August). Re-analysis of NAEP Math and Reading scores in states with and without high-stakes tests: Response to Rosenshine. *Education Policy Analysis Archives, 11*(25). Retrieved February 5, 2005 from http://epaa.asu.edu/epaa/v11n25/

Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Educational Policy Analysis Archives, 12*(1), 1-40. Retrieved January 20, 2016 from http://epaa.asu.edu/epaa/v12n1/

Dee, T., & Jacob, B. (2009, November).The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446.

Holme, J.J., Richards, M.P., Jimerson, J.B., & Cohen, R.W. (2010). Assessing the effects of high school exit examinations. *Review of Educational Research*, *80*(4), 476-526. doi:10.3102/0034654310383147

Jones, M.G., Jones, B., and Hargrove, T. (2003). *The unintended consequences of high-stakes testing.* Lanham, MD: Rowman & Littlefield.

Lavigne, A.L. (2014). Exploring the intended and unintended consequences of high-stakes testing evaluation on school, teachers, and students. *Teachers College Record*, 116, 1-29.

Lavigne, A.L., & Good, T.L. (2014). *Teacher and student evaluation: Moving beyond the failure of school reform.* NY: Routledge.

Nichols, S.L., Glass, G.V., & Berliner, D.C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives, 14*(1). Retrieved July 20, 2009, from http://epaa.asu.edu/epaa/v14n1/

Nichols, S.L., Glass, G.V, Berliner, D.C. (2012) High-stakes testing and student achievement: Updated analyses with NAEP data. *Education Policy Analysis Archives, 20*(20) Retrieved September 16, 2012, from http://epaa.asu.edu/ojs/article/view/1048

Rosenshine, B. (2003). High-Stakes testing: Another analysis. *Education Policy Analysis Archives, 11*(24). Retrieved January 20, 2016 from http://epaa.asu.edu/epaa/v11n24/

Winters, M.A., Trivitt, J.R., & Greene, J.P. (2010). The impact of high-stakes testing on student proficiency in low-stakes subjects: Evidence from Florida's elementary science exam. *Economics of Education Review, 29*, 138-146. doi: 10.1016/j.econedurev.2009.07.004

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

13   Braun, H. (2004). Reconsidering the impact of high-stakes testing. *Educational Policy Analysis Archives, 12*(1), 1-40. Retrieved January 20, 2016 from http://epaa.asu.edu/epaa/v12n1/

Dee, T., & Jacob, B. (2009, November).The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management, 30*(3), 418-446.

Nichols, S.L., Glass, G.V., & Berliner, D.C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives, 14*(1). Retrieved July 20, 2009, from http://epaa.asu.edu/epaa/v14n1/

Herman, J.L. & Haertel, E.H. (Eds.) (2005). *Uses and misuses of data for educational accountability and improvement*. The 104th Yearbook of the National Society for the Study of Education (part 2). Malden, MA: Blackwell.

Swanson, C.B., & Stevenson, D. L (2002). Standards-based reform in practice: Evidence on state policy and classroom instruction from the NAEP state assessments. *Educational Evaluation and Policy Analysis, 24*(1), 1–27.

Timar, T.B. and Maxwell-Jolly, J. (Eds) (2012). *Narrowing the achievement gap: Perspectives and strategies for challenging times*. Cambridge, Massachusetts: Harvard Education Press.

14   Education Counts indicators are available online and can be retrieved from their publically available website; http://www.edcounts.org/createtable/step1.php

15   Quote from p. 10: Boser, U. & Brown, C. (2016, January). *Lessons from state performance on NAEP: Why some high-poverty students score better than others*. Boulder, CO: National Education Policy Center. Retrieved January 20, 2016 from https://cdn.americanprogress.org/wp-content/uploads/2015/12/23090515/NAEPandCommonCore.pdf

16   Quote from p. 20: Boser, U. & Brown, C. (2016, January). *Lessons from state performance on NAEP: Why some high-poverty students score better than others*. Boulder, CO: National Education Policy Center. Retrieved January 20, 2016 from https://cdn.americanprogress.org/wp-content/uploads/2015/12/23090515/NAEPandCommonCore.pdf

17   Quote from p. 12: Boser, U. & Brown, C. (2016, January). *Lessons from state performance on NAEP: Why some high-poverty students score better than others*. Boulder, CO: National Education Policy Center. Retrieved January 20, 2016 from https://cdn.americanprogress.org/wp-content/uploads/2015/12/23090515/NAEPandCommonCore.pdf

18   Association for Supervision and Curriculum Development. (2012). *Fulfilling the promise of Common Core State Standards: Moving from adoption to implementation to sustainability*. Retrieved January 20, 2016 from http://www.ascd.org/ASCD/pdf/siteASCD/commoncore/CCSSSummitReport.pdf

Graham, S., & Harris K.R. (2015). Common Core and state standards and writing: Introduction to the special issue. *Elementary School Journal, 115*(4), 457-463.

Sforza, D., Tienken, C., & Kim, E. (2016, Winter). A comparison of high-order thinking between the common core state standards and the 2009 New Jersey content standards in high school. *AASA Journal of Scholarship and Practice, 12*(4), pp. 5-31. Retrieved February 1, 2016 from http://www.aasa.org/uploadedFiles/Publications/JSPwinter_FINAL.v10Jan16.v3.pdf

SUGGESTED CITATION:

Nichols, S.L. (2016). Review of "*Lessons From State Performance on NAEP: Why Some High-Poverty Students Score Better Than Others.*" Boulder, CO: National Education Policy Center. Retrieved [date] from http://nepc.colorado.edu/thinktank/review-CAP-standards