



REVIEW OF *THE LESSONS OF NO CHILD LEFT BEHIND*

Reviewed By

Bruce Fuller

University of California, Berkeley

August 2013

Summary of Review

This new report from the American Enterprise Institute, a free market think tank, attempts to tease out the effects of No Child Left Behind (NCLB) on North Carolina achievement scores. In particular, the report attempts to isolate the effects of NCLB's threat of sanctions placed on underperforming schools. While generally skeptical that NCLB has discernibly lifted learning over the past decade, the report does ascribe discrete yet small achievement gains to the sanction provisions of the law. Based on a brief literature review and the modest effects from sanctions found in North Carolina in this new study (effect size of 0.05 of a standard deviation), the authors infer that federal pressure and punishment are promising policy avenues. While the report claims that NCLB's specific policy levers can be definitively estimated amidst all the collateral policy noise, the reader learns little about how local educators comprehend or respond to federal and state accountability pressures. It is notable that this report from a prominent conservative think tank signals the importance of federal leadership and quality control from Washington. What's missing from the report—limiting the utility of the analysis—is that readers come away with little understanding of what elements of NCLB have lifted students and which have failed.

Kevin Welner

Project Director

William Mathis

Managing Director

Erik Gunn

Managing Editor

National Education Policy Center

School of Education, University of Colorado
Boulder, CO 80309-0249
Telephone: (802) 383-0058

Email: NEPC@colorado.edu
<http://nepc.colorado.edu>

Publishing Director: **Alex Molnar**



This is one of a series of Think Twice think tank reviews made possible in part by funding from the Great Lakes Center for Education Research and Practice. It is also available at <http://greatlakescenter.org>.

This material is provided free of cost to NEPC's readers, who may make non-commercial use of the material 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.

REVIEW OF *THE LESSONS OF NO CHILD LEFT BEHIND*

Bruce Fuller, University of California, Berkeley

I. Introduction

The title of a new study published by the American Enterprise Institute, *Were all those standardized tests for nothing? The Lessons of No Child Left Behind*, expresses skepticism while also trying to pinpoint what worked over the past decade.¹ This comes at a time when the Obama administration continues to disassemble key NCLB provisions and return to a federalist structure. States and, of late, urban districts are winning waivers that will

The AEI report's empirical study concludes that sanctions led to a minuscule bump in student achievement.

preserve authority of setting “proficient” levels of learning, the source of much mischief by governors and state education chiefs over the past decade. Meanwhile the administration has centralized its authority, without congressional authorization, to tie test scores to teacher evaluation, tell governors to lift caps on charter school growth, and incent states to adopt the Common Core curriculum.

The report already feels a bit out of date. NCLB was up for renewal in 2007, but Congress has been unwilling to seriously consider how to recast the federal role, so education secretary Arne Duncan’s reform agenda has largely moved past NCLB. Further dating the report’s policy relevance, House Republicans passed legislation this summer that would detonate any strong federal presence, trashing several of the No Child ingredients that Ahn and Vigdor are earnestly and carefully trying to weigh.

Amidst this shifting context, the report (1) unpacks the policy levers contained within No Child, (2) reviews summative empirical attempts to gauge effects of the law and (3) offers an original though narrow analysis of how punitive sanctions played out in low-performing North Carolina schools. Interestingly, the authors themselves warn that obsessing on punishments rather than incentives may work about as well in the policy world as does harsh parenting at home.

II. Findings and Conclusions of the Report

Discerning Discrete Effects from Noisy Policy

The report begins by arguing that state accountability regimes “and NCLB in particular have beneficial systemic effects on standardized test scores” (p.1). The authors recognize the collateral damage done by unbalanced high-stakes testing, such as inattention to science and the arts, along with the harm of hyper-didactic pedagogy (my phrase, not theirs). But the report never details evidence of the negative effects of high-stakes accountability inside classrooms, nor returns to this critical evidence when trumpeting small effects from punitive sanctions.

As detailed in the literature review discussion below, the report’s rather selective review highlights two broad studies that found discernible achievement gains—at least in mathematics among fourth-graders—in states that sported only weak or no standards-based accountability regimes. In contrast, an eye-opening historical analysis of student progress was released by NAEP this summer, just two months following the Ahn-Vigdor report, showing that, yes, important progress has been made in lifting achievement for fourth- and eighth-graders, and also showing dramatically narrowing racial achievement gaps since the early 1970s.² But much of this progress occurred in the early 1990s in the wake of school finance reform and earlier *state-led* standards efforts, long before enactment of No Child. More sobering, all progress in lifting achievement of the average child and narrowing learning disparities has stalled nationwide since 2008. Of course, drawing causal conclusions that link NAEP’s cross-sectional results to specific policies is problematic—as is the attempt made in the new AEI report, as discussed further below.

Unpacking the No Child Policy Bundle

A helpful contribution of the report concerns how the authors unpack the various policy levers that once made up the NCLB apparatus. Policy, the report points out, can boost student learning by jamming local educators to squeeze greater efficiency out of their production units (the classroom). In this light, the report delineates several elements of No Child that should be subjected to empirical assessment, examining whether those elements motivate students or teachers to lift the schools. They include the following:

1. the mere threat of sanctions presumably felt by low-performing schools;
2. how the sanctions are carried out by states and districts, such as reconstituting a school’s entire staff;
3. state-defined thresholds of “proficient” achievement and increased testing to gauge progress;
4. federally defined performance targets for districts, schools, and student subgroups (Annual Yearly Progress, or AYP, benchmarks), which “forced schools to change the way they allocated resources to educate students who were falling through the cracks” (p. 9);

5. the grant to parents of greater discretion, such as allowing students to transfer out of low-performing schools, and the provision of tutoring and supplementary services for students attending “program improvement” schools;
6. the push for states to purge uncredentialed teachers from classrooms, requiring that all teachers be ‘highly qualified’; and
7. the provision of additional resources to states to improve tracking students and their achievement levels.

The report highlights that positive incentives were created by some statehouses, including North Carolina’s collective cash payments to schools that show growth (not merely the attainment by pupils of “proficiency”). The report also candidly acknowledges when certain policy levers proved ineffective, such as the minimal level of interest by parents in moving their children out of “program improvement” schools. While the report does not explore the reasons for this ineffectiveness (in this case, the culprit may be the social milieu or school-level practices that proved too sticky, so unfathomable by Washington policy makers), the acknowledgement itself is important to the study.

Looking forward, the authors point out that, in the absence of engagement by Republican leaders, the Obama administration is no longer really trying to make this policy mix work better: “Federal initiatives in education policy have instead moved on, with the Obama administration’s championing its Race to the Top initiative rather than expending political capital on NCLB” (p. 3). This undoubtedly makes the analytic task of teasing-out specific effects of NCLB all the more difficult, as districts are now being pushed to follow the Obama/Duncan approaches: evaluate teachers more rigorously, expand charter schools, and innovate in untold ways—with few touch points back to the aging policy levers of NCLB.

Do Punitive Sanctions Work?

From this broad panorama the report then sets its empirical sights on the narrow question of whether the threat or execution of sanctions moved North Carolina schools to improve, or at least to lift a few more of their students over the “proficiency” hurdle. It examines, over the 2002-2010 period, the achievement response of schools that were threatened by, or were actually hit with, sanctions when they missed their federal growth target, compared with otherwise similar schools that barely avoided the trigger.

When schools failed to hit their federal growth target over four consecutive years, the state or local district was supposed to apply sharp penalties, including replacing the principal, reconstituting the teaching staff, or handing the school over to a charter management group. This regimen was required until the Obama administration granted North Carolina a waiver in May of 2012, as part of its “flexibility” policy that permits states to waive out of No Child provisions if they adopt a package of specified reforms.

The authors ask whether the imposition or threat of such sanctions led to one of two school responses thought to be valuable: the shifting of resources to low-performing

students, or organizational changes that resulted in achievement gains. The report does not, however, inform readers about specific ways in which school staff responded, which limits the utility of the authors' summative analysis.³ How is the threat of sanctions understood and responded to by principals and teachers inside schools? The report fails to shine a light on these crucial organizational processes.

The AEI report's empirical study concludes that sanctions led to a minuscule bump in student achievement. Comparing schools just above and below the sanction-trigger (using a regression-discontinuity design, or RDD), Ahn and Vigdor found that those actually sanctioned (not simply threatened) displayed a 0.05-standard-deviation gain in the following year, a statistically significant but minuscule boost for largely disadvantaged students. This effect size approximates the tepid impact of California's ill-fated effort to reduce class size, as estimated by RAND analysts.⁴

The report also found that gains were centered among children falling just below the proficiency line, raising the question of whether sanctions truly benefit school populations more broadly or if they merely yield temporary bumps for a targeted subset. The RDD design also sheds little light on the possible effects of NCLB policies on the bulk of low-performing schools that rest further from the cut-point that triggers sanctions. Inferences about these schools would stretch the limits of this quasi-experimental method. But the reader is left with the sinking suspicion that, given the small possible effects in schools in the sanction bull's-eye, the threat and execution of sanctions had tiny effects at best on other schools. It's also important to note that the outcomes measured in this study are not comprehensive. In addition to long-term benefits or harms and to the issues of hyperdidactic pedagogy and narrowed curriculum mentioned in the report, other possible policy effects include teacher turnover, student exit, and the shuffling of low-quality teachers to other poor communities.

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

The report builds its argument—that specific elements of NCLB may have yielded discrete benefits—using prior research (see Section IV, below). But the literature review only goes in-depth when it comes to broad summative assessment of one question: whether NCLB, writ large, moved test scores or not. Based on this brief literature review and the modest effects from sanctions found in North Carolina in this new study, the authors infer that federal pressure and punishment are promising policy avenues.

But again the sticky complexities of local implementation of sanctions, whether replacing a principal, reconstituting staff, or converting to a charter school, are of little interest to these authors. And how Ahn and Vigdor draw a simple causal line between policy made in Washington and pupil achievement effects feels naïve. If only large institutions were so translucent. That said, the encouraging rebound of schools under pressure of sanctions may inform adjustments to accountability policy.

The report also embraces a wider rationale that presents NCLB as incorporating “accountability and autonomy” (p. 3). Under this loose/tight organizational model, the role of centralized government is to define shared learning goals and monitor student progress, while leaving the means of accomplishing these goals up to local education agencies. The contention has some validity as regards NCLB itself, but in the larger sense it is unconvincing because many state governments continue to micro-manage what textbooks schools may purchase, frequently emphasize didactic delivery of official knowledge, and operate a regulatory thicket fed by categorical funding streams. In fact, some prominent recent reforms concerning, for instance, the evaluation of teachers and principals have intensified the mandates emanating from federal and state government. The report also does not address whether quasi-market actors, like charter or pilot schools, have shown positive spill-over benefits on traditional public schools. That is, the idealized model of a lean central government and deregulated freedom for the local actors is never set against institutional complications on the ground.

IV. The Report’s Use of Research Literature

The report provides a thin and incomplete review of the empirical literature, offered in two parts. The second part is less noteworthy, so I will address it briefly before moving to the first. The report gives cursory evidence on the specific components of NCLB that have received some empirical attention, especially prior work on the impact of school-level sanctions or incentives for growth. These summaries are helpful, but they uncritically assume that discrete effects can be tied to a single policy tool, when each is wielded amidst a variety of other policies. There’s no pristine experiment here, nor sufficient cases (states, districts) that lie close on either side of bright discontinuity line.

The first part summarizes national studies that have attempted to estimate discrete effects of NCLB, centering on states with weak standards-based accountability regimes prior to 2002. This review of summative estimation is a useful element of the report. Stanford economists Eric Hanushek and Margaret Raymond, for example, found that test scores, at least in fourth- and eighth-grade mathematics, rose in states that set in place consequences for schools that failed to meet federal growth targets, compared with states with no or low-stakes consequences prior to enactment of No Child.⁵

Hanushek and Raymond view NCLB enactment as a massive natural experiment, particularly in states with historically weak accountability policies, and they usefully draw on longitudinal testing data from the National Assessment of Educational Progress, NAEP. Their results are consistent with an earlier analysis of state-led accountability by Susanna Loeb and Martin Carnoy, years before No Child was enacted.⁶ Most recently, Tom Dee and Brian Jacob report significant gains in mathematics, but not in reading, among Black and Latino fourth-graders, based on additional years of post-NCLB experience.⁷ Unfortunately, these broad-brush summative studies help little in understanding how, or through what mechanisms, No Child helped to lift test scores.

My own early, descriptive work detailed how the initial NCLB bump observed in the post-2002 period was discernible in math and situated mainly among fourth-graders, also drawing on NAEP results, with weak to no discernible change in reading or among eighth- and 11th-graders.⁸ The long-term NAEP analysis released this past summer bears out these findings, except that discernible gains in reading did appear between 2004 and 2008 for eighth-graders, and racial achievement gaps have narrowed quite dramatically since 1971. The bad news is that all progress in elementary and middle schools has petered out since 2008, and only high school students in the lowest quartile of achievement have shown any progress over the past two generations.⁹

As noted above, the literature review failed to include or address the large body of literature that details the downside costs of high-stakes accountability tied to standardized testing. Although it briefly mentions concerns over teaching to the test and giving short shrift to non-tested subjects like science and civics, the underlying tenets of No Child are never questioned. Instead, the report frames a utilitarian problem of how to tinker with elements of federal policy that might boost test scores in English and mathematics.

V. Review of the Report's Methods

A naïve design assumption surfaces repeatedly throughout the report—that national or state-level achievement trends can be attributed to a complicated bundle of policy tools implemented simultaneously across the country. As described above in Section IV, the report mixes work that attempts to isolate a specific policy tool—e.g., threat of school

The authors usefully remind us of how complicated NCLB once was... But readers will not come away understanding how this pantheon of policies is understood and acted upon inside classrooms.

sanctions, positive rewards for growth, supplemental tutoring services—with the confident attempts by Eric Hanushek and others to estimate the big-bang, summative effect of a complicated federal policy bundle that unfolded in wildly diverse states, as governors kept piling on their own rules, materials, and programs.

The study's approach contributes to the body of research using the first kind of design, illuminating the utility of regression-discontinuity techniques to suggest discrete effects from a particular policy tool. Even here the report acknowledges possible confounding dynamics, stemming in particular from North Carolina's own positive incentive scheme for schools showing growth. But the report seems to accept uncritically the sweeping, estimations of studies that ask the comprehensive question: Did NCLB make a difference, yes or no? They may become historical studies, but going forward it would be nice to learn about the comparative effects and costs vs. benefits of the differing components of NCLB.

The report itself—published in manuscript form by the American Enterprise Institute—appears to be under peer review (endnote 12, appearing on page 27). Reputable think tanks often require peer review before publication of empirical findings.

VI. Review of the Validity of the Findings and Conclusions

So, what are the major take-aways from this report? The new chunk of evidence pertains to the specific question of whether punitive sanctions lifted student achievement in one state, North Carolina. The empirical estimates suggest, yes, the policy worked in this way at a very low level of magnitude and amidst a flurry of other policy activity. This narrow finding appears among collateral analyses that, while thin in supporting evidence, remain useful going forward.

First, the authors usefully remind us that NCLB was a complicated, at times orthogonal set of policy levers. These levers were roughly tied to standards-based accountability, to wildly varying state conceptions of “proficient” learning, and to a set of consequences for schools, their teachers, and students. The authors frankly acknowledge that the Obama administration has largely moved on, detaching its narrow reforms manifest in the Race to the Top initiative from the pressing question of whether a coherent, long-term federal role can be crafted in bipartisan fashion.

Second, the authors accent an eye-opening fact: while some evidence suggests that standards-based accountability has raised test scores, at least for younger students, we have little evidence concerning which elements of NCLB deserve credit for any such benefits. In fact, any noted gains may be attributable to other initiatives: expansion of early education, rising school attainment of Black mothers, or a panoply of state-level programs, all coincident with post-2002 implementation of No Child. The report provides no help in thinking about these dynamics causally, what discrete policies can be tested, and which are so locally variable that evaluation would be futile. The report focuses on testing possible effects of sanctioning low-performing schools, and it comes up short empirically. The estimated effect on achievement (0.05 standard deviation) is minuscule, and may result from collateral state policies, including North Carolina’s rewarding of achievement growth as the report acknowledges.

Third, the authors put forward recommendations for how to possibly improve NCLB—recommendations that are not really tied to the report’s empirical analysis (pages 22-26). Most of these recommendations agree that accountability regimes should focus on student growth, not between-cohort analyses. The recommendations also break with many current accountability advocates by arguing for the preferability of collective rewards to schools displaying growth, rather than bonuses for individual teachers. Other recommendations contend that school districts should retrofit poor teachers, not fire them; and districts should be given more freedom to fashion rewards, sanctions, and school improvement strategies. While these recommendations may merit serious consideration, most do not stem from the report’s empirical analyses.

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

This report is worthwhile, potentially benefiting readers no matter their political stripes or policy inclinations. The authors usefully remind us of how complicated NCLB once was, in all its glory. A careful reading of the report should leave readers with a good understanding of the crucial point that we don't know how this federal thrust may have contributed to test score gains between 2002 and 2008 (before scores hit a depressing plateau). This understanding is particularly important as Washington's cacophony of reform thrusts continues with little evidence of what did and did not work over the past decade. The authors' gallant effort to detect achievement blips from sanctioning schools in North Carolina confirms that punishing educators on the ground isn't likely to yield remarkable effects. Hopefully, the Calvinist-leaning architects of No Child will remember this lesson.

Equally notable is what's missing from the inside-the-beltway logic of the report. A tacit yearning for the past seems to call out to the authors—a desire to simply tweak certain policy levers, some of which have already been abandoned. The report's message to readers is that if we could just sharpen the threat of school sanctions, boost collective rewards to schools prompting growth, and allow states and districts to experiment with novel renditions of enforcement, we would likely see uplifting benefits.

But readers will not come away from this report understanding how this pantheon of policies is understood and acted upon inside classrooms. We see little recognition that schools are complex organizations staffed by diverse teachers seeking meaning and intrinsic rewards from their everyday work. Parents are to respond “rationally” to information and market options, to move their kids outside their neighborhoods or buy tutoring services when their schools fail. And somehow teachers will jump to avoid sanctions or win modest cash payments. From inside this Washington think tank, the nature of motivation and human-scale behavior inside schools apparently seems so simple, so easy to manipulate.

Overall, what's useful about this report is the nudge to disaggregate the key components of NCLB, empirically assess each, and incorporate those empirical lessons into the next generation of federal (or state) efforts. The report does not go very far in matching these aspirations, but the authors are asking the right questions.

Going forward, some lawmakers are finally admitting that calling out and slapping educators on the hand makes for dumb politics and even dumber motivational theory. Meanwhile, the House bill passed this summer would disembowel the federal role in education, permanently burying the persisting remnants of NCLB. Whether the political timing is right for bipartisan consensus on the federal role in education remains to be revealed. Perhaps Ahn and Vigdor contribute by considering a range of policy instruments that neither side in Washington is eager to fetch from their rusting tool kits, no longer so confident in how to lift the schools. Neither the House nor Senate bills speak of punitive sanctions, and student-growth regimes will likely be under the control of states, not Washington.

Readers of the AEI report will benefit from the information provided, but they will remain uninformed as to how the search for an effective federal role can be informed by scarce evidence on the effects of various elements of NCLB. While the report usefully points to positive rewards for schools showing growth, and it suggests that pulling the trigger on punitive sanctions may yield tiny benefits and that requiring tutoring interventions show nil effects.¹⁰ Yet the report's overall bearing does not veer from the status quo. It instead displays a faith in government pressure and top-down accountability, which will likely continue in more motivating forms, including Common Core standards and state incentives for student growth. The report's desire to learn empirically about what did work during the NCLB era is laudable. But we still await richer, more balanced assessments.

Notes and References

1 Ahn, T. & Vigdor, J. (2013). *Were all those standardized tests for nothing? The lessons of No Child Left Behind*. Washington DC: American Enterprise Institute. Retrieved August 26, 2013, from <http://www.aei.org/papers/education/k-12/accountability/were-all-those-standardized-tests-for-nothing-the-lessons-of-no-child-left-behind/>.

2 National Assessment of Educational Progress (2013). *Trends in academic progress: Reading, 1971-2012; Mathematics, 1973-2012*. Washington, DC: United States Department of Education, Report 2013-456.

3 An example of a report that does include such an analysis is:

Rouse, C., Hannaway, J., Goldhaber, D., & Figlio, D. (2007). *Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressures* (Working Paper 156). Princeton: Center for Economic Policy Studies.

See also:

Betebenner, D. (2008). *Review of "Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressures."* Boulder, CO: National Education Policy Center. Retrieved August 26, 2013, from <http://nepc.colorado.edu/thinktank/review-feeling-florida-heat-how-low-performing-schools-respond-voucher-and-accountability-/>.

Cecilia Rouse and her colleagues detailed school-level adjustments when Florida principals heard that their school may be placed under the state voucher program, organizational shifts that did benefit low-achieving students.

4 Stecher, B. & Bohrnstedt, G. (2000). *Class size reduction in California: Summary of the 1998-99 evaluation*. Santa Monica: RAND Corp.

Note that other, better designed, targeted and implemented class-size reduction reforms have shown substantially greater benefits. See:

Finn, J. D. (2002). Class Size Reduction in Grades K-3. In *School Reform Proposals: The Research Evidence* (Alex Molnar, Ed.) Information Age Publishing. Retrieved August 26, 2013, from <http://nepc.colorado.edu/files/Chapter02-Finn-Final.pdf/>.

5 Hanushek, E. & Raymond, M. (2005). Does school accountability lead to improved student performance? *Journal of Policy Analysis and Management*, 24, 297-327.

6 Carnoy, M. & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational Evaluation and Policy Analysis*, 24, 305-331.

7 Dee, T. & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30, 418-46.

8 Fuller, B., Wright, J., Gesicki, K., & Kang, E. (2007). Gauging growth: How to judge No Child Left Behind. *Educational Researcher*, 36, 268-278.

9 National Assessment of Educational Progress (2013). *Trends in academic progress: Reading 1971-2012, Mathematics 1973-2012* (NCES 2012-456). Washington, DC: National Center for Educational Statistics.

10 Deke, J., Dragoset, L., Bogen, K., & Gill, B. (2012). *Impacts of Title I Supplemental Educational Services on student achievement*, Report 2012-4054. Washington, DC: United States Department of Education.

DOCUMENT REVIEWED:

Were all those standardized tests for nothing? The Lessons of No Child Left Behind

AUTHORS:

Thomas Ahn and Jacob Vigdor

PUBLISHER/THINK TANK:

American Enterprise Institute

DOCUMENT RELEASE DATE:

May 2013

REVIEW DATE:

August 2013

REVIEWER:

Bruce Fuller, University of California, Berkeley

E-MAIL ADDRESS:

b_fuller@berkeley.edu

PHONE NUMBER:

(510) 642-9163

SUGGESTED CITATION: Fuller, B. (2013). *Review of "The Lessons Of No Child Left Behind."* Boulder, CO: National Education Policy Center. Retrieved [date] from <http://nepc.colorado.edu/thinktank/review-lessons-of-nclb/>.