

Evaluating NCLB

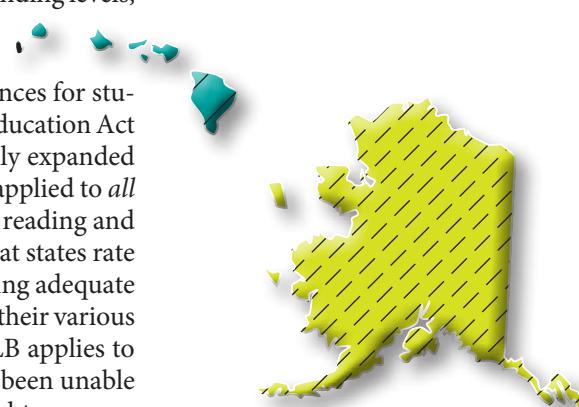
Accountability has produced substantial gains in math skills but not in reading

How has the No Child Left Behind (NCLB) Act affected student achievement? This is no idle question, as the landmark federal law is long overdue for reauthorization. The Obama administration has recently urged Congress to add the issue to its already crowded 2010 agenda, even going so far as to include an additional \$1 billion for K-12 education in its budget proposal if the law is reauthorized this year (a wholly symbolic gesture, given that it is Congress that sets spending levels, but one that indicates the administration's priorities).

Yet heightened attention to NCLB has not produced consensus over its consequences for students. No Child Left Behind was a reauthorization of the Elementary and Secondary Education Act (ESEA), the central federal legislation relevant to K-12 schooling. NCLB dramatically expanded the law's scope by requiring that states introduce school-accountability systems that applied to *all* public schools and students in the state. NCLB requires annual testing of students in reading and mathematics in grades 3 through 8 (and at least once in grades 10 through 12) and that states rate schools, both as a whole and for key subgroups, with regard to whether they are making adequate yearly progress (AYP) toward their state's proficiency goals. Supporters and critics, in their various approaches to discerning NCLB's impact, share a significant problem: because NCLB applies to all public school students, researchers lack a suitable comparison group and so have been unable to distinguish the law's effects from the myriad other factors at work over the past eight years.

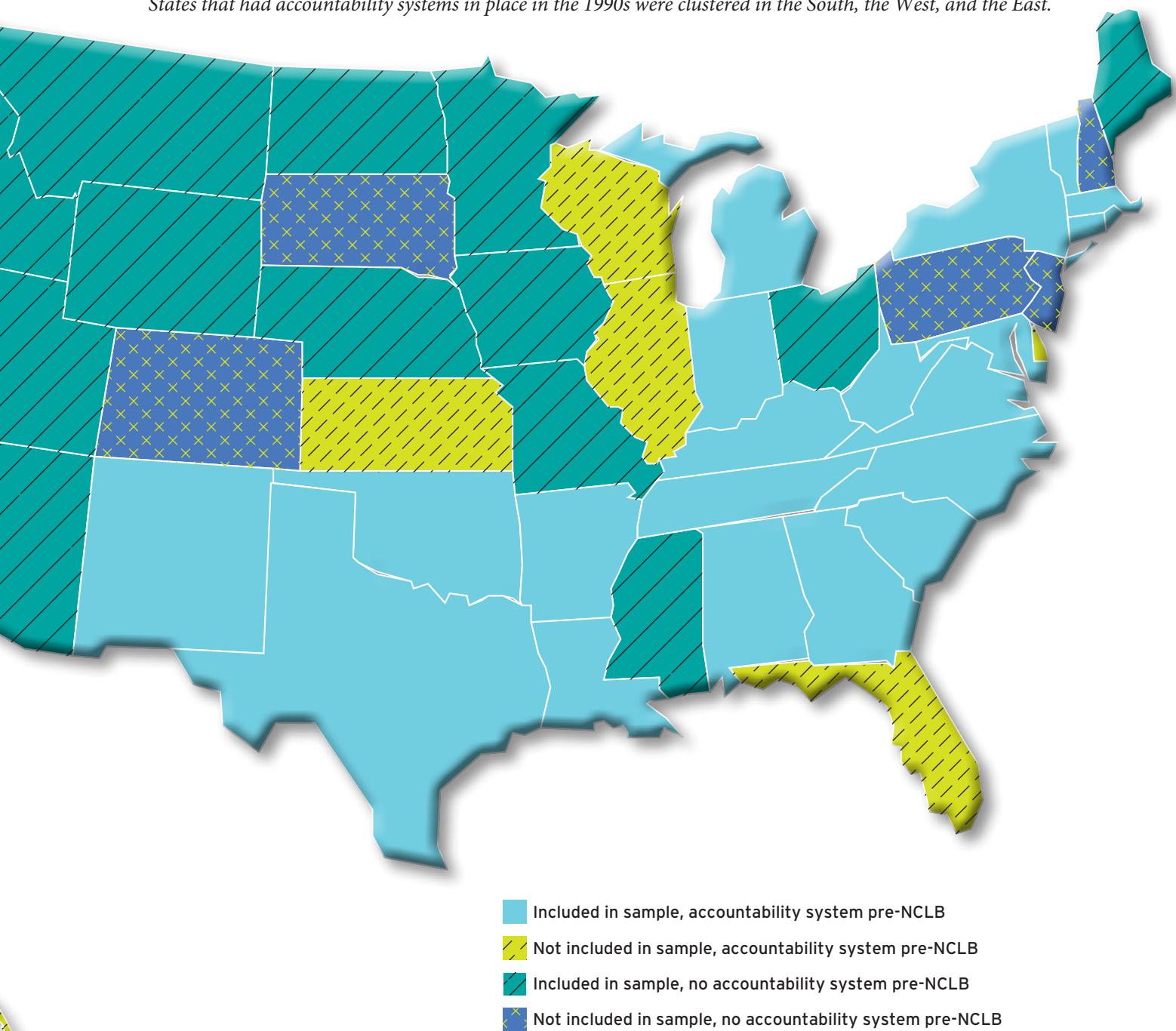
The new research we present below takes on this challenge. Our basic insight is that the test-based accountability provisions that are the defining characteristic of NCLB did not come from nowhere, but rather were modeled quite closely on reforms adopted by many states in the 1990s. For states with such accountability systems in place before 2002, NCLB's most important components may have created some logistical headaches but were largely irrelevant. In contrast, NCLB forced the remaining states to enact accountability systems for the first time. We can therefore estimate the impact of NCLB's accountability mandates by comparing test-score changes in states that did not

By THOMAS DEE and BRIAN JACOB



Sample Map (Figure 1)

States that had accountability systems in place in the 1990s were clustered in the South, the West, and the East.



Note: The map shows the analysis sample for grade 4 math. Due to differences in NAEP participation, the other samples differed from this base sample as follows. The analysis sample for grade 8 math included Illinois and excluded Iowa and Nevada. The analysis sample for grade 4 reading included Delaware, Florida, Kansas, and Washington and excluded Idaho, Indiana, Nebraska, North Dakota, Ohio, and Vermont. The analysis sample for grade 8 reading included Delaware, Florida, Kansas, and Washington and excluded Idaho, Indiana, Iowa, Michigan, Minnesota, Nebraska, North Dakota, Ohio, and Vermont.

SOURCE: Authors

ILLUSTRATION/ BRUCE SANDERS DESIGN

When Did NCLB Begin?

Exactly which academic year we should consider as the first one in which NCLB may have influenced school performance is a potentially important question. NCLB secured final congressional approval and was signed by President George W. Bush in the middle of the 2001-02 academic year. Our preferred approach is to view NCLB as first in effect during the next academic year (2002-03). NCLB is most often characterized as having been implemented during this year, in part because states were required to use testing outcomes from the prior 2001-02 year as the starting point for determining whether a school was making adequate yearly progress (AYP) and to submit draft

"workbooks" that described how school AYP status would be determined. Furthermore, state data collected during the 2002-03 year suggest that states moved quickly to adapt to NCLB's new testing requirements and to introduce school-level performance reporting.

However, one could reasonably conjecture that the discussion and anticipation surrounding the adoption of NCLB would have influenced school performance during the 2001-02 school year. Both major presidential candidates in the 2000 election had signaled support for school-based accountability, and President Bush sent a 26-page legislative blueprint titled "No Child Left Behind" to Capitol Hill within days of taking office

in January of 2001. Alternatively, it could be argued that NCLB should not be viewed as in effect until the 2003-04 academic year, when new state accountability systems were more fully implemented as well as more informed by guidance from and through negotiations with the U.S. Department of Education.

Assuming that NCLB began in 2002, or even 2001, rather than 2003, does not change our main results. However, assuming that NCLB began in the 2003-04 school year yields smaller effects (a statistically significant 0.09 standard deviations in 4th-grade math and a smaller and statistically insignificant effect in 8th-grade math).

have NCLB-style accountability policies in place when the law was implemented to test-score changes in those that did.

We find that the accountability provisions of NCLB generated large and statistically significant increases in the math achievement of 4th graders and that these gains were concentrated among African American and Hispanic students and among students who were eligible for subsidized lunch. We find smaller positive effects on 8th-grade math achievement. These effects are concentrated at lower achievement levels and among students who were eligible for subsidized lunch. We do not, however, find evidence that NCLB accountability had any impact on reading achievement among either 4th or 8th graders.

Assessing NCLB

The broad interest in understanding whether NCLB has influenced student achievement, both overall and for key subgroups, has motivated careful scrutiny of trend data from the National Assessment of Educational Progress (NAEP) and other sources. For example, the authors of a report commissioned by the U.S. Department of Education's Institute of Education Sciences (IES) note that achievement trends on both state assessments and NAEP were "positive overall and for key subgroups" through 2005. Using more recent data, a report by the Center on Education Policy concludes that reading and math achievement as measured by state assessments has increased in most states since 2002 and that there

have been smaller but similar patterns in NAEP scores. Both reports were careful to stress that these national gains are not necessarily attributable to the effects of NCLB.

Other studies have taken a less sanguine view of these achievement gains, arguing that they are misleading because states have made their assessment systems less rigorous over time. University of California scholar Bruce Fuller and colleagues, for example, document a growing disparity between student performance on state assessments and NAEP since the introduction of NCLB and conclude that "it is important to focus on the historical patterns informed by the NAEP." Using NAEP data on 4th graders, they conclude that the *growth* in student achievement has actually slowed since the introduction of NCLB.

Turning to the broader literature on school accountability, several researchers have evaluated the achievement consequences of the accountability systems states developed during the 1990s. One study by Martin Carnoy and Susanna Loeb of Stanford, which was based on state-level NAEP data, found that the within-state growth in math performance between 1996 and 2000 was larger in states with higher values on an accountability index, particularly for African American and Hispanic students in 8th grade. Another study, by Eric Hanushek and Margaret Raymond, both also at Stanford, evaluated the impact of school-accountability policies on state-level NAEP math and reading achievement measured by the difference between the performance of a state's 8th graders and that of 4th graders in the same state four years

earlier. They classified states as having either “report-card accountability” or “consequential accountability.” Report-card states provided a public report of school-level test performance. States with consequential accountability both publicized school-level performance and attached consequences to that performance. Hanushek and Raymond found that the introduction of consequential accountability within a state was associated with increases in NAEP scores.

Both of these studies suggest that NCLB-style accountability provisions may increase student achievement and also demonstrate how state-level NAEP data can be used to evaluate accountability systems. The analysis described below effectively extends this important work to cover the more recent state accountability reforms that were compelled by NCLB.

Research Design

Given the various social, economic, and educational factors at work before and after NCLB was implemented, it is difficult to draw strong conclusions about the policy’s impact from a simple comparison of achievement trends before and after enactment of the law. For example, the nation was suffering from a recession around the time NCLB was implemented, which one might expect would have reduced student achievement in the absence of other forces. At the same time, other national education policies and programs were in place that may also have influenced student achievement.

Perhaps the central challenge in evaluation research is to identify a plausible comparison group that was unaffected by the intervention under study. In the case of NCLB, this is

most distinctly in states that had not previously introduced similar policies.

This strategy relies on the assertion that pre-NCLB school-accountability policies were comparable to NCLB—that is, that the two types of accountability regimes are similar in the most relevant respects. The fact that many state officials criticized NCLB, arguing that it duplicated their prior accountability systems, suggests the functional equivalence of the two sets of policies. To ensure that this is the case and relying on a number of different sources, we evaluated the comparison states according to whether the features of their pre-NCLB accountability policies closely resembled the key aspects of NCLB. We found that they were in fact quite similar.

As an additional check on the validity of our treatment and comparison groups, we used our research design to estimate the impact of NCLB accountability on outcomes that we would not expect to be affected, such as the state-level average poverty rate and median household income. The fact that our method does not find any “effect” of NCLB on such outcomes suggests that these states can serve as a plausible comparison group for isolating the impact of NCLB accountability.

We implement our research design in a more fine-grained manner than simply comparing achievement trends in the treatment and comparison states. We define the treatment as the number of years *without* prior school accountability between the 1991–92 academic year and the onset of NCLB. Hence, states with no prior accountability have a value of 11. Illinois, which adopted its policy in the 1992–93 school year, would have a value of 2. Texas would have a value of

We find that the accountability provisions of NCLB generated large and statistically significant increases in the math achievement of 4th graders and that these gains were concentrated among African American and Hispanic students.

particularly difficult, as the policy simultaneously applied to all public schools in the United States.

We address this issue by comparing trends in student achievement across states that had varying degrees of prior experience with state school-accountability policies similar to those brought about by NCLB. The intuition behind this approach is that NCLB represented less of a “treatment” in states that had already adopted NCLB-like school-accountability policies prior to 2002. To the extent that NCLB-like accountability had either positive or negative effects on measured student achievement, we would expect, once NCLB had been implemented, to observe those effects

4 since its policy started in 1994–95, and Vermont would have a value of 9 since its program began in 1999–2000. This method implies that the larger the value of this treatment variable, the greater potential impact of NCLB. The total effect we report is the impact of NCLB accountability in 2007 for states with no prior accountability relative to states that adopted school accountability in 1997 (the mean adoption year among states that adopted accountability prior to NCLB).

It is important to note that this research design will capture the impact of the accountability provisions of NCLB, but not the impact of other NCLB components such as the

Reading First program or its Highly Qualified Teacher provisions. Additionally, our estimates will identify the impact of NCLB-induced school-accountability provisions on states without prior accountability policies. To the extent that one believes that states that expected to gain the most from accountability policies adopted them prior to NCLB, one might view the results we present as an underestimate of the average effect of school accountability.

Data

This analysis uses data on math and reading achievement from the state NAEP, which offers a representative sample of student achievement in each state at regular intervals. Participation in the state NAEP was voluntary prior to NCLB, although roughly 40 states did participate. NCLB made participation mandatory. The main advantage to using NAEP data for our analysis is that it is a low-stakes exam that is not directly tied to any state's standards or assessments. Instead, NAEP aims to assess a broad range of skills and knowledge within each subject area. Consequently, NAEP data should be relatively immune to concerns about accountability-driven test-score inflation, such as may result from "teaching to the test."

Because our research design depends on measuring achievement trends prior to NCLB, we limit our sample to states that administered the state NAEP at least twice prior to the implementation of NCLB. We include 2002 as a pre-NCLB data point in our analysis because, given the timing of the passage and implementation of the law,

Results

We find that the accountability provisions of NCLB increased 4th-grade math achievement by roughly 7.2 scale points (0.23 standard deviations) by 2007 in states with no prior accountability policies relative to states that adopted accountability systems in 1997. How large is this effect? As one point of reference, consider that the difference between the average scores of 4th and 8th graders in our sample suggests that students gain roughly 12 scale points per year. By this measure, the NCLB impact is equivalent to roughly two-thirds of the average annual gain in scale points. Consider also that the achievement gap between black and white 4th graders on the NAEP math exam is roughly 30 scale points (1 standard deviation), which means that the impact of NCLB is equivalent to about one-quarter of this difference. The effect for 8th-grade math is smaller (0.10 standard deviations) and falls just shy of achieving conventional levels of statistical significance. We find no effects for 4th- and 8th-grade reading.

The design of NCLB necessarily focused the attention of schools on helping students attain proficiency. Figure 2 presents our estimates of the effects of NCLB accountability on the percentage of students achieving at or above the basic and proficient performance levels on NAEP. Although states' definitions of proficient vary widely, very few set the proficiency bar as high as NAEP and most correspond more closely to NAEP's basic performance level. We find that NCLB accountability increased the share of students performing at or above basic in math by 10 percentage points among 4th graders and 6 percentage points among 8th graders. Math proficiency

We do not find that the introduction of NCLB harmed students at higher points on the achievement distribution. Indeed, NCLB accountability seemed to increase achievement among higher-achieving students, if by a smaller amount than it did among their low-achieving peers.

it seems unlikely that spring 2002 scores could have been substantially influenced by NCLB (see sidebar, page 56). All states administered NAEP in 2003, 2005, and 2007.

Our sample includes 39 states for 4th-grade math, 38 states for 8th-grade math, 37 states for 4th-grade reading, and 34 states for 8th-grade reading (see Figure 1). With a few exceptions, our analysis sample closely resembles the nation in terms of student demographics (e.g., percentage African American and percentage Hispanic), observed socio-economic traits (e.g., the poverty rate), and measures of the levels and pre-NCLB trends in NAEP test scores.

rates among 4th graders also increased by 6 percentage points. Again, however, we do not find consistent evidence that NCLB increased reading performance at either grade level.

Given NCLB's focus on proficiency, one would expect the law to disproportionately influence achievement among previously low-achieving students. Our results showing larger increases in the percentage of students reaching the performance level of basic on the NAEP are broadly consistent with this theory. However, in contrast with some previous research and commonly voiced concerns, we do not find that the introduction of NCLB harmed students at higher points

on the achievement distribution. Indeed, NCLB accountability seemed to increase achievement among higher-achieving students, if by a smaller amount than it did among their low-achieving peers. For example, in 4th-grade math, we find that NCLB increased scores at the 10th percentile by roughly 0.29 standard deviations compared with an increase of only 0.17 standard deviations at the 90th percentile (see Figure 3).

One of the primary objectives of NCLB was to reduce inequities in student performance by race and socioeconomic status. Indeed, this concern drove the requirement that, under the statute, accountability ratings be determined by subgroup performance in addition to aggregate school performance. Hence, it is of particular interest to understand the effect of NCLB accountability on specific student subgroups.

In 4th-grade math, these estimated effects are somewhat larger for Hispanic students relative to white students. Similarly, the effects were substantially larger among students who were eligible for subsidized lunch (regardless of race) relative to students who were not eligible. We also found relatively large effects for black students but only when our analysis weighted the state-year NAEP data by the corresponding enrollments of black students. This pattern suggests that NCLB generated more meaningful improvements in the achievement of black students in states where public schools served larger numbers of black students. The effects were roughly comparable for boys and girls.

In 8th-grade math, we find extremely large positive effects for Hispanic students and small, only marginally significant effects for white students. Unfortunately, the results for black students are too imprecisely estimated to warrant interpretation. The effects for students eligible for subsidized lunch are large and statistically significant. Interestingly, for 8th graders the effects are substantially larger for girls, with boys experiencing little if any benefit of accountability.

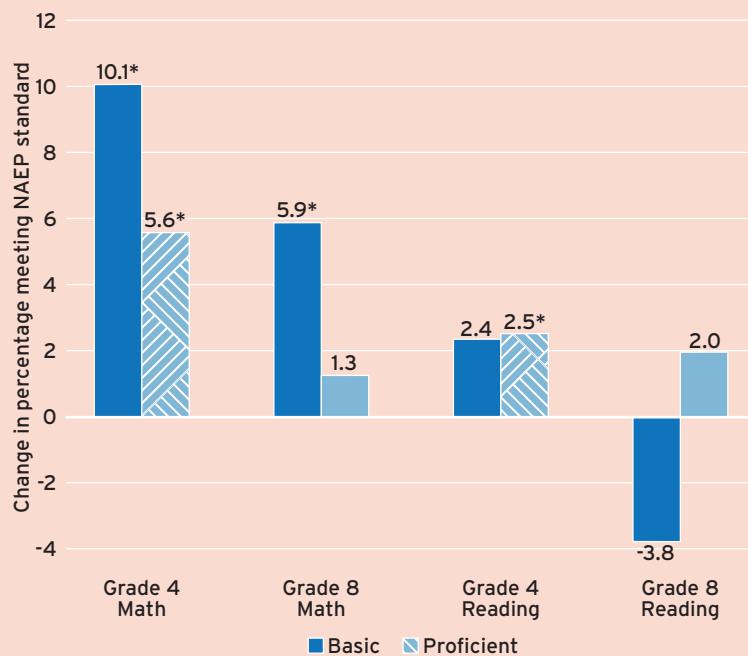
Unintended Consequences?

One concern about NCLB and most other test-based school accountability policies is that they may cause schools to neglect subjects other than math and reading. NAEP data offer some opportunity to test this hypothesis in the context of NCLB. A sizable number of states administered state-representative NAEP tests in science. Unfortunately, during

our analysis period, the 4th-grade science exam was only administered in 2000 and 2005 and the 8th-grade science exam was administered in 1996, 2000 and 2005. The lack of multiple pre- and post-NCLB measures of student achievement limit the power of our research design. Nonetheless, when we apply our research design to these data, we find no statistically significant effects at either grade level at any point on the achievement distribution. Moreover, we are able to

Clear Boost (Figure 2)

By 2007, NCLB's accountability provisions showed a positive impact on NAEP math, but not reading, scores for both 4th and 8th graders.



* Statistically significant at the 95 percent confidence level

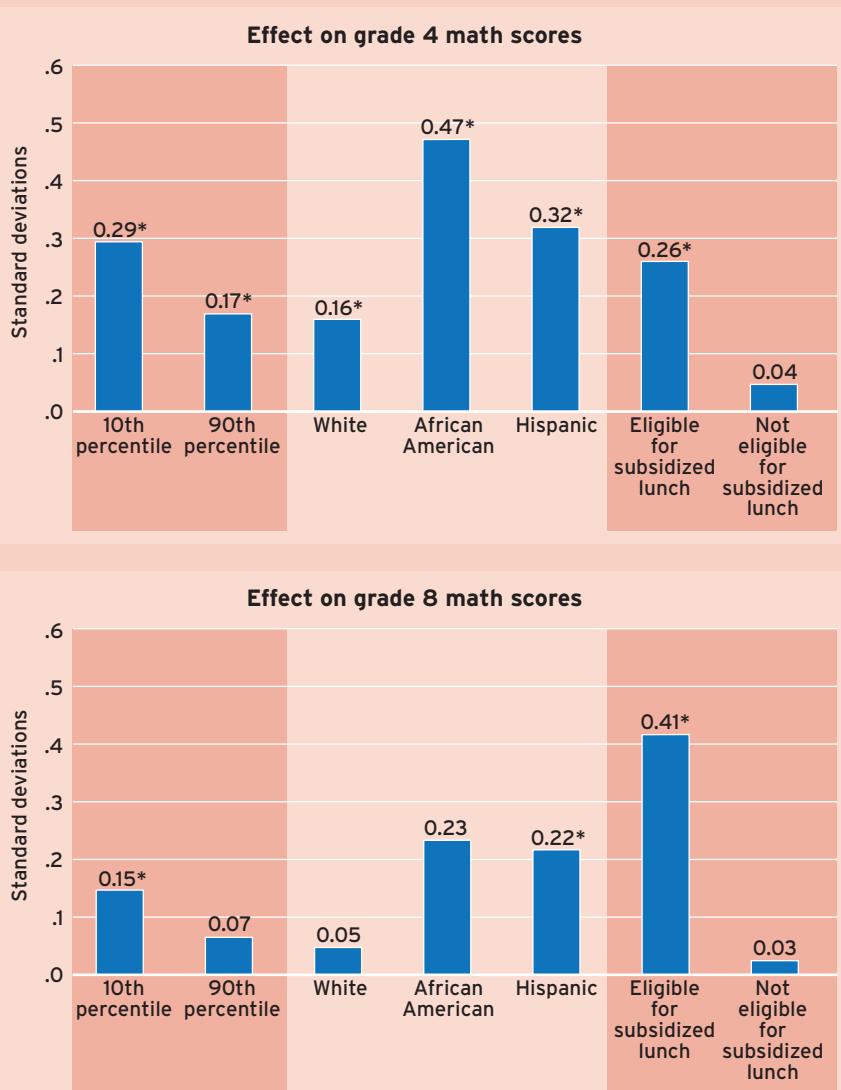
SOURCE: Author calculations

rule out effects larger than roughly 0.10 standard deviations. While these results should be taken with a grain of salt, they cast doubt on some claims that NCLB accountability has had an adverse impact on student performance in science.

Another major concern with test-based accountability, including NCLB, is that it provides teachers an incentive to direct energy toward the types of questions that appear most commonly on the high-stakes test and away from other topics within the tested domain. As noted above, one of the benefits of the analysis presented here is that it relies on

Reaching the Target (Figure 3)

NCLB accountability had the greatest impact on the math scores of African American and Hispanic students and those eligible for subsidized lunch.



* Statistically significant at the 95 percent confidence level

Note: The first two bars indicate effects on students scoring at the 10th and the 90th percentiles.

SOURCE: Author calculations

impact in data analysis and marginally significant effects in number properties and geometry.

The NAEP reading exam measures student competency in several skills related to comprehension: reading for information (i.e., primarily nonfiction reading), reading for literary experience (i.e., primarily fiction reading), and (for 8th grade only) the ability to perform a task (e.g., students apply knowledge from reading bus schedules or directions for repairing something). We find no significant differences in student achievement effects by topic area in reading; that is, NCLB accountability did not appear to have significant effects on student achievement in any of the three reading competencies. Keep in mind, however, that our research design does not allow us to comment on the effects of other aspects of the law, such as the Reading First program, that were explicitly designed to boost reading performance.

Summing Up

So how has NCLB accountability affected student achievement? Our results suggest that its consequences have been mixed. Specifically, we find that the accountability provisions of NCLB generated large and broad gains in the math achievement of 4th graders and somewhat smaller gains for 8th graders. Our results suggest that NCLB accountability had no impact on reading achievement for either group.

The mixed results presented here pose difficult but important questions for policymakers considering whether to “end” or “mend” NCLB. The evidence of substantial and almost universal gains in math is undoubtedly good news for advocates of NCLB. But the lack of any effect in reading, and the fact that the policy appears to have generated only modestly larger impacts among disadvantaged subgroups in math (and thus made only minimal headway in closing achievement gaps), suggests that the impact of NCLB has fallen short of its extraordinarily ambitious goals. Some commentators have argued that the failure of NCLB and earlier accountability reforms to close achievement gaps reflects a flawed, implicit assumption that

student performance on NAEP, which should be relatively immune from such test-score “inflation” since it is not used as a high-stakes test under NCLB or any other accountability system. It is nonetheless interesting to examine whether NCLB accountability has improved student achievement in any particular topic within math or reading. The NAEP math exam measures student performance in five specific topic areas: algebra, geometry, measurement, number properties and operations, and data analysis, statistics, and probability. Our results suggest that NCLB had a positive impact in all math topic areas for the 4th-grade sample. Among 8th graders, NCLB had a moderately large and statistically significant

These results cast doubt on some claims that NCLB accountability has had an adverse impact on student performance in science.

schools alone can overcome the achievement consequences of dramatic socioeconomic disparities.

An effective redesign of accountability policies like NCLB may need to pay more specific attention to the processes and practices operating within schools. Along those lines, it is interesting to note that our evidence of differential effects by grade and subject is broadly similar to the results from evaluations of earlier state-level school-accountability policies. Understanding the sources of these differences is likely to be particularly useful as policymakers discuss the future design and implementation of school-accountability systems. For example, the unique effectiveness of NCLB in improving the math skills of younger students could be related to the biological evidence that cognitive skills are more malleable at early ages. These outcomes may also result

from the specific ways in which schools and teachers have adjusted their instructional practices, perhaps differently for mathematics and reading. Much evidence suggests that school decisions about curricula (e.g., textbooks, instructional software, and the corresponding pedagogy) can have comparatively large effects on student achievement. Further research that can credibly and specifically examine how school and teacher responses have contributed to the achievement effects documented here would be a useful next step in identifying effective policies and practices that can reliably improve student outcomes.

Thomas Dee is associate professor of economics at Swarthmore College. Brian Jacob is professor of education policy and economics at the University of Michigan.

A Conference Organized By: The Harvard Program on Education Policy and Governance

For more information please visit: www.hks.harvard.edu/pepg/events.htm

*Hosted at the Harvard Kennedy School,
Cambridge, MA*



Paying Teachers for Performance June 2-4, 2010

Conference Sponsors: The Taubman Center for State and Local Government
Rappaport Institute for Greater Boston
Center for American Political Studies

The Program on Education Policy and Governance
John F. Kennedy School of Government
Harvard University
79 J.F. Kennedy Street, Cambridge, MA 02138
Email: PEPG_Administrator@hks.harvard.edu

