



Language Magazine on the latest study on the efficacy of No Child Left Behind

Policy Analysis for California Education (PACE), an independent research center based at UC Berkeley, has just released its latest study, “Is the No Child Left Behind Act Working? — The Reliability of How States Track Achievement.” Here is a brief summary of their findings:

Do state testing systems provide an accurate and consistent indication of the share of fourth-grade students who are deemed “proficient” in reading and mathematics?

Looking across 12 diverse states, there is a small improvement in the percentage of children achieving proficiency in reading, based on NAEP results between 1992 and 2005. But states estimated much higher shares of students reaching proficiency, compared with the NAEP results. Children made greater progress in math proficiency over this 13-year period, but, once again, state test results exaggerate the annual rate of improvement, compared with the federal NAEP results.

This historical disparity between state and federal estimates of proficiency levels has not previously been illuminated over this range of states. But the phenomenon is not new. The gap does not stem simply from NCLB’s unintended incentive for states to set low cut-offs for defining which students are deemed proficient. States have long claimed that a much higher share of students are proficient relative to NAEP results, even before NCLB created the incentive for states to set a low bar.

What was the achievement change during the three school years following enactment of NCLB (in January 2002)?

Some states have maintained their apparent momentum in raising the percentage of fourth-graders proficient in math, while reading performance leveled-off or slipped in several states, as gauged by state and NAEP exams. This analysis places earlier reviews of state test scores, post-NCLB, in the context of leveling NAEP scores after 2002.

Two states with weak accountability systems prior to NCLB (Arkansas and Nebraska) did experience gains in math proficiency after enactment of NCLB but not in reading.

We discuss how adjustments to federal reforms could help all states devise student assessment systems that yield more consistent benchmarks of children’s achievement over time.

Conclusions and Policy Options

Taken together, these findings illuminate the difficulty in answering the bottom-line question: Is NCLB working?

Recent claims based solely on state test results — either pre or post-NCLB — assert gains in some states and other states where fourth-graders have reached a plateau in reading or math performance (Education Trust, 2004; PACE, 2004; Paige, 2004).

Yet state results consistently exaggerate the percentage of fourth-graders deemed proficient or above in reading and math — for any

given year and for reported rates of annual progress, compared with NAEP results. For reading, this gulf between the two testing systems has actually grown wider over time. Any analysis conducted over the 1992-2005 period based solely on state results will exaggerate the true amount of progress made by fourth-graders.

State policy makers will likely stand by their testing regimes, given that some exam systems are closely aligned with curricular standards. But it remains unclear whether it's this tighter alignment that is driving higher estimates of the share of fourth-graders who are proficient, or whether the bar set for state tests is simply being set too low, relative to national standards.

Then, teachers and students adapt to state tests in ways that inflate actual levels of substantive learning. Both factors - low cut-points and inflated scores over time - are likely at work in many states.

State education officials at times create tests that are differentially sensitive to improvements at the low level, and states periodically change their testing programs, leading to jagged trend lines - where the share of students defined as proficient shoots upward or falls precipitously, compared with the steady trend lines that characterize NAEP results.

States should not be discouraged from carefully gauging progress at the low end. State and NAEP testing officials could do more to inform the public on how student demographics are changing and implications for interpreting achievement trends.

The rising proportion of English learners or minorities should not be used as an excuse for insufficient progress. But even the interpretation of NAEP trends is constrained by our inability to understand how achievement is moving, net the prior effects of student and family characteristics. In California, for instance, gains in fourth-grade NAEP scale scores for every subgroup of color were strong, even though state averages remained flat, 1998-2005, given steady change in student composition (Sigman & Zilbert, 2006).

The 12 focal states did show some gains in reading from 1992-2001, the decade prior to NCLB. Mean annual changes in NAEP scores ranged between -0.3 percentage points in Oklahoma (in the share of fourth-graders proficient or higher), to 1.1 points in Massachusetts.

These modest inclines were far smaller than average gains reported by state testing programs.

The inflated character of state test scores, following the establishment of low cut-points relative to the NAEP, also was apparent for math results.

Is NCLB working?

Many states continued to show progress after 2001, although the rate of growth fell, compared with the pre-NCLB period, even when gauged by their own testing results.

NAEP reading scores, in general, hit a plateau or declined over the three school years following enactment of NCLB. Gains in fourth-grade math were still apparent after 2001, but here too the gains reported by states were significantly higher than the pace of progress indicated by NAEP scores. Slowing rates of achievement growth, post-NCLB, also have been found in at least one tracking study of large numbers of students distributed across several states (Cronin,

Kingsbury, McCall, & Bowe, 2005).

Could improvements in NCLB rules and resources encourage states to develop more consistent student assessments, yielding truly comparable achievement levels over time?

Yes, on a technical level. Washington could first help raise confidence in state testing programs - and wider acceptance of NAEP results - by advancing a consensus as to where the proficiency bar should be set. State tests might be formally benchmarked to the NAEP, along with more transparent reports on the comparative rigor of state tests.

To combat the dumbing-down of cut-points that define proficiency, policy makers might return to the original vision of the challenging standards envisaged by the early architects of systemic reform: designing assessments that encourage analytic and writing skills and higher order thinking. Adding rigor and more complex skills to state tests could be motivating for teachers and students alike, and it would bring them into closer parity with proficiency levels yielded by NAEP assessments.

The federal government might provide resources to state education departments to conduct stronger equating exercises to link old and new tests. It's understandable that states may periodically want to alter who designs and runs their testing programs. But the inability of states to track achievement over time invites federal intervention and heavier reliance on the NAEP.

It's difficult to see how fourth-graders could be making real advances in reading and component skills without the NAEP detecting such learning.

If state education officials believe their tests are more closely aligned with curricular standards, or somehow more informative than the NAEP, they should address the technical problems that weaken the credibility of their current systems. Otherwise, the relevance of state testing programs may diminish as parents, voters, and policy makers struggle to discern whether school reforms are, in reality, raising student achievement.

The U.S. Department of Education took a notable step in 2005 — inviting states to propose a “growth model” under which schools would be recognized for raising achievement levels no matter where they started on the staircase toward universal proficiency (Anderson, 2005).

This intriguing adjustment raises a different set of issues, especially whether states should be awarded greater flexibility when it comes to gauging student progress.

On the other hand, the new policy discussion around rewards for growth opens the door to bring state and NAEP proficiency standards into closer alignment. This conversation will likely occur within the broader debate around the proper role of Washington and the states within a federalist system of school governance.

State officials might engage this discussion with serious ideas for how to improve their assessment regimes. Otherwise, the credibility of their testing results — and state leaders' claims of progress — may suffer. And Washington officials may want to closely inspect these achievement trends, put forward by the states, before declaring victory. ■

The full report may be downloaded in pdf format from www.pace.berkeley.edu.