

EDUCATION

Building on No Child Left Behind

Eric A. Hanushek

A dominant strand of U.S. educational policy for the past two decades has been incorporation of information about student achievement into management and regulation of schools. Although this policy idea is often attributed simply to the federal No Child Left Behind Act (NCLB), 44 states already had some form of test-based accountability (TBA) when the NCLB came into existence in 2002. NCLB transformed TBA into a national strategy, placed a clear goal on improvements in student achievement as reflected on certain standardized tests, and established a series of actions and penalties for failure to meet annual improvement goals on those tests (including school closure in the worst cases). More than 70% of the American public favors renewal of federal accountability legislation (1), and performance on similar tests is known to relate to important economic outcomes (2). In 2009, the U.S. Supreme Court focused on the importance of outcome accountability in a major school finance decision (3). Thus, TBA has become a fixture of American education. However, it is also clear that the current version could be improved significantly.

What Has Been Accomplished?

NCLB focuses on having all students proficient in reading, math, and science. All states had to develop learning standards and assessments of student performance. Individual schools are required to be on a path toward universal proficiency by 2014. Three lines of inquiry suggest that existing accountability systems have led to larger gains than would be expected in a world without them.

First, comparisons of math and reading performance across states from the National Assessment of Educational Progress (NAEP)—often called the Nation's Report Card—provide some insights. Other things being equal, states introducing accountability earlier showed larger gains on NAEP during the 1990s (4). Moreover, students in states with stronger accountability performed better (5). Students in Florida schools graded "F" for failing state accountability measures, and thus subject to sanctions, showed positive effects of school accountability when compared with similar schools scoring slightly

better at "D" (6). Results of individual state tests over time show that student achievement gains tend to be larger after the introduction of NCLB than before (7). For example, the success of Florida and Texas in raising the achievement of Hispanics and low-achieving students is attributed in part to their use of accountability systems (8–10). Because it is difficult to distinguish impacts of different simultaneous reforms and to establish causation, some uncertainty remains, but the combined picture shows improved student performance after introduction of TBA.

Second, accountability, particularly after NCLB, focused attention on achievement of disadvantaged populations. Evidence indicates this has changed the dynamic within schools, yielding improvements in previously low performing groups (4, 8, 7). In aggregate terms, for example, the black-white gap in mathematics achievement (measured by NAEP for 9- and 13-year-olds) significantly closed between 1999 and 2007 (11).

Third, U.S. evidence is consistent with a growing body of international evidence pointing to the value of central exit exams and more regular accountability. Particularly where there is more autonomy in local decision-making, schools facing accountability pressures do better on international math and science exams (12–14).

What Needs To Be Done?

Critics of NCLB, however, have noted a series of potential problems, including too much focus on basic versus higher-order skills, wide variation in state standards, narrowing of the curriculum, and other distortions in schools. Although the overall importance of each is difficult to pin down with available data and analyses, the underlying ideas suggest useful modifications in the reauthorization of NCLB.

NCLB has each state set learning standards, assessments, and proficiency levels independently, with the federal government determining what actions should be taken when schools fail to make sufficient progress. This division appears to be backward. Under NCLB, states have chosen widely dif-

A federally mandated system of test-based accountability for U.S. education can be made even better.



ferent cutoffs for "proficiency" (15), but in the face of national labor markets where someone from Georgia could well end up working in Arizona, these variations make little sense. History suggests stiff opposition to a national curriculum. As recently seen, however, nothing prevents states from voluntarily joining together to develop standards and assessments (16). The federal government could support and encourage this (17).

On the other hand, the diverse circumstances of schools indicate that centrally defined educational processes are unlikely to be effective (18). The federal government is not well equipped to determine precisely how schools do their jobs. Reforming NCLB could require states to develop their own plans for schools that were failing. Indeed, recognizing the heterogeneity of schools, the U.S. Department of Education has already permitted variation in plans ("differentiated accountability") in nine states (19). As noted above, permitting local autonomy with central testing is a successful strategy consistent with international evidence.

NCLB concentrates on the proportion of students below the state-determined proficiency level in each year. However, accountability should optimally be defined in terms of individual student learning growth, across different learning levels rather than just at the proficiency threshold. This implies that schools should be assessed according to their value added to learning, factoring in such environmental influences as family and neighborhood (20). Fifteen states are already authorized to use growth models for their accountability under NCLB (21). Additionally, this would eliminate incentives to ignore students already above proficiency or too far below to reach proficiency soon.

For testing efficiency, current tests are generally designed to measure most precisely a limited range of skills. An attractive alternative is use of adaptive testing, which can improve measurement in the range of higher-order skills. With adaptive testing (which underlies, for example, the Graduate Record Examination required for entrance to many graduate schools), performance on a set of

Hoover Institution, Stanford University, Stanford, CA 94305, USA. E-mail: hanushek@stanford.edu

initial screening questions moves the student to the range of test questions most relevant to their demonstrated performance—something easily done with computerized testing. The broadened range of testing would thus encourage expansion of topics and techniques introduced in the classroom rather than a focus on content of a more narrowly constrained test domain. Computerization has two additional advantages. First, it would provide immediate scoring of tests, getting around current delays in test scoring. (Note, however, that accountability testing will not be a substitute for formative assessments in the classroom that are designed to provide feedback to teachers in designing regular instruction plans). Second, having a large test bank would permit providing each student with a random selection of questions, minimizing any chance of cheating. Indeed with a large test bank covering the range of relevant material, it would even be possible to make questions available beforehand, with the notion that “teaching to the test” could actually be considered productive.

Research has found teacher quality to be the most important element of a good school (22), and this underlies the NCLB requirement that all schools have only “highly qualified teachers.” Unfortunately, there are severe measurement problems that make previous interpretations of this requirement hollow at best and harmful at worst. Teacher quality is not captured by characteristics such as master’s degrees, teaching experience, or even certification—things that states typically monitor. Fortunately, TBA produces student achievement data needed to assess the value added of teachers, a more appropriate focus of policy concerns (22).

Conclusions

TBA is a fixture of American education, but it has also become controversial. Clearly, TBA does not do everything, but it is a central part of almost all serious reform efforts. Thus, improving it rather than eliminating it is the only reasonable course.

References and Notes

1. W. G. Howell, M. R. West, P. E. Peterson, *Education Next* 7, 12 (2007).
2. E. A. Hanushek, L. Woessmann, *J. Econ. Lit.* 46, 607 (2008).
3. U.S. Supreme Court, *Horne v. Flores* (2009).
4. E. A. Hanushek, M. E. Raymond, *J. Policy Anal. Manage.* 24, 297 (2005).
5. M. Carnoy, S. Loeb, *Educ. Eval. Policy Anal.* 24, 305 (2002).
6. D. N. Figlio, C. E. Rouse, *J. Public Econ.* 90, 239 (2006).
7. Center on Education Policy, *Answering the Question That Matters Most: Has Student Achievement Increased Since No Child Left Behind?* (Center on Education Policy, Washington, DC, 2007).

8. M. Ladner, D. Lips, *Education Next* 9, 20 (2009).
9. D. Grissmer, A. Flanagan, *Exploring Rapid Achievement Gains in North Carolina and Texas* (National Education Goals Panel, Washington, DC, 1998).
10. P. E. Peterson, in *Reforming Education in Florida*, P. E. Peterson, Ed. (Hoover Institution Press, Stanford, CA, 2006), pp. 49–66.
11. A. Vanneman, L. Hamilton, J. B. Anderson, T. Rahman, *Achievement Gaps: How Black and White Students in Public Schools Perform in Mathematics and Reading on the National Assessment of Educational Progress* (NCEES 2009-455, U.S. Department of Education, Washington, DC, 2009).
12. L. Woessmann, in *No Child Left Behind? The Politics and Practice of School Accountability*, P. E. Peterson, M. R. West, Eds. (Brookings, Washington, 2003), pp. 292–323.
13. J. H. Bishop, in *Handbook of the Economics of Education*, E. A. Hanushek, F. Welch, Eds. (North-Holland, Amsterdam, 2006), pp. 909–944.
14. L. Woessmann, E. Luedemann, G. Schuetz, M. R. West, *School Accountability, Autonomy, and Choice Around the World* (Edward Elgar, Cheltenham, UK, 2009).
15. National Center for Education Statistics, *Mapping 2005 State Proficiency Standards onto the NAEP Scales* (U.S. Department of Education, Washington, DC, 2007).
16. M. McNeil, *Education Week* 28, 16 (2009).
17. J. E. Chubb, *Learning from No Child Left Behind: How and Why the Nation’s Most Important but Controversial Education Law Should be Renewed* (Education Next Books, Hoover Institution Press, Stanford, CA, 2009).
18. E. A. Hanushek, A. A. Lindseth, *Schoolhouses, Courthouses, and Statehouses: Solving the Funding-Achievement Puzzle in America’s Public Schools* (Princeton Univ. Press, Princeton, NJ, 2009).
19. U.S. Department of Education, Administrators: Differentiated Accountability States (U.S. Department of Education, Washington, DC, 2008); www.ed.gov/admins/lead/account/differentiatedaccountability/dapstates.html.
20. P. E. Peterson, M. R. West, *Education Next* 6, 77 (2006).
21. U.S. Department of Education, Administrators: Lead and Manage My School: Growth Models [elementary and secondary education]; www.ed.gov/admins/lead/account/growthmodel/index.html.
22. E. A. Hanushek, S. G. Rivkin, in *Handbook of the Economics of Education*, E. A. Hanushek, F. Welch, Eds. (North-Holland, Amsterdam, 2006), pp. 1051–1078.

10.1126/science.1177458

EDUCATION

Moving Past No Child Left Behind

Daniel Koretz

No Child Left Behind is a poorly designed program, with serious side effects and little evidence of benefit, in need of fundamental changes.

Weaknesses in the U.S. educational system are clear. U.S. students do not compare well to peers in many other nations in their mastery of mathematics and science (1). Inequities in educational resources and outcomes are glaring. Although policy responses to these problems should include holding educators accountable for student performance, No Child Left Behind (NCLB) is a poorly designed test-based accountability (TBA) system that requires fundamental changes.

Score Inflation

A fundamental problem with TBA systems—one that NCLB fails utterly to address—is score inflation, increases in scores substantially larger than warranted by true gains in students’ learning. Research suggests that inflation, although not ubiquitous, is common and can be very large (2, 3). This phenomenon of corrupted outcome measures is not particular to TBA. It has been seen in many fields where performance-based incentives are imposed (4).

For example, in the early 1990s, Kentucky instituted a TBA system that in many respects

foreshadowed NCLB. During the first 2 years for which teachers were held accountable for scores on the state test, fourth-grade scores on the state’s reading test rose by 0.76 standard deviation (a truly remarkable gain, an order of magnitude larger than historical data would have suggested reasonable). During the same period, the state’s fourth-grade reading scores on a federally managed test [the National Assessment of Educational Progress (NAEP)] did not improve (5), even though the two tests were designed to measure similar material.

Score inflation has several roots, but most important is that achievement tests are necessarily small, somewhat predictable samples of larger domains of achievement. Sample-based testing can work well when educators and students have no strong incentive to focus on the specific sample measured. In TBA systems, however, such incentives are strong (6, 7), and scores sometimes become inflated. Research (8) has indicated that many teachers reallocate instructional time in an effort to focus on tested material (and even on the particular forms in which content appears on the test) at the expense of other content.

Focus on the Proficiency Threshold

These problems must be confronted in the design of any TBA system, but NCLB also

Graduate School of Education, Harvard University, Cambridge, MA 02138 USA. E-mail: daniel_koretz@harvard.edu