

June 7, 2011

Panel Finds Few Learning Gains From Testing Movement

By Sarah D. Sparks

Nearly a decade of America's test-based accountability systems, from "adequate yearly progress" to high school exit exams, has shown little to no positive effect overall on learning and insufficient safeguards against gaming the system, a blue-ribbon committee of the National Academies of Science concludes in a new report.

"Too often it's taken for granted that the test being used for the incentive is itself the marker of progress, and what we're trying to say here is you need an independent assessment of progress," said Michael Hout, the sociology chair at the University of California, Berkeley. He is the chairman of the 17-member committee, a veritable who's who of national experts in education law, economics and social sciences that was launched in 2002 by the National Academies, a private, nonprofit quartet of institutions chartered by Congress to provide science, technology and health-policy advice. During the last 10 years, the committee has been tracking the implementation and effectiveness of 15 test-based incentive programs, including:

- National school improvement programs under the No Child Left Behind Act and prior iterations of the Elementary and Secondary Education Act;
- Test-based teacher incentive-pay systems in Texas, Chicago, Nashville, Tenn., and elsewhere;
- High school exit exams adopted by about half of states;
- Pay-for-scores programs for students in New York City and Coshocton, Ohio and;
- Experiments in teacher incentive-pay in India and student and teacher test incentives in Israel and Kenya.

On the whole, the panel found the accountability programs often used assessments too narrow to accurately measure progress on program goals and used rewards or sanctions not directly tied to the people whose behavior the programs wanted to change. Moreover, the programs often had insufficient safeguards and monitoring to prevent students or staff from simply gaming the system to produce high test scores disconnected from the learning the tests were meant to inspire.

“I think there are some real messages for school districts on accountability systems” in the report, said Kevin Lang, an economics professor at Boston University who, during his time on the committee, also served as a district school board member in Brookline, Mass.

“School boards need to have a means for monitoring the progress of their school systems, and they tend to do it by looking at test scores,” he said. “It’s not that there’s no information in the objective performance measures, but they are imperfect, and including the subjective performance measures is also very important. Incentives can be powerful, but not necessarily in the way you would like them to be powerful.”

Gaming the System

Among the most common problems the report identifies is that most test-based accountability programs use the same test to apply sanctions and rewards as to evaluate objectively whether the system works. As a result, staff and students facing accountability sanctions tend to focus on behavior that improves the test score alone, such as teaching test-taking strategies or drilling students who are closest to meeting the proficiency cut-score, rather than improving the overall learning that the test score is expected to measure. This undercuts the validity of the test itself.

For example, New York’s requirement that all high school seniors pass the Regents exam before graduating high school led to more students passing the Regents tests, but scores on the lower-stakes National Assessment of Educational Progress, which was testing the same subjects, didn’t budge during the same time period, the report found.

“It’s human nature: Give me a number, I’ll hit it,” Mr. Hout said. “Consequently, something that was a really good indicator before there were incentives on it, be it test scores or the stock price, becomes useless because people are messing with it.”

In fact, the report found that, rather than leading to higher academic achievement, high school exit exams so far have decreased high school graduation rates nationwide by an average of about 2 percentage points.

The study found a growing body of evidence of schools and districts tinkering with how and when students took the test to boost scores on paper for students who did not know the material—or to prevent those students from taking the test at all.

Recent changes to federal requirements for reporting graduation rates, which require that schools count as dropouts students who “transfer” to a school that does not award diplomas, may help safeguard against schools pushing out students to improve test scores or graduation rates. Still, the National Academies researchers warned that state and federal officials do not provide enough outside monitoring and evaluations to ensure the programs work as intended.

AYP and Academics

For similar reasons, school-based accountability mechanisms under NCLB have generated minimal improvement in academic learning, the study found. When the systems are evaluated—not using the high-stakes tests subject to inflation, but using instead outside comparison tests, such as the NAEP—student achievement gains dwindle to about .08 of a standard deviation on average, mostly clustered in elementary-grade mathematics.

To give some perspective, an intervention considered to have a small effect size is usually about .1 standard deviations; a 2010 federal study of reading-comprehension programs found a moderately successful program had an effect size of .22 standard deviations.

Moreover, “as disappointing as a .08 standard deviation might be, that’s bigger than any effect we saw for incentives on individual students,” Mr. Hout said, noting that NCLB accountability measures school performance, not that of individual students

Committee members see some hopeful signs in the 2008 federal requirement that NAEP scores be used as an outside check on achievement results reported by districts and states, as well as the broader political push to incorporate more diverse measures of student achievement in the next iteration of ESEA.

“We need to look seriously at the costs and benefits of these programs,” said Daniel M. Koretz, a committee member and an education professor at Harvard University Graduate School of Education in Cambridge, Mass. “We have put a lot into these programs over a period of many years, and the positive effects when we can find them have been pretty disappointing.”

Jon Baron, the president of the Washington-based Coalition for Evidence-Based Policy and the chairman of the National Board for Education Sciences, which advises the Education Department’s research arm, said he was impressed by the quality of the committee’s research review but unsurprised at minimal results for the various incentive programs.

Incorporating diverse types of studies typically reduces the overall effects found for them, he noted, adding that the study also addresses a broader issue. “One of the contributions that this makes is that it shows that looking across all these different studies with different methodologies and populations, some in different countries, there are very minimal effects in many cases and in a few cases larger effects. It makes the argument that details matter,” Mr. Baron said.

“It’s an antidote to what has been the accepted wisdom in this country, the belief that performance-based accountability and incentive systems are the answer to improving education,” Mr. Baron said. “That was basically accepted without evidence or support in NCLB and other government and private sector efforts to increase performance.”