



Week of May 31-June 3, 2011

National Research Council Casts Doubt on High-Stakes Testing

By UCLA IDEA

In an era when education policy and practice are intertwined with standardized testing, a new report makes official what has been known for years—that high-stakes testing is not an effective “lever” for improving teaching and learning.

“Incentives and Test-Based Accountability in Education,” published by the highly-regarded Washington, D.C.-based National Research Council, reports on testing and sanctions in school programs over the last 10 years. Drawing on leading scholars from around the nation, the National Research Council creates independent expert reports that synthesize the consensus understandings of the research community. This report concluded that high-stakes standardized tests have created environments where teachers emphasized test-taking skills and limited instruction to what they thought would appear on the tests. Among the “unintended consequences” was that students could improve scores while actually learning less.

UC Berkeley professor Michael Hout, chairman of the committee that wrote the report, said, “It’s human nature: Give me a number, I’ll hit it. . . 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.”

This week’s *Theme* will post the report’s two powerful conclusions and three recommendations. The full report is a must-read for everyone who wants to know how high-stakes tests influence education opportunity. These data and findings from the National Research Council can be powerful tools for discussion and starting points for consensus in future debates about the role of education testing and sanctions.

Posted unedited from ‘Incentives and Test-Based Accountability in Education’:

Conclusion 1: Test-based incentive programs, as designed and implemented in the programs that have been carefully studied, have not increased student achievement enough to bring the United States close to the levels of the highest achieving countries. When evaluated using relevant low-stakes tests, which are less likely to be inflated by the incentives themselves, the overall effects on achievement tend to be small and are effectively zero for a number of programs. Even when evaluated using the tests attached to the incentives, a number of programs show only small effects. Programs in foreign countries that show larger effects are not clearly applicable in the U.S. context. School level incentives like those of NCLB produce some of the larger estimates of achievement effects, with effect sizes around 0.08 standard deviations, but the measured effects to date tend to be concentrated in elementary grade mathematics and the effects are small compared to the improvements the nation hopes to achieve.

-more-

For links to articles, read our *Themes in the News* online.

www.ucla-idea.org



THEMES in the NEWS

IDEA's weekly commentary on education news

Week of May 31-June 3, 2011

Page 2

Conclusion 2: The evidence we have reviewed suggests that high school exit exam programs, as currently implemented in the United States, decrease the rate of high school graduation without increasing achievement. The best available estimate suggests a decrease of 2 percentage points when averaged over the population. In contrast, several experiments with providing incentives for graduation in the form of rewards, while keeping graduation standards constant, suggest that such incentives might be used to increase high school completion.

Recommendation 1: Despite using them for several decades, policymakers and educators do not yet know how to use test-based incentives to consistently generate positive effects on achievement and to improve education. Policymakers should support the development and evaluation of promising new models that use test-based incentives in more sophisticated ways as one aspect of a richer accountability and improvement process. However, the modest success of incentive programs to date means that all use of test-based incentives should be carefully studied to help determine which forms of incentives are successful in education and which are not. Continued experimentation with test-based incentives should not displace investment in the development of other aspects of the education system that are important complements to the incentives themselves and likely to be necessary for incentives to be effective in improving education.

Recommendation 2: Policymakers and researchers should design and evaluate new test-based incentive programs in ways that provide information about alternative approaches to incentives and accountability. This should include exploration of the effects of key features suggested by basic research, such as who is targeted for incentives; what performance measures are used; what consequences are attached to the performance measures and how frequently they are used; what additional support and options are provided to schools, teachers, and students in their efforts to improve; and how incentives are framed and communicated. Choices among the options for some or all of these features are likely to be critical in determining which—if any—incentive programs are successful.

Recommendation 3: Research about the effects of incentive programs should fully document the structure of each program and should evaluate a broad range of outcomes. To avoid having their results determined by the score inflation that occurs in the high-stakes tests attached to the incentives, researchers should use low-stakes tests that do not mimic the high-stakes tests to evaluate how test-based incentives affect achievement. Other outcomes, such as later performance in education or work and dispositions related to education, are also important to study. To help explain why test based incentives sometimes produce negative effects on achievement, researchers should collect data on changes in educational practice by the people who are affected by the incentives.

####