

# **SAVE OUR SCHOOLS**

## **Education Research Brief**

### **US National Research Council Fails Test-Based Accountability Programs**

A decade or more of test-based accountability programs in the United States has had little to no effect on student achievement according to a report just published by an expert panel of the US National Research Council. The report is a devastating indictment of the focus on testing in US education policy and, by implication, Australian education policy which now closely follows the US model.

The report found that “the overall effects on achievement tend to be small and are effectively zero for a number of programs” [p.4-26]. It says that the impact of test-based accountability programs has been far below what was hoped and far less than what is needed to bring the US up to the results of the top achieving countries.

It cautions about the use of these programs, saying that the “available evidence does not justify a single-minded focus on test-based incentives as a primary tool of education policy” [p.5-2].

The significance of this report cannot be over-estimated. It is one of the most comprehensive reviews of the experience with test-based accountability measures to date and was carried out by a panel of some of the most eminent scholars in the US. Their conclusion is emphatic:

We conclude.... that the available evidence does not give strong support for the use of test-based incentives to improve education and provides only minimal guidance about which incentive designs may be effective. [p.5-1]

The report also describes a range of negative impacts of test-based accountability measures. These include narrowing of the curriculum by focusing on the areas and grades tested at the expense of other areas of the curriculum and grades. Others include gaming of results by test preparation, focusing on students just below performance benchmarks, excluding low performing students and cheating.

Test-based accountability measures dominate the education policy agenda in the US at all levels of government. In recent years there have been increasing efforts by the federal government and the states to devise systems based on test results that make students, teachers, principals, or whole school systems accountable for how much students learn. They include publishing school results, rewards and sanctions for schools, incentive pay schemes for teachers, reward payments for students and high school exit exams.

Versions of many of these schemes have been, or are about to be, adopted in Australia by the Labor Government, so the results of this major study are very relevant.

The expert panel was convened by the US National Research Council, the research arm of the National Academies of Sciences, Engineering and Medicine. The 17-member panel comprised a veritable who’s who of national experts in education law, economics and

social sciences. It has been tracking the implementation and effectiveness of 15 test-based incentive programs for the past 10 years.

The panel reviewed studies in three categories: school-level policies related to the No Child Left Behind Act (NCLB) and its predecessors; experiments with teachers and students that use rewards, such as performance pay; and high school exit exams.

### **NCLB and its predecessors**

A number of US states instituted test-based incentives during the 1990s, with consequences for schools that anticipated those implemented for all states in 2001 under the NCLB. The NCLB defines consequences for schools that do not show adequate yearly progress in improving student achievement. The initial requirements are to file improvement plans, make curriculum changes, and offer their students school choice or tutoring. If there is not adequate progress under these measures, schools are required to restructure in various ways. The consequences are based on state tests in reading and mathematics that use state-defined targets for student proficiency.

The panel found that, at first sight, the evidence from studies of the impact of the NCLB and its predecessors appear to be substantial. However, it found that the statistically significant effects were concentrated in 4th grade mathematics. In contrast, the results for 8th grade math and for reading for both grades were often not statistically significant and sometimes negative. It also noted that the two estimates showing the largest improvements came from analyses that excluded results for 8<sup>th</sup> grade reading, and thus gave an unbalanced overall picture of the effects of the incentives on achievement.

The panel concluded that overall effect on student achievement for both the pre-NCLB and the NCLB periods was very small, amounting to only about three percentile points.

...the evidence related to the effects on achievement of test-based incentives to schools appears to be modest, limited in both size and applicability....This gain is small, both by itself and in comparisons across nations....To achieve an increase of the magnitude needed to match the high performing countries would mean that students currently at the 50th percentile in the United States would have to increase their scores to the current 84th percentile. For underachieving groups, far more improvement would be needed because of the large achievement gaps in the United States [p. 4-24 – 4-25].

The panel also examined the studies of other effects of the NCLB and its predecessors. It found changes in teaching practice which involved an increased focus on tested subjects, on lower performing students at the threshold of attaining proficiency, and on material that closely mimics the tests.

In addition to changes in instruction in the subject area, there is also evidence of attempts to increase scores in ways that are completely unrelated to improving learning. The attempts included teaching test-taking skills, excluding low-performing students from tests, feeding students high-calorie meals on testing days, providing help to students during a test, and even changing student answers on tests after they were finished.

### **Reward incentives**

The panel examined a number of US and overseas studies on the impact of various forms of incentive payments to teachers and students based on test results.

The US studies analysed schemes operating in Chicago, Nashville, New York City, Ohio, and Texas. The overall effects on achievement were found to be very small. Many effects were positive, but only some were statistically significant. There were also a number of negative effects, that is, the incentive scheme lead to a decline in student achievement.

The panel also analysed studies of incentive payment programs in India, Israel, and Kenya. The effects were generally positive, ranging from very small to modestly large, and were statistically significant. However, the review noted that the programs in India and Israel measured achievement using high-stakes tests attached to the incentives and this is likely to upwardly bias the observed effects. It said that there was evidence that the incentive programs focused teaching on test preparation. It also noted that the education context in these countries is quite different from that of the US and questioned what can be learned from these studies that would be applicable to the use of incentives in the United States.

### **High school exit exams**

Use of exit exams in the US has been growing over the past three decades and now includes 25 states and two-thirds of public high school students. Exit exams typically involve tests in multiple subjects, all of which must be passed before a student can graduate, though many states provide alternate paths that can be substituted for a failure on one or more subject tests.

Only one causal study has investigated the effect of the introduction of these exams on student achievement. It found that they had no overall effect on student achievement for either lower or higher performing students. The effects on different combinations of subject and grade were evenly divided between small positive and negative effects, and none was statistically significant.

Two studies have examined their impact on student graduation rates. Both found that states adopting more difficult exit exams showed a statistically significant decrease in graduation rates.

### **Test-based incentives narrow education**

The report shows that standardised tests do not adequately measure education performance and quality, and can actually harm education beyond the effect of test scores on the tested subjects.

The panel said that the tests typically used in the US to measure performance in education fall short of providing a complete measure of desired educational outcomes in many ways. First, the tests obviously only measure performance in selected subjects and grades. Second, they omit entirely advanced levels of performance in the two tested subjects of reading and mathematics as well as areas of performance in those subjects that are hard to assess with standardized tests. Third, they do not measure the learning of such important characteristics as creativity, curiosity, persistence, values, collaboration, and socialization.

The report says these omissions are significant because the use of incentives for performance on tests is likely to reduce emphasis on the outcomes that are not measured by the test. Even if test scores on the tested portion of the curriculum increase, understanding of the untested portion may decrease.

In addition, the report notes that the use of tests for high-stakes accountability has meant that inappropriate forms of test preparation are becoming more widespread and problematic. The

consequence is that test results may become increasingly misleading as measures of achievement in a domain when instruction is focused too narrowly on the specific knowledge, skills, and test question formats that are likely to appear on the test. Overly narrow instruction might include such practices as drilling students on practice questions that were released from prior years' tests, focusing on the limited subset of skills, knowledge and question formats that are most likely to be tested, and teaching test-taking tricks.

The report warns that inappropriate test preparation can harm education:

Bad or inappropriate test preparation is instruction that leads to test score gains without increasing students' mastery of the broader, intended domain, which can result from engaging in the types of inappropriate strategies discussed above. These practices are technically permissible and can even be appropriate to a limited degree, but they will not necessarily help students understand the material in a way that generalizes beyond the particular problems they have practiced. Mastering content taught in test-like formats has been shown not to generalize to mastery of the same content taught or tested in even slightly different ways. In this kind of situation, test scores are likely to give an inflated picture of students' understanding of the broader domain. [p.3-3]

### **Future policy and research**

The report finds that 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. It says alternative approaches are needed. Use of incentives should be rigorously evaluated to determine what works and what does not.

It recommends that further use of incentive system should be based on multiple performance measures. Given that tests are necessarily incomplete measures of valued educational goals, designers of incentive systems should recognize the potential problems inherent in having strong consequences based on test scores alone and should experiment with the use of systems of multiple measures that reflect desired outcomes.

The report also emphasises that there are other aspects of improving education that would benefit from development. In particular, investments to improve standards, curriculum, instructional methods, and educator capacity are all necessary for improving educational outcomes.

Trevor Cobbold

Michael Hout & Stuart W. Elliott (eds), [Incentives and Test-Based Accountability in Public Education](#), National Research Council, National Academies Press, Washington DC, 2011.

**SOS - Fighting for Equity in Education**

<http://www.saveourschools.com.au>