

# Standardized tests with high stakes are bad for learning, studies show

By Carolyn J. Heinrich

SPECIAL TO THE AMERICAN-STATESMAN

Published: 6:43 p.m. Saturday, March 10, 2012 | Updated: 6:47 p.m. Saturday, March 10, 2012

Standardized achievement tests have long been a routine part of our efforts to measure the educational progress of students. In the distant past, testing days came and went with little notice or fanfare for students, parents and teachers alike. And in those days and times, the tests probably provided fairly accurate assessments of students' progress in learning from one year to the next.

But those days of relatively relaxed test-taking for students and limited stakes for school districts and teachers are long gone. Test-based accountability systems that attach weighty consequences to student test results for school district staff, teachers, students and public officials are becoming increasingly institutionalized in the education system. There are probably few other places where the stakes attached to these tests are as high as they are in Texas.

There is a clear rationale for tying incentives for educational improvement to student achievement tests. We know from a variety of economic, psychological and management studies that people are highly responsive to incentives, even those that do not necessarily have individual rewards or sanctions linked to them or that may merely accord some form of public recognition (or shame) based on the results. Unfortunately, what the research has also definitively shown is that people will respond to these incentives in both intended and unintended ways, and the less control they feel they have over the measured outcomes and the more stringent the targets or performance tests, the more likely they are to respond perversely.

We have observed these patterns of unconstructive responses to performance incentives across a number of domains — health, workforce training, public assistance programs and more — but the evidence of serious problems has piled up faster in public education than in any other policy arena.

I was part of a National Academies of Science committee that was asked to carefully review the nature and implications of America's test-based accountability systems, including school improvement programs under the No Child Left Behind Act, high school exit exams, test-based teacher incentive-pay systems, pay-for-scores initiatives and other uses of test scores to evaluate student and school performance and determine policy based on them. We spent nearly a decade reviewing the evidence as it accumulated, focusing on the most rigorous and credible studies of incentives in educational testing and sifting through the results to uncover the key lessons for education policymakers and the public.

Our conclusion in our report to Congress and the public was sobering: There are little to no positive effects of these systems overall on student learning and educational progress, and there is widespread teaching to the test and gaming of the systems that reflects a wasteful use of resources and leads to inaccurate or inflated measures of performance.

Before high stakes are attached to a particular performance measure, such as math scores, it may very well correlate well with positive student outcomes that we are trying to encourage and build up. Indeed, the National Assessment of Educational Progress test is often used as a gauge for actual student performance specifically because it is a low-stakes test. But once a measure of performance is activated in a system that attaches significant consequences to its attainment, individuals are motivated to pursue all possible ways to raise measured performance, including those that do not contribute to the genuine goals of the system — goals such as increasing student knowledge and learning capabilities.

Studies published in the best economics and education journals have shown unequivocal evidence of excessive teaching to the test and drilling that produces inflated measures of students' growth in learning; cheating on tests that includes erasing incorrect answers or filling in missing responses; shifting of students out of classrooms or other efforts to exclude anticipated poor performers from testing, or alternatively, concentrating classroom teaching efforts on those students most likely to increase their test scores above a particular target, and other even more subtle strategies for increasing testing averages.

This type of behavior, which narrows the focus of classroom education and frequently diverts time and resources from more innovative and interactive approaches to teaching, has been characterized in academic literature and policy circles alike as "hitting the target and missing the point."

What we have come to understand to date about test-based accountability does not bode well for the new policy introduced by the Texas Legislature that will make the new STAAR student achievement tests count toward 15 percent of a student's course grade. After a wave of objections, state Education Commissioner Robert Scott announced last month that school districts may wait until a year from now to begin applying the 15 percent rule. If and when it kicks in, these are very high stakes to attach to a test, and this will undoubtedly have implications for how teachers and students spend their time in the classroom.

I am a parent of a freshman in an Austin public high school who is already fretting about what she understands to be very serious and formidable consequences of this new policy for her future. She found the questions on a practice exam to be largely unrelated to what she was learning in the classroom, and this was reflected in the scores she attained. How can she continue to spend five to six hours a night working on her regular schoolwork and preparing for exams created by her teachers and also find time to prepare for taking a separate set of tests that will count toward 15 percent of her course grade?

The empirical research again speaks to the unintended effects this policy is likely to generate: fear, reduced student motivation, increased withdrawals and lower graduation rates are examples of well-documented negative effects that this type of high-stakes testing induces.

In an assembly of parents, teachers, school staff and district officials, a presentation was made showing how the district will be working to improve the students' test-taking skills. The example suggested that if students do not know the meaning of a particular word in a test item, they would be taught to replace it with an "X" and focus instead on grasping the logic of the question phrasing that will give them a better chance of selecting the correct answer. In other words, if you do not understand the content, you can still improve your "guessing" skills through these efforts to help you become a better test-taker. Is this the way we want our children to spend their time in the classroom?

Maybe we should shift to the model widely used in some Asian countries where, in addition to classroom time spent on test-taking, the students spend 10-hour days on the weekends and their holiday breaks in test preparation classes that drill them in precisely this way. As a university professor, I have seen the results of this extreme focus on test-taking: These students score at the highest levels on tests that are reported in their admissions applications, but they score considerably lower on writing assessments, and most importantly, their performance in the classroom does not measure up to the test scores.

Incentivizing this type of intensive focus on improving test-taking capabilities is not going to help produce the better educated, more highly skilled and innovative workforce that business leaders and other employers assert is essential to competing effectively in the global economy.

Texas should be applauded for its tireless efforts to develop new policies that will increase educational effectiveness and to set high achievement goals for its students. But the latest rendition to incentivize better student performance in the form of a policy that ties 15 percent of a student's course grade to these tests is a step backward, not forward. It ignores a now broad base of evidence that these policies produce minimal or no positive effects on student learning and are likely to induce costly, negative responses in and beyond the classroom. I hope that the deferral of its requirement will give the state the time needed to revisit and retract this step in the wrong direction.

<http://www.statesman.com/news/news/opinion/standardized-tests-with-high-stakes-are-bad-for--1/nRk77/>

---

Carolyn J. Heinrich is Sid Richardson Professor of Public Affairs and affiliated professor of economics at the University of Texas LBJ School Of Public Affairs. She is also president of the Public Management Research Association and an elected member of the National Academy of Public Administration. Her research focuses on education and social welfare policy, public management, performance management, and econometric methods of social program evaluation.