

A Research Report

The Effects of High-Stakes Testing on Student Motivation and Learning

Do high-stakes testing policies lead to increased student motivation to learn? And do these policies lead to increased student learning? No, according to four independent achievement measures.

Audrey L. Amrein and David C. Berliner

The current generation of policy-makers did not invent high-stakes testing. Tests of various sorts have determined which immigrants could enter the United States at the turn of the 20th century, who could serve in the armed forces, who was gifted, who needed special education, and who received scholarships to college. But the No Child Left Behind Act of 2001 aims to make high-stakes testing more pervasive than ever before, mandating annual testing of students in grades 3-8 in reading and math.

The federal legislators who overwhelmingly passed this act into law apparently assumed that high-stakes tests would improve student motivation and raise student achievement. Because testing programs similar to those required by No Child Left Behind already exist in many states, we can put that assumption to the test.

Eighteen states currently use exams

to grant or withhold diplomas: Alabama, Florida, Georgia, Indiana, Louisiana, Maryland, Minnesota, Mississippi, Nevada, New Jersey, New Mexico, New York, North Carolina, Ohio, South Carolina, Tennessee, Texas, and Virginia.¹ As Figure 1 shows, most of these states also attach to their state assessments a broad range of other consequences for students, teachers, and schools. The experiences of these states can help us predict how the new nationwide program of high-stakes testing will affect student achievement.

Unfortunately, the evidence shows that such tests actually decrease student motivation and increase the proportion of students who leave school early. Further, student achievement in the 18 high-stakes testing states has not improved on a range of measures, such as the National Assessment of Educational Progress, despite higher scores on the states' own assessments.

Effects on Motivation to Learn

High-stakes testing assumes that rewards and consequences attached to rigorous tests will "motivate the unmotivated" to learn (Orfield & Kornhaber, 2001). The "unmotivated" are usually identified as low socioeconomic students in urban schools, often African Americans and Latinos.

Yet researchers have found that when rewards and sanctions are attached to performance on tests, students become *less* intrinsically motivated to learn and less likely to engage in critical thinking. In addition, they have found that high-stakes tests cause teachers to take greater control of the learning experiences of their students, denying their students opportunities to direct their own learning. When the stakes get high, teachers no longer encourage students to explore the concepts and subjects that interest them. Attaching stakes to tests appar-



© David Paul Productions/Getty Images

ently obstructs students' path to becoming lifelong, self-directed learners and alienates students from their own learning experiences in school (Sheldon & Biddle, 1998).

Wheelock, Bebell, and Haney (2000) investigated the degree to which external tests motivated students to learn by examining the self-portraits of students in testing situations. Students depicted themselves as anxious, angry, bored, pessimistic, and withdrawn from high-stakes tests. Older students were more disillusioned and hostile toward tests than were younger students.

As Sacks writes,

Test-driven classrooms exacerbate boredom, fear, and lethargy, promoting all manner of mechanical behaviors on the part of teachers, students, and schools, and bleed schoolchildren of their natural love of learning. (1999, pp. 256-257)

In sum, the assumption that high-stakes tests motivate students appears to be seriously flawed. In fact, such tests often decrease student motivation and

Attaching stakes to tests alienates students from their own learning experiences in school.

lead to higher student retention and dropout rates.

High School Dropouts

Dropout rates are climbing throughout the United States, and many researchers hold high-stakes tests at least partly to blame (Rothstein, 2002). Some researchers found that dropout rates were 4 to 6 percent higher in schools with high school graduation exams. Another study reported that students in the bottom quintile in states with high-stakes tests were 25 percent more likely to drop out of high school than were their peers in states without high-stakes tests (Jacob, 2001). Researchers in yet

another study found that failing these tests also significantly increased the likelihood that even the students with better academic records would drop out (FairTest & Massachusetts CARE, 2000).

We calculated that 88 percent of the states with high school graduation tests have higher dropout rates than do states without graduation tests. In 62 percent of these states, dropout rates increased in comparison with the rest of the nation after the state implemented high-stakes graduation exams. In addition, the top 10 states with the weakest grade 9-12 continuation ratios all administered high-stakes tests over the years for which data were available (Amrein & Berliner, 2002a).

Students Earning Alternative Degrees

More and more teenagers are exiting formal schooling early to earn a General Educational Development (GED) credential (Murnane, Willett, & Tyler, 2000). Although young people who have earned such alternative degrees do not technically count in dropout statistics, many of them undoubtedly left school because of their concerns about passing rigorous graduation tests.

Sixty-three percent of the states with high school graduation tests posted decreases in the average age of students who took the GED exam after the high-stakes tests were implemented, according to our analysis of KidsCount data (Amrein & Berliner, 2002a). Other studies confirm our finding (Haney, 2001; Murnane, Willett, & Tyler, 2000). In North Carolina, for example, the proportion of students under the age of 20 taking the GED increased by 73 percent between 1986 and 1999, 43 percent more than the proportion for the nation during the same time. And in Georgia, almost twice as many teenagers earned GEDs in 2000 than in 1990 (American Council on Education, 2001).

Student Retention

Students who repeat a grade are significantly more likely to drop out of school (Goldschmidt & Wang, 1999). In states where promotion to the next grade hinges on passing the state exams, high-stakes testing policies also contribute to

A Research Report

higher dropout rates in the long run.

Nonpromotion policies have been implemented in Chicago and in Louisiana. The experiences of these localities foreshadow what will happen as policies designed to end social promotion grow in popularity.

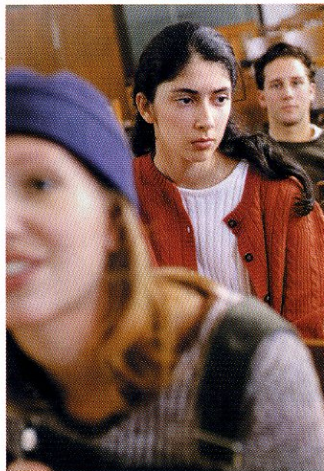
In 1997, Chicago initiated a district policy designed to base grade 3, 6, and 8 promotion and retention decisions on test scores. In its first year, almost 26,000 students—32 percent of 3rd graders, 31 percent of 6th graders, and 21 percent of 8th graders—failed the test. After summer school, 15 percent of the 3rd graders, 13 percent of the 6th graders, and 8 percent of the 8th graders were retained. Since 1997, about 50,000 students have been retained in Chicago because of low test scores. Researchers found that Chicago students retained before high school were 12 percent more likely to drop out before graduating (Hauser, 2001; Woestehoff, 2000).

In Louisiana, between 10 and 15 percent of 4th and 8th graders were retained in 2000 because they failed the state's high-stakes test (Robelen, 2000). The great majority of them were from racial minority and economically disadvantaged backgrounds.

Even before they actually take the test, struggling students are more likely to be retained in grade if they attend schools in high-stakes testing environments. By holding low-achieving students back, schools ensure that these students have more of the knowledge necessary to perform well on high-stakes tests the next year—and also keep the low-performing students' test scores out of the composite test performance in the grades in which high-stakes tests matter.

In Texas, students from racial minority and low socioeconomic backgrounds are being retained in grade 9 at very high rates before taking the Texas Assessment of Academic Skills (TAAS) in grade 10. Many teachers retain students if they doubt their potential to pass the TAAS the following year. McNeil (2000) estimated that half of all minority students enrolled in Texas high schools are technically enrolled as freshmen.

Although some of them are 9th graders for the first time, thousands of others have been retained in the 9th grade once or even twice. Other researchers (Haney, 2000, 2001; Klein, Hamilton, McCaffrey, & Stecher, 2000; Yardley, 2000) have verified her numbers. In 1998, one in every four African American and Latino 9th graders in Texas was retained (Fisher, 2000). After these



© Doug Menza/Getty Images

High-stakes tests often decrease student motivation and lead to higher student retention and dropout rates.

students are retained, thousands of them drop out of school.

Retention in grade does not motivate students to learn more or perform better. Instead, retention motivates many students to leave school early. In some ways, this problem may be worse than the problems that the high-stakes testing policies are designed to fix.

Effects on Student Learning

After a state implements high-stakes testing policies, scores on the state's assessments often improve. Students can easily be trained so that scores on

the state tests go up. For example, scores can be made to rise by narrowing the curriculum. Art, music, creative writing, physical education, recess, ROTC, and so forth are all reduced in time or dropped from the curriculum when schools need to increase their scores on the state tests. Even in the curriculum areas that are tested, schools may drop sub-areas if they are unlikely to appear on the test. So if quadratic equations are not tested in the state's own mathematics tests, then quadratics may not be taught as a part of algebra. Instructional time is shifted to the curriculum areas that will appear on the tests, and consequently scores on the state tests go up.

High-stakes tests cause other problems for the schools as well. Schools often emphasize drill activities and use district funds to buy test preparation materials that are supposed to increase scores, regardless of the fact they undermine the validity of the tests. Unfortunately, the tests also corrupt some teachers, administrators, and students so that they even feel compelled to cheat.

These common problems of high-stakes testing programs are quite likely to affect the breadth and depth of student learning. If schools narrow the curriculum they teach; make heavy use of drill activities tied to the state test; cheat by over-identifying language-minority and special education students and then keeping these students from taking the tests; retain poorly performing students in grade; and encourage those who are least likely to pass the state's tests to drop out, then scores on state tests will almost certainly go up. But have students really learned any more than they did before high-stakes testing policies were instituted?

Results from our 18 high-stakes testing states allow us to study this question. If statewide high-stakes testing policies actually improve student learning, we should see that improvement reflected not just in the states' own test scores but also in independent measures. For each of the 18 states, we looked at four well-respected student achievement measures: the SAT, the

ACT, Advanced Placement (AP) tests, and the National Assessment of Educational Progress (NAEP).

What did we find when we undertook these analyses? Nothing! Nothing much seemed to be happening on these measures of student learning. In fact, we can make a much stronger case that high-stakes testing policies hurt student learning instead of helping it. Here is how we came to this conclusion.

For each of the four independent measures of student learning (the SAT, ACT, AP tests, and NAEP) we did an archival time-series analysis² using the state data on each measure and comparing it to the national data for each measure. Figure 2 provides an example. Here we plotted national SAT scores and state SAT scores for New York from 1977 through 2001. New York's first high-stakes high school grad-

uation exam went into effect for the class of 1985; its second high-stakes high school exam first affected the class of 1995. In Figure 2, we have highlighted the year before the high-stakes tests were implemented with a diamond (◆). Right after these points, high-stakes testing policies should demonstrate a visible impact on student learning.

Instead, SAT scores in New York actually lost ground compared with

FIGURE 1

States That Make Extensive Use of High-Stakes Testing

State	Stakes							
	Graduation contingent on a high school graduation exam	Grade-to-grade promotion contingent on a promotion exam	State publishes annual school or district report cards	State identifies low-performing schools according to whether they meet state standards or improve year-to-year	Monetary awards given to high-performing or improving schools	State has authority to close, reconstitute, or revoke accreditation or take over low-performing schools	State has authority to replace school personnel—principals or teachers—due to low test scores	State permits students in failing schools to enroll elsewhere
Alabama	X	—	X	X	X	X	X	—
Florida	X	—	X	X	X	—	—	X
Georgia	X	2004	X	X	X	2004	—	—
Indiana	X	—	X	X	X	X	—	X
Louisiana	X	X	X	X	—	X	X	X
Maryland	X	—	X	X	X	X	X	—
Minnesota	X	—	X	—	—	—	—	—
Mississippi	X	—	X	X	2003	2003	—	—
Nevada	X	—	X	X	—	X	X	—
New Jersey	X	—	X	X	X	—	—	—
New Mexico	X	X	X	X	X	X	X	—
New York	X	—	X	X	—	X	X	—
North Carolina	X	X	X	X	X	X	X	—
Ohio	X	2002	X	X	X	—	—	—
South Carolina	X	2002	X	X	X	X	X	—
Tennessee	X	—	X	X	X	X	—	—
Texas	X	2003	X	X	X	X	X	X
Virginia	X	—	X	X	—	X	—	—

Source: Information gathered through interviews with the state testing directors and other testing personnel. Each testing director verified his/her state's data in the table.

A Research Report

those throughout the United States. After implementation of the first high-stakes graduation test, students in New York who took the SAT lost 3 points from 1984–1985 compared with students in the United States overall. In the longer term, from 1984–1994, New York students who took the SAT lost 11 points compared with students in the nation as a whole. Thus, the first high-stakes test apparently did not raise student achievement as reflected on the SAT.

After the second high-stakes graduation test went into effect, the short-term results from 1994–1995 again show that students in New York lost 3 points compared with the rest of the nation. The long-term results were also negative, with New York students losing 6 points compared with the nation from 1994–2001.

We performed this type of analysis for each of the 18 states, looking at each of the four achievement measures and also considering changes in participa-

High-stakes testing policies have resulted in no measurable improvement in student learning.

tion rates in each. In this way, we evaluated the effects of the high-stakes testing policies in a particular state on student learning, as measured by the four independent assessments of learning available.

SAT Changes

When we looked at the SAT results across the 18 states in which one or more high-stakes tests were implemented over time, we found 17 short-term positive effects (that is, SAT scores went up), 13 short-term negative effects (SAT scores went down), and one case in which scores did not change. In the long-term analyses, SAT scores went up

in 15 cases and down in 16. These results indicate that high-stakes testing policies have no systematic effects on student learning.

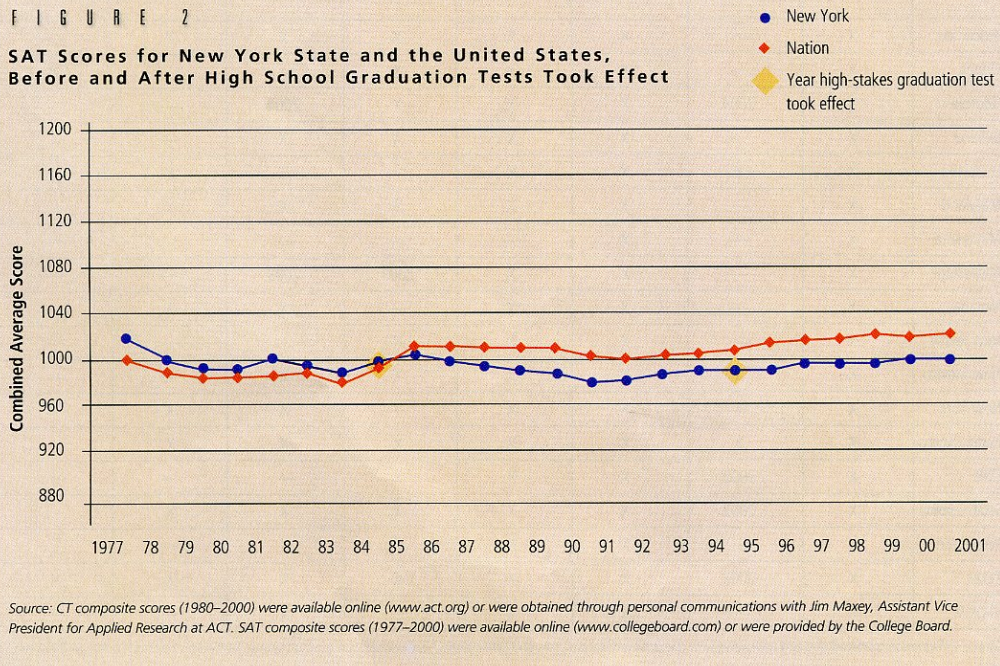
We also found that students participated in the SAT testing program at lower rates after high-stakes high school graduation exams were implemented than before. SAT participation rates, compared with those of the United States as a whole, increased in 7 states and decreased in 11 states after the point at which the graduation exams were implemented.

ACT Changes

Sixty-seven percent of the states that use high school graduation exams posted *decreases* in ACT performance after they implemented such exams. On average, the academic achievement of college-bound students as measured by the ACT decreased in states with high-stakes high school graduation exams. ACT participation rates, compared with national rates, increased in 9 states,

FIGURE 2

SAT Scores for New York State and the United States, Before and After High School Graduation Tests Took Effect



decreased in 6 states, and stayed the same in 3 states from 1994–2001.

AP Test Changes

Controlling for participation rates, from 1995–2000, 57 percent of the states with high-stakes high school graduation exams posted losses in the percentage of students passing AP exams with a grade of 3 or better (out of 5). Moreover, compared with the nation, 67 percent of the states with these high-stakes policies posted losses in the percentage of students who participated in AP programs. These data provide no evidence of increased learning or increased motivation to take the rigorous AP courses.

NAEP Changes

When we looked at the 4th grade NAEP mathematics test from 1992–2000, 50 percent of the states with high-stakes testing policies posted increases in composite math performance compared with the rest of the nation; 50 percent of the states showed either losses or no effects. On the 8th grade NAEP mathematics test from 1990–2000, only 36 percent of the states with high-stakes testing policies posted increases in composite math performance compared with the rest of the nation; 64 percent of the states showed either losses or no effects.

On the 4th grade NAEP reading test from 1992–1998, only 46 percent of the states with high-stakes testing policies posted increases in composite reading performance compared with the rest of the nation; 54 percent of the states showed either losses or no effects.

We also looked at the same kinds of students over time, as they moved from 4th to 8th grade in mathematics and reading. In mathematics, from 1996–2000, 62 percent of the states with high-stakes testing policies posted student losses in mathematics achievement compared with the United States as a whole. In reading, however, from 1994–1998, 69 percent of states with high-stakes testing policies posted student *gains* in comparison with the rest of the nation. This is the only case showing some evidence of increased

learning in connection with high-stakes testing policies. In all other cases, the data do not support the idea that high-stakes testing policies increase student learning.

Our analysis of the NAEP data also revealed that changes in states' test scores were affected by the exclusion rates the states used. State scores went up or down depending on the numbers of students who were kept out of the pool of eligible test-takers. Thus, the great growth in NAEP test scores in both North Carolina and Texas turns out to be a function of the fact that these states excluded more students from NAEP testing than did the other states.

It's time to abandon high-stakes policies and substitute more formative testing programs.

In summary, when we look at 18 states with high-stakes testing policies, we find that such policies have resulted in no measurable improvement in student learning, as indicated by four different independent measures.

A Better Way

If the data for 18 states with high-stakes testing policies foreshadow what will happen as we implement the high-stakes policies written into current federal legislation, we risk reducing student motivation to learn, driving more students and teachers out of our schools, and becoming a less educated, less learned people. Although test scores will rise and our politicians will be placated, we will have hurt our public education system.

As we think about testing policies, we should remember the wisdom in the farmer's comment that weighing a pig every day won't ever make the pig any fatter. Eventually, you have to feed the pig.

Weighing or assessing may not work, but everyone looks busy doing it, and it costs much less than providing all

students, including poor and minority students, with high-quality preschools, small class sizes in the early grades, well-qualified teachers, adequate medical attention, and so forth. It's time to abandon high-stakes policies and substitute more formative testing programs which, when they uncover poor school performance, result in fiscal, intellectual, and social reforms that will make a difference for the students in those schools. ■

¹ The number of states with high-stakes high school graduation tests has escalated almost linearly in the past 20 years, rising from just 3 in 1983 to 18 in 2002. In addition, nine states currently have graduation exams under development: Alaska, Arizona, California, Delaware, Hawaii, Massachusetts, Utah, Washington, and Wisconsin.

² Time-series studies are particularly well suited to determining the impact of large-scale social or government policies. In time-series designs, strings of observations of the variables of interest are made before and after some policy is introduced. The effects of the policy, if any, are shown by the rise and fall of scores on the variable of interest (Amrein & Berliner, 2002b).

References

- American Council on Education. (2001). *Who took the GED? GED 2000 statistical report*. Washington, DC: Author.
- Amrein, A. L., & Berliner, D. C. (2002a). Figures calculated using 1998 data from *KidsCount data book online* [Online]. Available: www.aecf.org/kidscount/kc2001
- Amrein, A. L., & Berliner, D. C. (2002b). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives*, 10(18). [Online]. Available: <http://epaa.asu.edu/epaa/v10n18>
- FairTest & Massachusetts CARE. (2000). MCAS: Making the dropout crisis worse [Online]. Available: www.fairtest.org/care/MCAS%20Alert%20Sept.html
- Fisher, F. (2000). *Tall tales? Texas testing moves from the Pecos to Wobegon*. Unpublished manuscript.
- Goldschmidt, P., & Wang, J. (1999). When can schools affect dropout behavior? A longitudinal multilevel analysis. *American Educational Research Journal*, 36(4), 715–738.
- Haney, W. (2000). The myth of the Texas miracle in education. *Education Analysis Policy Archives*, 8(41). [Online]. Available: <http://epaa.asu.edu/epaa/v8n41>

A Research Report

Haney, W. (2001). *Revisiting the myth of the Texas miracle in education: Lessons about dropout research and dropout prevention*. Paper prepared for the Dropout Research: Accurate Counts and Positive Interventions Conference, June 13, 2001, sponsored by Achieve and the Harvard Civil Rights Project, Cambridge, Massachusetts. [Online]. Available: www.law.harvard.edu/civilrights/publications/dropouts/dropout/haney.pdf

Hauser, R. M. (2001). Should we end social promotion? Truth and consequences. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Foundation Press.

Jacob, B. A. (2001). Getting tough? The impact of high school graduation exams. *Education Evaluation and Policy Analysis*, 23(2), 99-121.

Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives*, 8(49). [Online]. Available: <http://epaa.asu.edu/epaa/v8n49>

McNeil, L. (2000). *Contradictions of school reform*. New York: Routledge.

Murnane, R. J., Willett, J. B., & Tyler, J. H.

(2000). Who benefits from obtaining a GED? Evidence from High School and Beyond. *Review of Economics and Statistics*, 82(1), 23-37.

Orfield, G., & Kornhaber, M. L. (Eds.). (2001). *Raising standards or raising barriers? Inequality and high-stakes testing in public education*. New York: The Century Foundation Press.

Robelen, E. W. (2000, May 24). La. set to retain 4th, 8th graders based on state exams. *Education Week*, p. 24. [Online]. Available: www.edweek.org/ew/ewstory.cfm?slug=37louisiana.h19

Rothstein, R. (2002, October 9). Dropout rate is climbing and likely to go higher. *New York Times*, p. 8.

Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus Books.

Sheldon, K. M., & Biddle, B. J. (1998). Standards, accountability, and school reform: Perils and pitfalls. *Teachers College Record*, 100(1), 164-180.

Wheelock, A., Bebell, D. J., & Haney, W. (2000). What can student drawings tell us about high-stakes testing in Massachusetts? *Teachers College Record* [Online]. Available: www.tcrecord.org/Content/asp?ContentID=10634

Woestehoff, J. (2000). Chicago's flunking policy gets an F. In K. Swope & B. Miner (Eds.), *Failing our kids: Why the testing craze won't fix our schools*. Milwaukee, WI: Rethinking Schools.

Yardley, J. (2000, October 30). Critics say a focus on test scores is overshadowing education in Texas. *New York Times*, p. 14.

Authors' note: A longer version of this article is available online at <http://epaa.asu.edu/epaa/v10n18>. The work on this project was funded in part by the Rockefeller Foundation. The opinions expressed are entirely those of the authors and may not represent those held by the Rockefeller Foundation.

Copyright © 2003 Audrey L. Amrein and David C. Berliner.

Audrey L. Amrein is a research professional at the College of Education of Arizona State University, Tempe, AZ; amrein@asu.edu. **David C. Berliner** is Regents Professor of Educational Leadership and Policy Studies at the College of Education of Arizona State University; berliner@asu.edu.

If attending the ASCD conference in San Francisco, please visit us at booth #1912

SRA

Everyday Mathematics
The University of Chicago School Mathematics Project

Bridging the years with mathematics excellence

Since 1988, *Everyday Mathematics* has proven to educators across the country what great teaching methods coupled with great materials can accomplish. This dynamic math program, created by the University of Chicago School Mathematics Project, ranks among the most successful math programs available.

We'd like to say a heartfelt Thank You to these long-time user districts:
Edina, MN since 1989 • Newton, MA since 1991 • Evanston, IL since 1992

We would also like to welcome the newest members to the *Everyday Mathematics* family in 2002, including:
Adams 12 Five Star, Thornton area, CO
Newark, NJ • Washoe County, Reno, NV

We look forward to a long relationship with you!

Questions? Call 1-888-SRA-4543 for the answers.
www.sra4kids.com



The McGraw-Hill Companies