

December 28, 2002 THE NEW YORK TIMES

## More Schools Rely on Tests, but Study Raises Doubts

By GREG WINTER

**R**igorous testing that decides whether students graduate, teachers win bonuses and schools are shuttered, an approach already in place in more than half the nation, does little to improve achievement and may actually worsen academic performance and dropout rates, according to the largest study ever on the issue.

With calls for accountability in public education mounting, such make-or-break exams have become cornerstones in at least 28 states in the drive to improve public schools. The idea is that by tying test scores to great consequences, the learning process will be taken that much more seriously and tangible progress will be all the more likely.

The approach is also central to some of President Bush's sweeping education overhaul, lending even greater momentum to the movement known as "high stakes" testing.

But the study, performed by researchers at Arizona State University and financed by teachers' unions that have expressed skepticism about such tests, found that while students show consistent improvement on these state exams, the opposite is typically true of their performance on other, independent measures of academic achievement.

For example, after adopting such exams, twice as many states slipped against the national average on the SAT and the ACT as gained on it. The same held true for elementary-school math scores on the National Assessment of Educational Progress, an exam overseen by the United States Department of Education.

Trends on Advanced Placement tests were also worse than the national average in 57 percent of those states, while movement in elementary-school reading scores was evenly split — better than the national average in half the states, worse in the other half. The only category in which most of the states gained ground was middle-school math, with 63 percent of them bettering the national trend.

"Teachers are focusing so intently on the high-stakes tests that they are neglecting other things that are ultimately more important," said Audrey Amrein, the study's lead author, who says she supported high-stakes tests before conducting her research. "In theory, high-stakes tests should work, because they advance the notions of high standards and accountability. But students are being trained so narrowly because of it, they are having a hard time branching out and understanding general problem-solving."

The study was commissioned by the Great Lakes Center for Education Research and Practice, a Midwestern group of six state affiliates of the National Education Association,

which has opposed using any one test to determine when students graduate, schools get more money and teachers are replaced. The research is sure to be a subject of fierce debate among educators, and its methodology has already drawn some criticism, though an independent panel of researchers at other universities has concluded that the findings are valid.

Perhaps most controversial, the study found that once states tie standardized tests to graduation, fewer students tend to get diplomas. After adopting such mandatory exit exams, twice as many states had a graduation rate that fell faster than the national average as those with a rate that fell slower. Not surprisingly, then, dropout rates worsened in 62 percent of the states, relative to the national average, while enrollment of young people in programs offering equivalency diplomas climbed.

The reason for this is not solely that struggling students grow frustrated and ultimately quit, the study concluded. In an echo of the findings of other researchers, the authors asserted that administrators, held responsible for raising tests scores at a school or in an entire district, occasionally pressure failing students to drop out.

In lawsuits, educators have testified that students were held back rather than promoted to a grade in which high-stakes tests were administered, and that others were expelled en masse shortly before testing days. But neither those witnesses nor this study has been able to quantify that circumstance nationally, or prove that it has substantially influenced dropout rates.

As the popularity of do-or-die exams has increased, educators have vehemently argued their merits and drawbacks, focusing mainly on individual states, like Texas and Massachusetts, where their adoption has spurred the most controversy.

But this study is the first to have looked at the issue nationally. The study examined graduation rates and scores from a variety of tests in more than two dozen states that have turned to do-or-die exams over the last two decades in hopes of raising academic performance.

"This is not research by press release, this is serious work," said Sherman Dorn, a historian of education at the University of South Florida who reviewed it. "What's very clear is that the study challenges the conventional wisdom that high-stakes testing improves academic achievement and does not have unwanted consequences beyond that."

The study has drawn its share of detractors, in no small part because one of its authors, David Berliner, has been a critic of school vouchers and other education proposals often championed by conservatives.

"I've gotten this reputation of being outspoken," Mr. Berliner said. "Some call it ideological; I call it honest. Either way, the data speaks for itself."

Soundness of the data aside, some of Mr. Berliner's critics question whether such tests are to blame for the poor showings.

"You almost never have a pure cause-and-effect relationship," said Chester E. Finn, assistant secretary of education in the Reagan administration. "Yes, you're introducing high-stakes tests, but maybe you're also changing the way you license teachers, or extending the school day, or changing textbooks. There's always a lot of things going on concurrently, so you really cannot peg everything to the high-stakes tests."

Other skeptics challenged the fairness of holding up the SAT, the ACT, Advanced Placement tests and the national math and reading exams as indicators of academic performance, even if they are the only nationally administered tests with which to measure one state against another.

For example, the National Assessment of Educational Progress test, administered every four years, "gives us a nice eyeball assessment, but the problem is it's given infrequently," said Jay P. Greene, a senior fellow at the Manhattan Institute who is working on a similar though more limited study.

"And the college entrance tests are very bad at judging learning," he said, "because only a modest number of students actually take the test."

The criticism notwithstanding, many researchers said the study fell within the bounds of what was already known about make-or-break exams. Educators have long complained that the threat of serious consequences means that teachers focus on little else, sometimes building their lesson plans entirely around the contents of the test.

That would not necessarily be a problem if the state exams were based on a comprehensive curriculum, said Eva Baker, co-director of the National Center for Research on Evaluation, Standards and Student Testing at U.C.L.A. But as often as not, the state exams are given in the absence of such a framework, leaving teachers to fill in the gaps on their own, sometimes with an overzealous reliance on test-taking drills.

"The most perverse problem with high-stakes tests," Ms. Baker said, "is that they have become a substitute for the curriculum instead of simply a measure of it."

Some researchers suggested that the study might have actually understated the consequences of high-stakes tests, particularly for dropout rates, because it relied on government statistics. "Officially reported dropout statistics are pretty suspect in a lot of places," said Walter Haney, a professor at the Lynch School of Education at Boston College, pointing out that students who leave school to get a G.E.D. are not always counted as dropouts. "The real results are probably worse."

A larger question raised by the study is what effect, if any, it will have on the public debate over high-stakes testing. While many educators will most likely hold it up as proof that such exams are flawed, largely because they appear to offer inadvertent

encouragement to schools to constrain the curriculum and squeeze out underachievers, others see the issue as more open-ended.

"Should we just make better tests," asked Anthony G. Rud Jr., associate professor of education at Purdue University, or "is there something fundamentally wrong with testing in this matter?"