HIGH STAKES TESTING

"This is the year that U. S. schools went test-crazy." Thus reads the opening line of an April 16, 2000, article by David Bacon in the *Oakland Tribune*. The statement contains one inaccuracy--the schools didn't so much go crazy for the tests, as they went crazy trying to cope with the tests that were imposed on them from governors, legislators, and state boards of education, all cheered on by business and industry. Other than that, Bacon captured the feelings of many people observing education.

To be sure, this was not a sudden madness. In the 1970's and early 1980's, some 35 states had adopted some version of a "minimum competency test" to assure that high school diplomas were not based on "social promotion" and/or "seat time." And in 1977, a report on the apparent SAT decline had made every minute change in that test score front page news (at least, when the scores went down; when the scores went up, that result got buried with local news).

The new illness, though, was more virulent. When Bacon penned his article, all but one state had adopted or created standards for public school students and 41 had adopted or constructed tests for measuring and passing judgment on student performance.

Students now are more at risk of not graduation than in the MCT days because the tests are tougher and/or the scores needed to pass are unrealistically high. Fully 90% of the students in Arizona failed at the first administration as did 98% of Virginia's schools. In addition, students are being retained in grade or forced to attend summer school based on test scores. Proposals exist to start testing students in kindergarten. Teachers are warned that their raises, bonuses or even their jobs are on the line. Principals and superintendents suffer similar threats. While the emphasis has been on the negative, on rarer occasions, the bonuses of teachers, principals and superintendents are tied to specific test score gains. Whereas tests were once used largely as monitoring devices, they now have enormous consequences for many people. Hence the catch-phrase "high-stakes testing."

Looking at the frenzy about testing, two questions immediately come to the fore: Why? And "Are the Testing programs having their desired impact? The short answer to the first question is "A loss of trust in teachers and administrators." The answer to the second question is "No." Let us consider each question in detail in turn.

WHY DID AMERICANS BECOME NERVOUS ABOUT THEIR PUBLIC SCHOOLS?

American public schools have always suffered from criticism, but the critics became more numerous and more vocal shortly after World War II. As the nation moved towards universal secondary education, it was also engaged in a space and arms race with the Soviet Union. For the first time, schools were perceived by people such as Admiral Hyman Rickover as integral to national defense. Colleges would, of course, prepare the engineers, scientists and mathematicians needed to meet the Red Menace, but those colleges and universities had to start with the products of high schools and some felt that these schools were not up to snuff. The very fact that schools were holding and graduating a larger and larger proportion of students increased the anxiety as people feared that the increased graduation rate might reflect a decrease in rigor. When the Russians launched Sputnik in 1957, the beeps emitted by that small satellite proved to the critics that they had been right.

Schools never fully recovered from Sputnik, but received another severe blow in 1983 with the publication of "the paper Sputnik", "A Nation At Risk." In 1996, the *New York Times* observed that that document was merely propaganda, but it was not recognized as such in many quarters at the time (Applebome, 1997). Its highly selective and negatively spun statistics were used as a "clarion call" to overhaul the schools.

The anxiety people might feel about their schools was heightened by the fact that, as journalist Peter Schrag observed, good news about schools served no one's political education reform agenda. The Reagan and Bush administrations, pushing privatization, vouchers and tuition tax credits, actively suppressed positive data where it could and ignored positive data where it could not actually control the flow of information.

Thus, a 1992 international study in mathematics and science which found American ranks mostly (but not entirely) low, was given a large press conference by the U. S. Department of Education. An international study in reading that found American students second in the world was ignored. A large analysis of the U. S. public school system by Sandia National Laboratory engineers was suppressed for being too positive. U. S. Department of Education officials denied that the report was suppressed, but Lee Bray, the now-retired Vice President of Sandia National Laboratories responsible for the report is emphatic that it was (personal communication, August 200).

The Clinton-Gore years have seen an increased press for additional resources for public schools, but they have emphasized the problems of schools that require the resources. American universities use a similar approach in their attempts to obtain funding from governments and foundations. And, as it has for the last 100 years, business and industry has found American education wanting and has tried to prescribe what is to be taught.

The consequence of this negativity coming from so many sources is that virtually everyone is willing to believe the worst about the schools. For instance, in the mid-1980's two lists appeared showing the most serious problems in the schools in the 1940's and in the 1980's. In the 1940's, schools were plagued by students talking out of turn, not raising their hands, chewing gum in class, and breaking in line. In the 1980's, drugs, violence, gangs and teen pregnancy had become the most serious problems. Yale University professor, Barry O'Neill found that many people along the entire political spectrum assumed that the lists were based on research and were true. O'Neill revealed them as a hoax (O'Neill, 1994).

All of the above events contributed to a feeling that the people running the schools could not be trusted to provide accurate information on what students were or were not learning. Something more "objective" was needed, something that did not depend on the "subjective" judgments of teachers. That something in most instances turned out to be a test.

Is the public's nervousness warranted? Not according to the data--mostly test data--that exist. There are many aspects of schooling that cannot be measured with tests, but tests are the major source of data we currently have that permit comparisons of schools, states, or nations. What do these test show?

----That standardized achievement tests attained record high levels in the mid to late 1980's and remain there.

----That scores on the National Assessment of Educational Progress have risen to all-time highs. Gains have been especially dramatic for blacks and Hispanics.

----That the proportion of students scoring above 650 on the SAT mathematics section attained record levels around 1995 and has remained at the all-time high. This cannot be accounted for by Asian-American students who are too few in number. Of the 75% increase between 1981 and 1995, black, white, Hispanic and Native Americans accounted for 57%.

----That the number of students taking Advanced Placement examinations his risen from just over 1000 in 1961 to over 1,000,000 currently.

----That American students are second in the world in reading.

It would thus seem that the condition of public education, insofar as it can be adequately assessed by existing test instruments, shows no cause for alarm. Even if the raison d'etre of the high stakes testing programs were missing, they could still be acceptable programs if they were shown to be causing achievement to increase. A check of the data, though, not only fails to find such improvements, but uncovers a gaggle of unfortunate outcomes.

Examining these outcomes brings to the extended answer of the second question: Are the tests having their intended impact?

Before beginning to answer this question in specific terms, we can do no better than to quote from an article by Robert L. Linn in the March, 2000 *Educational Researcher.* Linn is probably the most respected psychometrician in the nation with a reputation for reasoned discourse and objectivity. After exhaustively examining the evidence surrounding the impact of high-stakes testing. Linn concluded thusly:

> As someone who has spent his entire career doing research, writing, and thinking about educational testing and assessment issues, I would like to conclude by summarizing a compelling case showing that the major uses of tests for student and school accountability during the past 50 years have improved education and student learning in dramatic ways.
>
> Unfortunately, that is not my conclusion. Instead, I am led to conclude that in most cases the instruments and technology have not been up to the demands that have been placed on them by high-stakes accountability. Assessment systems that are useful monitors lose much of their dependability and credibility for that purpose when high stakes are attached to them. The unintended negative effects of the high-stakes accountability uses often outweigh the intended positive effects.

Before examining these "unintended negative consequences," we should make the general point that there is a gap between what test scores reveal and what people want to know. In a June 11, 2000 article, *New York Times* reporter, Anemona Hortocollis put it this way: "In the war of perception against reality, almost nothing can be harder to gauge than the meaning of test scores…Yet parents and teachers are encourage [to use tests] to judge their children and schools the way investors watch the Dow industrials."

Thus, one of the negative consequences of high-stakes testing is to drive a wedge between parents and their children. Parents, having watched their children for years, have a feel for "what they are about." But the test might say otherwise. Fortunately, most parents are skeptical about what tests say. A poll by the American Association of School Administrators found that two thirds of parents say a test can't measure a child's progress and half say that tests don't reflect what children know.

On a more societal level, high stakes testing is increasing social stratification. These tests have been presented by persons such as Diane Ravitch and E. D. Hirsch, Jr., as engines of social justice. No one has explained how, given the enormous differences in scores for affluent and majority students, such tests will improve the chances of success for poor and minority children. For instance, on the Virginia Standards of Learning U. S. History test, required for graduation, only 13% of black students and 23% of Hispanic students passed, compared to 40% of white students. And this was on second administration, after a year of intense preparation for the test. Similar gaps were found on all tests. For instance, 76% of white students passed the Algebra I test, while only 36% of blacks and 49% of Hispanics scored high enough to pass. When statewide tests were introduced in Texas, the dropout rates for black and Hispanic students rose sharply and have not returned to previous levels.

Under the gun of the tests, teachers are abandoning their usual curricula and modes of teaching to lecture about test-oriented material. In many instances, they are omitting aspects of the curriculum not on the test. One local school board in a large Virginia district held a special session to determine if they needed to mandate recess for their elementary schools because so many of them had abandoned it in favor of test preparation. In Texas, where science and social studies were not initially included in testing, teachers reported that those subjects virtually disappeared. When the science and social studies tests appeared, science and socials studies were quickly geared to what those tests tested.

Tests can easily misrepresent the achievements of a school. For instance, six high schools in Miami-Dade and Broward County, Florida, made the College Board's list of the top 100 high schools in the entire nation, based on the number of Advanced Placement examinations taken per student. Yet, in the Florida state accountability system, which grades schools from A to F, all six received a grade of C.

To be sure, there are limitations to using AP exams as a measure of quality, but the differing pictures painted by the different measures point to another problem afflicting many high-stakes programs: severe judgments are being made on the basis of a single test score. The standards for test use promulgated jointly by the American Psychological Association, the American Educational Research Association and the National Council on Measurement in Education say clearly that no decisions about human beings should be made in such a way. Even the commercial test developers who are realizing enormous profits from the test boom, concur.

Parents, teachers and students are rebelling against these tests in various ways, another indication that the people most affected by the tests do not find them to be healthy for children. Some parents simply refuse to permit their children to take the tests, while others openly organize for their repeal. Some students, on pain of suspension, refuse to take the tests. And some principals and teachers, faced with instruments they do not believe validly assess what their students know, cheat.

Everyone in the nation supports efforts to improve schools. But there is a growing realization among many, perhaps most people, that the imposition of high stakes tests carries, as Linn wrote, unintended negative consequences that defeat that purpose of improvement. Any presidential candidate should provide a well thought out program for developing a more thoughtful and humane program of school accountability.

Applebome, Peter, "Dire Prediction Deflated: Johnny Can Add After All." *New York Times,* June 11, 1997, p. A31.

Bacon, David, "Value of School Testing Open to Question." *Oakland Tribune,* April 16, 2000.

Hartocollis, Anemona. "Test Scores Are Up. From This We Can Conclude?" *New York Times,* June 11, 2000.

Linn, Robert, "Testing and Accountability." *Educational Researcher,* March, 2000, p. 4-15.

O'Neill, Barry, "Anatomy of a Hoax." *New York Times Sunday Magazine,* March 6, 1994, pp. 46-49.