

Media Report

California's Stanford 9/STAR Testing System: Rampant Cheating Invalidates the Results by Alan Bonsteel, M.D., July 6, 2000

INTRODUCTION

Governor Gray Davis has declared that testing is his "polar star of education" and has announced that he will not run for re-election as governor in 2002 if test scores do not improve. However, the state's current testing system, the Stanford 9/STAR test, is deeply flawed. Championed by Governor Davis's predecessor, Pete Wilson, and put into place by a reluctant California Department of Education, this slipshod testing system may tell us little about how California's K-12 students are doing — including even whether student achievement is improving or worsening.

An initiative for school choice has now qualified for the November ballot. Given that kind of pressure, it is likely that public school officials will do anything they can to be able to claim that test scores are improving and that fundamental reform of the school system is therefore unnecessary.

BACKGROUND

The predecessor of the current system, the CLAS (California Learning Assessment System) test, was given to California children in 1992 and 1993. It was ended after only two years, owing to its statistical flaws, its intrusive questions, to which Governor Wilson and Republican legislators objected, its failure to provide parents with individual test scores, and the embarrassment of the California public education establishment with the results. From 1994 through 1997, California had no statewide testing system at all; there was only a patchwork approach in which different districts used different tests.

During his final years in office, Governor Wilson expended considerable political capital to achieve his vision of a rigorous statewide test that provided individual test scores to parents. The result, the Stanford 9/STAR test, was first given in 1998, Wilson's final year in office. This year will be the third time it has been given in California.

Opposition from the public education establishment to the kind of rigorous testing that Wilson wanted was so great that a number of compromises were required to make the testing politically feasible. The most important of these compromises is that test scores for individual classrooms are not available. Thus, the Stanford 9/STAR test is of little or no use to principals and school boards in evaluating the performance of individual teachers. It is also of little use to parents who would like to support their school administrators in holding teachers accountable.

COMPONENTS OF THE STANFORD 9/STAR TEST

The Stanford 9/STAR test consists of the following two components, both published by Harcourt Educational Measurement:

? The nationally normed Stanford 9 (Stanford Achievement Test, Ninth Edition, Form T) test, administered to all students in grades 2 through 11. Sample questions from the Stanford 9 test are included in the appendix of this report.¹

? The STAR (Standardized Testing and Reporting) augmentation, intended to reflect California's academic standards as enacted by the California State School Board. This test is also given in grades 2 through 11, except for the mathematics part in grades 8, 9, and 10, where only students enrolled in certain math classes are tested.

The SABE/2 (Spanish Assessment of Basic Education, Second Edition) test is intended to be administered to Spanish-speaking students who have been in California schools for 12 months or less. It is published by CTB/McGraw-Hill. There are no tests given statewide in any language other than Spanish, putting Limited English Proficient students whose home language is other than English or Spanish at a considerable disadvantage.

DISCONNECT BETWEEN STANFORD 9/STAR TEST AND SCHOOL CURRICULUM

The Stanford 9 test is a "nationally normed" test. No one claims that it is aligned with the curriculum actually taught in California.² An obvious major discrepancy is that although U.S. history is normally taught in the 11th grade in California schools, it is tested in the 10th grade on the Stanford 9 test.¹²

The STAR "augmented" portion of the test was intended to correspond to the statewide standards established by the California State Board of Education. However, only the reading and English questions are considered to be reliable for that purpose.³

The disconnect between the Stanford 9/STAR test and the California curriculum is a result of California's system of educational governance, which has been characterized as "hydra-headed" since at least the 1920s. Responsibility for California's public schools is divided among the governor, the governor's secretary for education, the legislature, a California State Board of Education appointed by the governor, an elected superintendent of public instruction overseeing the California Department of Education, and the Commission on Teacher Credentialing. Under the California constitution and state statutes, the California State School Board adopts instructional materials and is the policy-maker for California's schools, with the elected superintendent of public instruction designated to carry out the policies.

In recent years, however, the California State School Board and the superintendent

have warred. After establishing academic standards for California's schools, the board sought to implement those standards by choosing instructional materials aligned to the standards and mandating tests based on the standards, hoping that this pressure would force the schools to teach to the standards. The California Department of Education, however, has not clearly communicated those standards to the state's teachers, nor do the majority of current textbooks (when available) teach to the standards, as a result of use of obsolete textbooks or of districts simply ignoring the established standards. Furthermore, at least in theory the Stanford 9 and STAR tests are "secure" tests, with the limited number of Stanford 9 questions reproduced in this report the only ones that have been made available to the public — or to teachers! In theory, when teachers pass out the tests, they are supposed to avert their eyes and not note what questions are being asked. Thus, teachers are doubly in the dark as to what the statewide academic standards are supposed to be.

The following letter to be the editor of the San Francisco Chronicle, published on July 5, 2000, well expresses the profound frustration of one teacher, who, after he had done his best to teach his students the official curriculum for that grade level, discovered that they had been tested on material that was officially placed in the following year:

CHEATING IS RAMPANT

Believe it or not, the Stanford 9 test questions are repeated exactly the same from year to year! Not surprisingly, therefore, copies of actual examinations are in wide circulation. The STAR test is also in wide circulation. Unlike the Stanford 9, however, this year, for the first time, one-quarter of the questions on the STAR test are new — leaving 75% unchanged.¹

California has already seen two widely publicized episodes in which public school teachers were caught giving advance copies of the Stanford 9 test to their students: in Los Angeles Unified^{4,5} and in Woodland.⁶ One has to wonder how many more such episodes have gone undetected or unreported.

In preparing this report, I interviewed three public school teachers. While all three stated that they would not allow cheating in their own classrooms, two of them said that they could easily obtain old copies of the Stanford 9 test if they wished. In addition, a former district school board member said that he had possession of several grade levels of the Stanford 9 test.

A recent Sacramento Bee article stated that 12 of California's 992 public school districts reported "irregularities" in the Stanford 9 test.⁶ As a result, in early May, the California State Board of Education asked staff members to clarify what constitutes appropriate student preparation for the Stanford 9 test. Unfortunately, this clarification will come too late for this year's test, which has already been administered.

Many California school districts administer practice tests for the Stanford 9. However, Harcourt Educational Measurement claims that it does not produce any practice tests and that the only sample questions that have been made public are those reproduced in the appendix of this report.¹ One must ask, therefore, where these practice tests came from and what the differences are — if any— between the sample questions and those on the actual Stanford 9 test.

The same Sacramento Bee article cited above stated that "several districts have purchased other states" versions of the Stanford 9, " and quoted Delaine Eastin, the current superintendent of public instruction, as admitting " I cant say that every district that bought it used it appropriately."

Eastin was also quoted as saying, "In the end, there is a certain amount of honor in all this, and we hope and pray the system works."

What would make the system work would be to vary the questions from year to year. The decision to use the same questions every year, in a state in which they pass through the hands of a quarter of a million teachers, is exactly what would be expected from an education bureaucracy that wants to see test scores rising each year, even if true student achievement does not. That decision to use the same questions year after year was made on Delaine Eastin's watch.

The cheating will get worse, because it is now clear that the California Department of Education is looking the other way, and because, now that the current state budget has been passed, lucrative teacher bonuses for improving test scores will be at stake. Teachers are now eligible for \$5000, \$10,000, or even \$25,000 for raising test scores. All employees in schools in which test scores go up, including even custodians, are eligible for \$1600 bonuses. And \$227 million has been earmarked for bonuses for schools that meet growth targets.¹⁵

EXCLUSION OF LOW-PERFORMING STUDENTS SKEWS THE SCORES

The integrity of any test also depends on whether all eligible students are tested. The Lexington Institute (Arlington, Virginia), for example, has documented that the "improvements" in the National Assessment of Educational Progress in four states — North Carolina, South Carolina, Kentucky, and Connecticut— were likely to be entirely the result of excluding low-performing students from the testing.⁷ For example, Kentucky excluded 10% of its students from the test as learning disabled (LD) in 1998, whereas it excluded only 4% as LD in 1994. Its 6-point percentile "improvement" in reading scores in 1998 compared to 1994 appeared to have been entirely the result of excluding greater numbers of low-performing students.

Here in California, the San Francisco Examiner has documented that virtually all of the "improvement" in test scores claimed by the San Francisco Unified School District was due to excluding ever-greater numbers of students over a six year period.⁸ Most of the remainder of the increases appeared to be a result of changing demographics

as San Francisco became a more expensive place to live and the poorer families were increasingly forced out.

The Contra Costa Times has noted that whereas the state claimed that 95% of eligible Contra Costa County students took the 1999 Stanford 9/STAR test, only 84%-85% of eligible students actually took all of the parts of the test they were supposed to take.⁹ In some schools, such as Castlemont High School in Oakland (Alameda County), fewer than 45% of the students took all of the parts of the test they were supposed to.

The Contra Costa Times article went on to say,

Second-graders statewide took 449,954 Stanford 9 spelling tests, but only 427,720 reading tests. That means at least 22,234 second-graders weren't full participants in STAR.

And while 317,536 of the state's 390,742 high school juniors took the Stanford 9's basic math skills test, those numbers dropped by more than 50,000 when it came time for the more rigorous math exam that is aligned with new state standards. In other words, only about 67% of 11th graders took the standards-based math test.

California law requires those who have spent at least one full year in the public school system and who should therefore be expected to be fluent in English to take the Stanford 9/STAR test in English. But in 1998, then-Superintendent Ruben Zacarias was caught sending letters to Spanish-surnamed parents of students in the Los Angeles Unified School District who had been in California more than one year suggesting that they withdraw their children from the Stanford 9 test.¹⁰

In fact, one district — San Francisco Unified — is illegally excluding all limited English-proficient (LEP) Spanish-speakers from the test, regardless of their tenure in the system. (Another district, Oakland Unified, excluded all LEP students until this year.) Such defiance of the law not only impacts the usefulness of data from SFUSD compared to that of districts in compliance with the law, but, because of the large size of this district, also impacts on the usefulness of the statewide data.

DECLINING ACHIEVEMENT IN HIGHER GRADES

The Stanford 9 test scores in California drop off in the higher grades, as seen in the following table:

Percentage of California Students Scoring Above 75th Percentile in Stanford 9 Compared to the National Average in 1999

	Percentage Scoring Above 75th Percentile	
Subject	2nd Grade	11th Grade
Reading	21%	16%

Math	27%	25%
Language	26%	19%

While it's important to see how students do in the lower grades, of course, what really matters is their preparation once they have completed high school — in other words, the end result. And the sad truth is that, the closer California students approach graduation, the worse they do compared to students in other states.

TEST SCORES ARE NOT ADJUSTED TO REFLECT HIGH DROPOUT RATES

California is in a statistical tie with Texas for the highest dropout rate in the nation among the large-population states, and it is still among the worst even when all states are considered. For the 1998-99 school year, California's graduation rate was a dismal 68.3% — meaning that 31.7% of our students did not graduate, a number that climbs much higher still when adjusted for our very rapid enrollment growth over the last four years.

Students who drop out tend to be among the least academically adept. If the Stanford 9/STAR test were to test not just those who have made it to the later years of high school, but also those who didn't make it, California would look far worse still compared to the national average.

SCHOOLS OF "SIMILAR" SOCIO-ECONOMIC STATUS ARE NOT COMPARED HONESTLY

The results of the Stanford 9/STAR test are supposed to be one part of the California Academic Performance Index (API). However, that "Index" currently consists only of one indicator, the Stanford 9/STAR test. State education officials have stated that other indicators, such as dropout rates, are currently not accurate enough for comparisons of schools, and none are expected to be included in the index for years to come.

The demographic data for use in comparing the results of the Stanford 9/STAR test are obtained by means of a form that is attached to each answer sheet. Last year, approximately 40% of all the forms were not filled in at all. Of those that were filled in, many were completed by the students themselves. Students who were sometimes quite young were asked about their parents' education levels and incomes — data that young children would not be expected to know. Even when the information was supplied by the schools, it is unclear if it was accurate, as most public schools do not normally ask their parents about professional status, educational attainment, or income.

This year, the complete results of the Stanford 9/STAR test were supposed to be released in mid-July. However, about two-thirds of the state's schools again submitted incomplete demographic data on the parents of test-takers. Therefore, only the absolute scores will be available on July 17, while the "similar schools" rankings won't appear until mid-August. ¹³

For purposes of the Academic Performance Index, each school in California is assigned two rankings on a 1-to-10 scale. The first ranking compares schools statewide. The second attempts to compare each school with 100 others serving students of the same socioeconomic background.

After last year's "similar schools" ranking was released, many questioned the validity of these rankings. In February of this year Superintendent Delaine Eastin refused to release data used to compile them, explaining that "its release now would make people angry because the Department of Education might have placed schools in the wrong peer group."¹¹

Even the revised "similar schools" rankings still raises questions about the validity of the demographic data supporting them.

There are also questions about favoritism. Of all of California's districts, the one widely regarded as the worst, regardless of size, is Compton Unified. Taken over by the state in 1993 because of financial irregularities and academic disarray, the district has continued to deteriorate to the point that the California Department of Education refuses to release a timetable for returning the district to local control.

An analysis of Compton's "similar schools" rankings (enclosed) indicates that, while its high schools were appropriately labeled extremely low-performing, its elementary schools were given implausibly high marks. Of Compton's 23 elementary schools, 14, or 61%, reported similar-schools scores of 6 or better — in other words, above average— including 8 schools that scored a "perfect 10."

These implausible rankings were followed in June 2000 with even more implausible dropout rates. After having reported an 18.6% 4-year derived dropout rate for the 1997-98 school year, Compton officials reported a miraculous improvement to a 4-year derived dropout rate of only 7.6% for 1998-99, far lower than the state's (phony) claim of an 11.1% statewide 4-year derived dropout rate. Compton Unified enrolled 1911 students as freshmen in 1995 and graduated 848 seniors in 1999, for a 4-year attrition rate of 55.6%, so even its earlier claim of an 18.6% dropout rate was wildly implausible.

With Compton Unified the only school district now run directly by the California Department of Education, it's hard to believe that CDE officials aren't cooking the books for their own benefit. One has to ask why, if Compton now has a dropout rate even lower than the state average (however phony that figure may be) and test scores better than average for schools with students of similar socioeconomic status, the wishes of the people of Compton are not respected and the school district returned to local control. On the contrary, CDE officials insist that the situation in Compton is so grave that they anticipate continuing to run that district for many years to come. They seem to want to have it both ways: spectacular improvements, but an ongoing crisis.

HIGHER STANDARDS REQUIRED FOR STUDENTS THAN FOR TEACHERS

The sample questions from the Stanford test that are available to the public are reproduced in the appendix. Also enclosed are sample questions from the CBEST (California Basic Educational Skills Test), the only test that teachers have to take to become certified in California. We think that readers will agree that the CBEST questions are easier than even the 10th grade questions from the Stanford 9 test.

The CBEST is once again being challenged as racially discriminatory, since pass rates are higher for whites than for certain minorities. On June 19, a state attorney described the CBEST in court as being "an 8th- to 10th-grade test of the three Rs," the first time an official of the state of California has been willing to give a public assessment of the level of difficulty of this exam.¹⁴

One "spin" that might be put on this paradox is that some of the questions on the Stanford 9 test are quite challenging. In fact, for those who have been out of school for some time, the mathematics part of the Stanford 9 test would indeed represent a challenge. However, it is probably more reasonable to wonder if the dropoff in Stanford 9 test score in the higher grades of high school is due to public school teachers with a weak command of their subject areas.

CONCLUSION

Many school districts have already "pre-released" their Stanford 9/STAR test results, invariably claiming dramatic — and utterly implausible— improvements. It is already clear, before the release of the statewide data, that the Stanford 9/STAR test is fundamentally flawed and that no conclusions can be reasonably drawn from the results— not even the direction of student achievement in California.

What would serve the people of California in attempting to re-establish accountability of our public schools would be a statewide test aligned to the state's curriculum, and with the questions changed from year to year to prevent the kind of wholesale cheating that now is all too common. What we have instead is a test that is exactly what one would expect from a public education establishment that would like to be able to announce "improvements" regardless of the true level of achievement of our children.

The author, Alan Bonsteel, M.D., is a 21-year veteran of the school reform movement in California. He is the co-author of the book *A Choice for Our Children: Curing the Crisis in America's Schools*, and he is the President of California Parents for Educational Choice.

1. The official spokesman for Harcourt Educational Measurement is Bob Rayburn, an educator who can be reached at (916) 326-5372.

2. Personal communication with Les Axelrod, Research and Evaluation Consultant, California Department of Education, (916) 657-5198.
3. Policy Analysis for California Education (PACE) report, May 28, 2000, by Michael Kirst and Elizabeth Burr, quoted in the San Francisco Chronicle, May 26, 2000, by Nanette Asimov.
4. "Teachers in Cheating Probe Face Discipline," Los Angeles Times, January 22, 2000, Martha Groves.
5. "Stressing School Test Scores Brings New Threat: Cheating," Sacramento Bee, January 29, 2000, Marjie Lundstrom.
6. "Fear of Cheating on Statewide Test Grows as Stakes Rise," Sacramento Bee, Pamela Martineau, May 29, 2000.
7. "Can Americans Trust their National Report Card?" School Reform News, March 2000, Robert Holland, Lexington Institute.
8. "SF Test Scores Questioned: Improvement May Be Smoke and Mirrors," San Francisco Examiner, April 12, 1999, Julian Guthrie.
9. "Nagging Problems of STAR," Contra Costa Times, January 9, 2000, Lisa Shafer
10. "Zacarias' Action Shows Panic in LAUSD," Daily News (Los Angeles), March 26, 1998, Alan Bonsteel.
11. "Eastin Refuses to Give Data on 'Similar Schools,'" Contra Costa Times, February 3, 2000, Lisa Shafer.
12. " Personal communication with Bill Lucia, " former Executive Director of the California State School Board, (916) 445-9600.
13. " Glitches to Delay Release of Some State Test Data, " Los Angeles Times, June 22, 2000, Martha Groves.
14. "Teachers" Test Getting Graded by Appeals Court," San Francisco Examiner, June 20, 2000, Bob Egelko.
15. "Millions for Schools Tied to Stanford 9 Test Scores," Los Angeles Times, July 1, 2000, Richard Lee Colvin.