



Clarifying
Complex
Education
Issues

A Primer on STAR Test Scores

More than 4 million California students in grades 2 through 11 take tests in the spring as part of the state's Standardized Testing and Reporting (STAR) program. These scores lead to school-by-school rankings and comparisons, awards for schools that improve, interventions in low performing schools, and financial rewards to teachers in the low performing schools with the greatest improvement. In California, test scores matter.

The STAR program tests basics and standards-based knowledge

California's STAR program has two major components:

1) **The Stanford Achievement Test, Ninth Edition, Form T (Stanford-9)** is a standardized, multiple-choice test produced and scored by Harcourt Educational Measurement. The test is nationally normed; that is, a sample of 500,000 to 600,000 students who reflect the demographic profile of all U.S. students took it in 1995. Their scores are the benchmark against which results from California and other states are compared.

The test topics in grades 2 through 8 are reading, mathematics, language, and spelling, with science and social science substituted for spelling in grades 9 through 11. The tests remain the same each year for a given grade.

2) **The standards-based (augmented) tests** in English/language arts and math are keyed to California's K-12 academic content standards adopted

by the State Board of Education. Both tests include items from the Stanford-9 that reflect state standards plus newly developed questions. In grades 8-10, the math questions are linked to the standards for a particular subject, such as Algebra 1, Geometry, Algebra 2, or Integrated Math 1, 2, 3.

Not all students are tested in the STAR program. A student's Individual Education Program (IEP) through the Special Education program could specify an exemption or special accommodation, such as a test in Braille or with extra time. Further, parents may request, in writing to the school principal, a student's exemption from the test.

In addition to the Stanford-9 and standards-based tests, the state requires that the Spanish Assessment of Basic Education, Second Edition (SABE/2), published by CTB/McGraw Hill, be given to Spanish-speaking students who have been enrolled in a California public school for less than one year. SABE/2 scores are reported for very few students compared to the total school population. These scores cannot be compared to the Stanford-9 scores because the tests are completely different.

Stanford-9 results provide multiple comparisons

The California Department of Education (CDE) reports Stanford-9 scores in three ways, each of which is useful for particular purposes. Parents and teachers receive additional breakdowns for individual students.

1) **National percentile rank (NPR)**. Both individual and group scores are matched against the scores of the national sample of students. The NPR is the percent of students in the national group with scores less than or equal to a student's Stanford-9 score. For example, when a student or group has a reading score at the 60th percentile, it means that 60% of the students in the norming sample scored lower. A national percentile rank is reported to parents and teachers for individual students and is listed publicly each July for groups of 10 or more students by grade, school, district, county, and the state. This reporting method is useful for comparing the scores of individual students to each other and to state or national averages.

2) **Percent at or above the 50th percentile**. This is used to report group performance. It tells what percent of the group performed at or above the national average (50th percentile). The 50th percentile is the median: by definition, half the students in the national group are above and half below that point. California's student performance is compared to the national sample at each grade. Students who score in the top two national quartiles are at or above the

A Short History of Tests in California

A statewide test—the California Assessment Program, or CAP—began in 1972. Each student took only a sample of questions from the overall test. This process, called matrix sampling, does not produce individual student scores, but it does permit an in-depth assessment in each subject. Many districts also gave their students an additional commercial test to assess their individual achievement against a national norm.

In the early 1990s, CAP was replaced by CLAS (California Learning Assessment System), a statewide test for grades 4, 5, 8, and 10. This multiple-choice and written-response test was discontinued in 1995, in part because of controversy over portions of the test.

For the next two years each school district selected its own national tests from a state-approved list to provide individual student scores for grades 2 to 10. In 1997 the Legislature adopted the Standardized Testing and Reporting (STAR) program. It required the selection of a single nationally normed test for all students in grades 2 to 11 that would produce comparable individual, grade, school, district, county, and state scores.

50th percentile. The California Department of Education recommends using this method for year-to-year comparisons. Although it tells the percentage of students in each quartile, it does not indicate the specific level of their performance.

3) **Scaled scores.** The third reporting method converts individual raw scores—a simple count of the number of items answered correctly—to a single scale or scoring system across all grade levels. Unlike the other types of scores, the scaled score has the same meaning in terms of achievement for each grade, making it the best indicator of students' growth over time.

California's standards-based test scores are reported separately

Student scores on the standards-based English/language arts and math tests are reported as the percent of correct answers. Each question has the same weight, with no distinction on the difficulty or the content of the test items. Because the questions were prepared for California students, the results can only be compared with the previous year within the state.

The state is working on how to evaluate students' performance on these questions, with the English/language arts performance standard expected to be ready in 2001. Until then, the results are difficult to assess.

Analyses of STAR results reveal areas of concern about testing

In the first three years of STAR, the results raised several red flags about areas that need further examination and attention.

One big concern is in high school, where the scores drop, especially in reading, compared to earlier grades. In addition, very few students in grades 8 to 10 take the high school math classes for which the standards-based questions were written, and therefore only a small number are tested. Although all 11th graders took the same math test, the results were low. This raises questions about the alignment of the testing system, academic content standards, and the curriculum.

When the test data is disaggregated according to English language ability, family income, or ethnic groups, troubling disparities appear. The gap in scores between English learners (EL) and low-income students and their classmates is large. Among ethnic groups, African Americans and Hispanics have considerably lower scores than whites and Asians.

By happy contrast, test scores of students in the early grades are steadily improving, on the average. Early evaluation indicates some possible effects from nearly universal small classes in those grades. Many EL students at this level are also showing growth comparable to their English-speaking classmates.

In expressing concerns about how test scores are used, educators point out the limitations of multiple-choice questions—as well as many other factors that affect scores and their comparison. These range from classroom conditions during the test, to student attitudes toward and skill at test-taking, to the quality of instruction in the classroom.


Comparing test scores year to year in a school may be misleading. So many California students are mobile that the 4th grade profile at the school could be quite different from the previous year's 3rd graders. Similarly, the demographic makeup of 3rd graders, even in the same district, could vary considerably from one year to the next.

Finally, the Stanford-9 itself comes under fire because of the huge difference between the national test-takers and California's student population. EL students were only 2% of the national group, whereas in California they account for 23% of the test-takers.

The emphasis will be on standards

California's ambitious accountability system is based, for now, on the mandated national Stanford-9 test. The scores drive the allocation of millions of dollars in awards to teachers and other school employees when students improve their scores. This is expected to change over the next several years as more indicators are added to make the comparisons and rankings of schools broader and, presumably, more fair. The standards-based tests will be given greater weight. Meanwhile, school district administrators and teachers are under pressure to examine their practices in pursuit of improved Stanford-9 scores and potential rewards.

The state has more plans for student testing, including a High School Exit Exam and standards-based tests in high school science and social science. The State Board of Education's three-year plan to develop performance standards and to integrate the variety of tests sets 2003 as the date to determine that all content standards are covered. That is also the year the state re-examines its choice of a national norm-referenced test. The goal for the following year, 2004, is to determine whether local curriculum matches the state standards.

California's experience with the use of test scores echoes the experiences in several other states. In Texas, Massachusetts, and Ohio, officials are confronting public criticism of judging schools and students on a single test, which leads to teaching to the test. If California is to avoid some of these controversies, it will have to put together a coherent system of state standards, assessments, and accountability programs with strong enough public and educator support to fulfill their intended purpose. 



EdSource is a not-for-profit 501(c)(3) organization established in California in 1977.

Independent and impartial, EdSource strives to advance the common good by developing and widely distributing trustworthy, useful information that clarifies complex K-12 education issues and promotes thoughtful decisions about California's public school system.

This EdFact may be reprinted, with credit to EdSource.

For general STAR test results by school, district, county, and state, visit the Ed-Data website: www.ed-data.k12.ca.us

For detailed STAR results, see: <http://star.cde.ca.gov>

For more information about our organization, a catalog of our publications, or details on how to subscribe to our Information Service, please contact us at:

4151 Middlefield Rd., Suite 100
Palo Alto, CA 94303-4743

650/857-9604 Fax 650/857-9618
e-mail: edsources@edsources.org
website: www.edsources.org

EdSource thanks the William and Flora Hewlett Foundation for investing in our work.

© Copyright 2000 by EdSource, Inc.