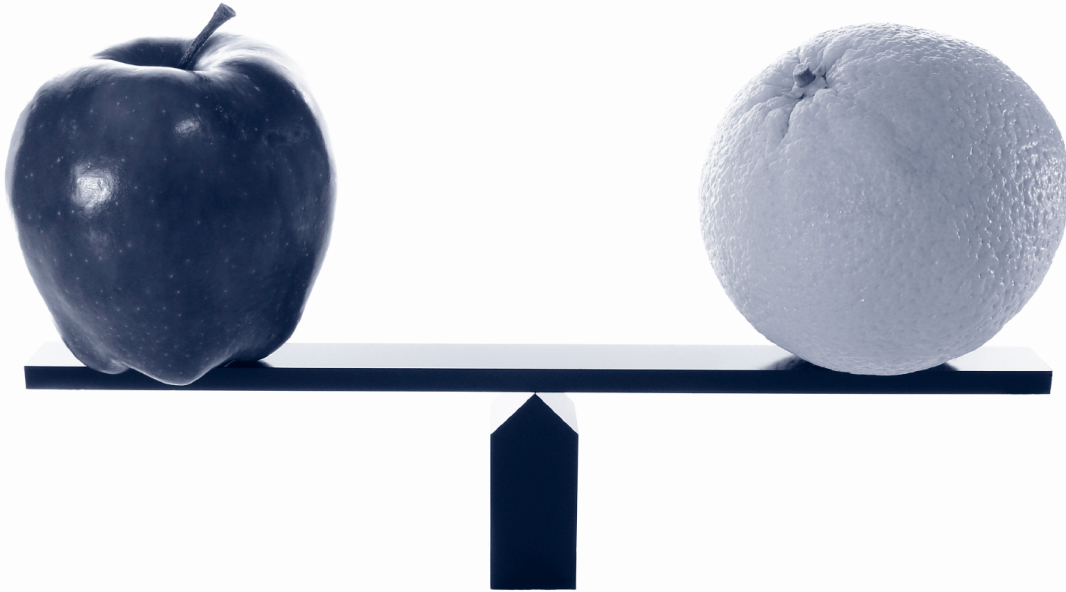


# NAEP and the California Standards Tests: A Case of Apples and Oranges



**POLICYMAKERS, REPORTERS, PARENTS, AND** educators often turn to test results as a convenient, if limited, way to consider questions of overall student achievement and school performance. But for California, different tests seem to offer different answers.

Results from the 2007 administration of the National Assessment of Educational Progress (NAEP) and that year's California Standards Tests (CSTs) appear to tell very different stories about California students' academic performance. Headlines on CST results trumpet improvement in test scores and student achievement. Simultaneously, other reports cite NAEP results to assert that California's schools and students are doing poorly and not improving fast enough, if at all.

The results of the 8th grade NAEP reading test and the most closely related CST (English language arts) illustrate this difference. According to the 2007 NAEP results, only 22% of California's 8th graders scored proficient or advanced in reading.

CST results show that 41% of 8th graders scored proficient or above in English language arts the same year. The 4th grade math tests illustrate a similar trend: 27% of students achieved at the proficient or advanced level on the NAEP math assessment, while 56% of 4th graders achieved at those levels on the CST in math.

So how can the results be so different? The common assumption is that if two tests are labeled 8th grade reading or 4th grade math, they ought to show similar results. However, there are several reasons why this assumption is not necessarily valid. The tests, and thus the results, can be like comparing apples to oranges if they:

1. do not assess the same content.
2. do not assess the same students.
3. do not determine and report success (or proficiency) in the same way.

This report explores the similarities and differences between NAEP and the CSTs and some of the factors that explain how the

two tests can yield different but valuable portraits of student achievement. It also touches on California's performance on the NAEP and clarifies conclusions that may be drawn about California's student achievement based on NAEP results.

## Understanding the fundamentals of the NAEP testing system

The National Assessment of Educational Progress (NAEP) is an ongoing assessment of academic achievement administered to students nationwide. Also called the Nation's Report Card, NAEP provides estimates of student performance in several subjects at the national, state, and large-urban-district levels by testing a scientifically selected representative sample of students from each jurisdiction. NAEP results are calculated to permit comparisons over the course of several years of student performance among states and certain urban districts.

NAEP has been administered in varying forms on a voluntary basis since the 1970s. These national tests have garnered more attention since 2003, when the No Child Left Behind Act (NCLB) began requiring states with districts receiving federal Title I funds to administer the NAEP reading and math tests to their 4th and 8th graders every two years. (The specific schools that participate in the testing are chosen from a stratified random sample.) In addition to the required 4th and 8th grade math and reading tests, California has, since 1992, also voluntarily participated in other NAEP state assessments in subjects such as writing and science and in subjects with only national results, such as U.S. history.

The design of NAEP differs in many fundamental ways from other standardized tests that California students take during the school year, most notably the California Standards Tests (CSTs). Differences in test content and structure, the students who are tested, and the way the tests report achievement are central to understanding what information each set of assessment results can provide.

## Inside This Report

<b>Test Results Raise Critical Questions About How NAEP and the CSTs Differ</b> .....	3
Do NAEP and the CSTs Test the Same Thing? .....	5
Do the Students Tested on NAEP in California Differ from Those Tested in Other States? .....	6
Is “Proficient” on NAEP the Same as “Proficient” on the CSTs? .....	8
<b>Interpreting NAEP Results</b> .....	9
<b>Both Tests—NAEP and the CSTs—Help Illuminate a Complex Subject</b> .....	10

*This report was researched and written by:*  
**Heather Barondess**

*Edited by:*  
**Mary Perry**

*EdSource thanks the **Bill & Melinda Gates Foundation, the William and Flora Hewlett Foundation, and the James Irvine Foundation** for their investment in our core work.*

### Who takes NAEP?

Although states with districts receiving Title I funds must administer NAEP reading and math tests to their 4th and 8th graders, NAEP tests just a small group (known as a sample) of California students who are selected as representatives of the larger student population. The National Center for Education Statistics (NCES) uses demographic data from the Common Core of Data (CCD) to select a stratified sample of schools whose combined student population represents the demographic profile of California’s students. (The statistical process of stratification classifies schools into groups with similar characteristics in order to find a representative sample.) Students at schools selected to be in the sample—including those identified as English learners and students with disabilities—are then randomly chosen for NAEP testing.

According to the Nation’s Report Card, 10,600 students in grade 4 (2.3% of grade 4 enrollment) and 8,900 students in grade 8 (1.8% of grade 8 enrollment) constituted the California sample for reading in 2007. Due to oversampling and the participation of Los Angeles and San Diego unified school districts in the Trial Urban District Assessment, California’s sample is larger than other states’ samples. In 2007, 320 California public schools with students in grade 4 and 310 schools with students in grade 8 participated in the 2007 reading administration. This represented 9% of all the elementary and middle schools in the state. Students in nonpublic schools were also tested, but they are included only in the national-level results.

It is common practice for students with disabilities and English learners to take standardized tests. Testing accommodations on NAEP are provided to students as permitted by NAEP guidelines and the accommodation policies each state has established for its own testing program. Depending on state policies—which differ—and determinations made by officials at the school level, certain students can be excluded from the assessment. Because of its stringent inclusion

guidelines, California tends to include a higher proportion of English learner students in its assessments than most other states. For example, California requires English learners to be assessed in English after one year of enrollment in California schools; other states test English learners in their native language for the first several years. (For more information, see the section on students tested in California on page 6.)

### When and how is NAEP administered and scored?

#### Administration

NAEP assessments in reading and math are administered to 4th and 8th graders between January and March, and initial results are released six months later. Tests in other subjects and those given to 12th graders are administered and scored on a different timeline. A federally funded NAEP administrative team conducts the assessment in each school, which generally involves just one day of testing. NCES collects demographic data the preceding winter in order to inform the selection of schools that will make up the sample.

#### Test Structure

NAEP tests include both multiple-choice and “constructed-response” (open-ended) questions, which are meant to require higher-level thinking. The tests themselves are also split into sections so that no one student takes an entire exam in either math or reading. This type of test structure is called a matrix design: each assessment is divided into multiple sections, and each student takes only two sections, which last a total of 50 minutes. By combining test results from many students, NAEP ends up with results from a much longer test than any individual student could take.

#### Scoring

NAEP tests are scored with both a numerical scale score (from 0 to 500 for 4th and 8th grade reading and math) and corresponding achievement levels of advanced, proficient, and basic. The percentage of

## What exactly is NAEP?

NAEP consists of two corresponding assessments: a “main” test that reports results at the national, state, and/or district levels, and a “long-term trend” national test.

The “**main**” NAEP currently tests students in grades 4, 8, and 12. Assessments are administered every year, with state-level assessments conducted every other year in odd-numbered calendar years. Reading and math are tested every two years. Other subjects, such as science, writing, and disciplines in the social sciences and humanities, are tested less frequently. Results for certain subjects are reported only at the national level. The main NAEP reports its results in several ways:

- **State NAEP:** State-level results are based on scores from a sample of students designed to be representative of the students in that state. Results are reported for public school students only and for certain subgroups of students.
- **National NAEP:** National-level results are based on scores from a nationally representative sample of students, including public and nonpublic students. Results in math and reading reported in odd-numbered years are estimated by combining all of the state samples of public school students and a national sample of nonpublic school students. The results of certain assessments (civics, U.S. history, geography, the arts, and all 12th grade results) are reported at the national level only.
- **Trial Urban District Assessment (TUDA):** District-level results are based on scores from a representative sample of 4th and 8th grade public school students in 11 urban districts. Los Angeles Unified and San Diego Unified are the only California districts that participate in the TUDA.

The “**long-term trend**” NAEP is a separate assessment that tests students ages 9, 13, and 17 in math and reading every four years to track changes in student achievement nationally since the inception of NAEP in 1969. The long-term trend NAEP reports results only at the national level for major demographic groups. The content of this test remains constant in order to produce results that can be compared over longer time intervals. To learn more about the long-term trend NAEP, go to: <http://nces.ed.gov/nationsreportcard/ltr>

*The scope of the discussion in this report is limited to the main NAEP in 4th and 8th grade reading and math. These tests are taken by more students, are more widely referenced than other NAEP assessments, and provide state-level results.*

students scoring below basic is also reported. Achievement levels correspond to scale scores, but the “cut scores”—the score required to reach a given achievement level—differ among tests. They do remain stable over time, however. For example:

- On the grade 4 reading test, *proficient* represents scores between 238 and 267.
- On the grade 8 reading test, *proficient* represents scores between 281 and 323.

Scores on any given test may be compared with the same test over time and among groups of students. (The main NAEP assessments do change slightly every decade to reflect changes in curriculum in the nation’s schools.) Scale scores are

developed independently for every subject tested (reading, math), but then scores from all three grades of testing (4, 8, and 12) are placed together on one reporting scale, known as a cross-grade scale. This allows for comparisons of overall student performance across grade levels within a certain test subject. It is important to keep in mind that tests increase in difficulty with higher grade levels, so different things are measured at different grades even though they are reported on the same cross-grade scale.

### Reporting results

NAEP results are estimates of the average performance of large groups of students and

are considered valid and reliable only when they are reported at state, national, and large-urban-district levels. No results are reported at student and school levels because no one student takes an entire test. In addition, there are not enough students within a school to accurately estimate school-level results. This also serves to keep student and school performance confidential, as required under the NAEP legislation.

Efforts to make the complicated NAEP results “easy to understand” have been criticized for oversimplifying their meaning. The widely distributed Nation’s Report Card omits many of the details of test design, sampling, and scoring in order to provide clarity; but these omissions can lead to incorrect assumptions about NAEP results. (These details are all available, however, on the NAEP website: <http://nces.ed.gov/nationsreportcard>)

This approach has been particularly problematic when policymakers, reporters, or researchers have tried to use NAEP results to establish causal relationships between student achievement and other factors, such as teacher quality or parent involvement. NAEP is not designed for this task, and NAEP results typically contain cautions against doing this. Similarly, comparing student performance among states using NAEP-defined achievement levels instead of scale scores—though much easier to understand—can lead to misunderstandings, especially when the NAEP results are compared with state-defined achievement levels. (This topic is discussed further on page 8.)

### Test results raise critical questions about how NAEP and the CSTs differ

As noted above, the reported results on the California Standards Tests and NAEP appear to tell different stories about student achievement. California is not the only state in which this happens: this trend occurs in many states and has led testing officials and researchers to search for explanations. The Center on Education Policy (CEP), a Washington, D.C.–based nonprofit organization that promotes public education, released a study addressing this issue. Looking at all 50

**figure 1 | The design and administration of NAEP and the California Standards Tests (CSTs) systems differ considerably**

	NAEP	CSTs
<b>Purpose</b>	<ul style="list-style-type: none"> <li>● Provide a national measure of student performance and how performance is changing over time.</li> <li>● Make valid comparisons among states and between a state and the nation.</li> <li>● Provide results for the nation, states, and some large urban districts.</li> </ul>	<ul style="list-style-type: none"> <li>● Measure students' achievement of state academic content standards.</li> <li>● Track progress toward school growth targets for state accountability and toward federal NCLB requirements for adequate yearly progress.</li> <li>● Provide state, county, district, school, and individual student data.</li> </ul>
<b>Overseeing Agency</b>	U.S. Department of Education, with policy direction from the National Assessment Governing Board	California Department of Education, with oversight from the State Board of Education
<b>Who Is Tested</b>	A sample of students in grades 4, 8, and 12 in public and private schools	All public school students in grades 2–11, with few exceptions
<b>Subjects for Which There Are Tests</b>	Math, reading, writing, science, U.S. history, geography, arts, economics, civics, foreign language, and world history. (Frequency of administration varies.)	Math, English language arts, science (general as well as biology, chemistry, physics, and earth science), and history/social studies
<b>When Tests Are Given</b>	Every year between January and March; every two years for state NAEP. (Exact testing window varies by state.)	Every year in the spring when 85% of the instructional year is complete.
<b>When Testing Began</b>	“Long-term trend” NAEP began in 1969; voluntary state participation in “main” NAEP began in 1990.	The CSTs began reporting results in 2002–03.
<b>Achievement Levels</b>	Four levels: advanced, proficient, basic, below basic	Five levels: advanced, proficient, basic, below basic, far below basic
<b>Range of Possible Scores</b>	0 to 500 (4th and 8th grade reading and math)	150 to 600 (all the CSTs)
<b>Required Participation Rate</b>	States must test at least 85% of the schools selected for the sample.	Every school must test 95% of its students to meet federal requirements.
<b>Structure of Test</b>	Multiple-choice and constructed-response matrix-design test. (Each student takes a portion of the exam lasting 50 minutes.)	Multiple-choice only, except 4th and 7th grade writing. Each student takes the complete test, lasting 150 to 195 minutes.
<b>Accountability</b>	State-level results publicly reported. Results also reported for Los Angeles and for San Diego unified districts. No consequences or sanctions for performance.	Contributes to state accountability system (Academic Performance Index). School- and district-level results publicly reported. Can lead to sanctions under NCLB if the district receives federal Title I funds.
<b>Release of Results</b>	National, state, and large-urban-district levels only	State, county, district, and school levels, with results publicly reported. Parents and schools receive student-level reports.

DATA: THE NATION'S REPORT CARD, CALIFORNIA DEPARTMENT OF EDUCATION (CDE)

EdSource 5/08

states between 2003 and 2005, CEP found that many states reporting moderate-to-large gains on state tests showed flat or declining performance on NAEP (including California on grades 4 and 8 reading tests).

A number of researchers have examined some of the factors that independently and

together might explain these observed differences. Some of the factors include the knowledge and skills assessed, the content and format of the test, the demographics of the students who take the tests, and how the tests are scored and the results reported. Their work sheds some light on why

California's NAEP and CST results seem so different.

### *How does NAEP differ from the CST testing system?*

NAEP is one indicator of California's student achievement, and the California

Standards Tests, the state's standardized assessment, are another. The CSTs assess all public school students each year to evaluate their progress toward learning the state performance standards. As Figure I documents in detail, the purpose and design of NAEP and the CSTs differ in many ways. Some of these differences help explain the variations in performance that are reported on these two tests.

For example, although NAEP tests only a sample of students, nearly all students in grades 2–11 take the CSTs. Because all students take an entire CST, interpreting the results is fairly straightforward. However, sophisticated statistical methods are required to yield valid results from the subset of students who take a portion of the NAEP. NAEP's methodology is widely accepted by those in the field of assessment and permits NAEP to measure student performance on a broader range of skills than would be possible with other testing models. However, understanding and interpreting NAEP results can be challenging for even the most skilled researchers, much less for parents, teachers, policymakers, and reporters.

### ***Do NAEP and the CSTs test the same thing?***

It is clear that NAEP and the CSTs differ—to varying extents—in content and structure. Although little systematic research has been conducted to determine how NAEP and state tests compare, experts have stressed that these differences could be important. A 1999 report by the National Research Council, *Uncommon Measures*, which discussed the equivalence of NAEP and state tests, concluded: “Unless the test to be linked to NAEP is very similar to NAEP in content, format, and uses, the resulting linkage is likely to be unstable and potentially misleading.”

### ***NAEP does not systematically align with the academic content tested on the CSTs***

The guiding frameworks that inform the content of NAEP and the CSTs have distinct origins and purposes, which may lead to differences in the content that the

### **Student motivation seems to have little effect on NAEP results**

Researchers hypothesize that a student's motivation to perform well on a test—as well as teachers' motivation to prepare their students—can vary depending on the stakes. For example, students might try harder or take a test more seriously if their performance will affect their grade in a course or if their parents will see the test results. The CSTs are high-stakes tests because the scores are reported to parents and schools and are tied to the state and federal accountability systems. In contrast, NAEP is a no-stakes test, with no consequences related to either a student's or a school's performance. Individuals and schools who participate in NAEP testing are completely anonymous. Although the hypothesis about student motivation may seem intuitively correct, existing research on this issue has not found motivation to have much effect on a student's NAEP performance, even for middle and high school students. (For an example of such research, see the study by Vonda L. Kiplinger and Robert L. Linn at: [http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/13/8e/ed.pdf](http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/13/8e/ed.pdf))

tests cover. The CSTs are based on California's academic content standards, which were developed through an extensive consultative process within the state of California. The initial purpose of California's content standards was to serve as a guide for curriculum and instruction. Their use in determining the content and structure of the CSTs took place later.

In contrast, the NAEP frameworks have only one purpose, which is to guide the design of the assessments. These frameworks were developed through a national consensus process involving the review of states' curriculum frameworks and standards and consultation with state officials. The NAEP frameworks are not themselves a set of academic content standards, but rather they outline content that national experts in a particular subject area consider to be important.

Little research has been done to systematically compare California's content standards with the NAEP frameworks. However, on the face of it, they do not appear to be consistently aligned with each other. The case of 8th grade math provides a clear example. In 2007, about half of California's students took Algebra I in 8th grade, while the other half took General Math. In effect, there are two different math curricula—and two different math CSTs—for California 8th graders, each of which is taken by about half the students. Meanwhile, NAEP administers a single assessment to 8th graders in math.

Although the test contains algebraic principles, it is not an algebra test.

The reading/English language arts tests also do not appear to align with one another. The 4th and 8th grade CSTs for this subject assess all aspects of English language arts, including a writing sample in the 4th grade. In contrast, the content of the comparable NAEP assessments focus exclusively on reading comprehension. Writing is assessed on a different NAEP test. (It is interesting to note that out of all the NAEP assessments, most states' content standards for 4th grade reading are the ones most closely aligned with the skills tested on a NAEP test, according to a 2007 study from the Center on Education Policy.)

### ***NAEP and the CSTs are structured differently***

Even if the content being assessed is similar for both NAEP and the CSTs, the structure of the tests can vary in important ways. If a certain content area is weighted more heavily on one test than the other or if the format of the tests is not similar, student performance can be affected. For example, Brian A. Jacob's 2007 study of Texas' standards found that the difference in 4th grade math scores on the state test and NAEP was largely explained by the dissimilar math skills examined and the inconsistent presentation/formats of the two tests.

The NAEP Validity Studies (NVS) Panel (a group charged with providing a technical review of NAEP plans and

**figure 2 | The structure and format of NAEP and the CSTs differ in notable ways**

Allocation of Test Questions by Content Area, 4th Grade Math, 2007					
NAEP Content Category	Number of Questions	% of Questions	CST Content Category	Number of Questions	% of Questions
Number Properties and Operations	65	40%	Number Sense	31	48%
Measurement	35	20%	Measurement and Geometry	12	18%
Geometry	26	15%			
Data Analysis and Probability	20	10%	Statistics, Data Analysis, and Probability	4	6%
Algebra	20	15%	Algebra and Functions	18	28%
Reasoning and Process Skills	n/a	Embedded in other questions	Mathematical Reasoning	n/a	Embedded in other questions
<b>Total</b>	<b>166</b>	<b>100%</b>	<b>Total</b>	<b>65</b>	<b>100%</b>

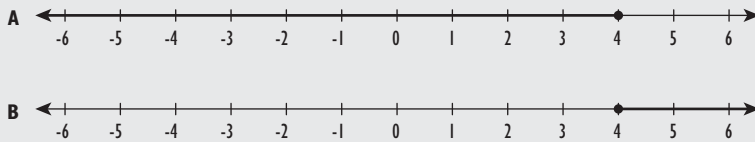
Note: The content analysis framework used in this figure came from "Linking a Statewide Assessment to the 2003 National Assessment of Educational Progress (NAEP) for 4th and 8th Grade Mathematics," Delaware Department of Education.

DATA: NAEP VALIDITY STUDIES (NVS) PANEL  
CALIFORNIA DEPARTMENT OF EDUCATION (CDE)

EdSource 5/08

### Multiple-choice Compared with Constructed-response Questions

A multiple-choice question might ask students to identify which of the two line graphs below correctly expresses the equation  $x \geq 4$ .



A constructed-response question would not offer these choices, instead asking the student to draw a graph that correctly expresses the equation  $x \geq 4$ .

products and with identifying technical concerns) reviewed the content of NAEP and of the CSTs for 4th grade math and found differences. It compared the NAEP *Mathematics Frameworks* to California's blueprints for its state tests and to those of five other states and Singapore. Drawing from that comparison and other studies, Figure 2 shows how the content areas on the NAEP differ from what is on the CST for 4th grade math.

The format and presentation of the two tests are also different. The 4th grade NAEP in math consists of 46 short constructed-response questions, six extended constructed-response questions, and 114 multiple-choice questions. In contrast, the CST in grade 4 math is composed of 66 multiple-choice questions. The dissimilarity in structure and format of the CST and NAEP for 4th grade math might explain some of the difference in student performance.

### Do the students tested on NAEP in California differ from those tested in other states?

The 2007 NAEP results report that California students, as a whole, are generally performing worse on most NAEP assessments than students in the majority of other states. It also appears that the demographics of the students who take the test in California compared to other states contribute to California's poor overall rankings on the national scale. Evaluating and comparing student achievement among subgroups, rather than for students overall, can help control for the effects of the state's unique demographics. It can also help to identify gaps in achievement among different subgroups within the states.

### Based on student demographics, California stands apart

California has the highest proportion of English learners of any state in the nation. One-third of 4th graders and 22% of 8th graders in the state were identified as English learners in 2007. Under NCLB, states set their own policies for how students will be classified as English learners. In California, English learner students are defined as those whose home or primary language is not English and whose district has not reclassified them as "fluent English proficient" based on state test scores and other criteria.

Other states, however, may have different policies for identifying English learners. Florida identified only 9% and Texas only 16% of their students as English learners. These lower percentages may reflect the demographics of a state's students, but they may also be affected somewhat by the state's policies used to classify students as English learners and to redesignate them as fluent English proficient.

### A large proportion of California test takers are English learners

States also have the autonomy to decide who is excluded from state tests, who may be eligible for testing accommodations, and what accommodations are used, in accordance with NAEP testing regulations. These policies generally apply to students with

**figure 3 | More English learners take NAEP in California than in other states**

Participation of English Learners (ELs) on 2007 NAEP, Grade 4 Reading			
	Identified as EL	Excluded Due to EL Status	% of EL Students Excluded
Jurisdiction	% of All Students Sampled		
Nation	11%	2%	18%
California	33%	2%	6%*
Florida	9%	4%	44%
New York	9%	2%	22%
Texas	16%	5%	31%

\* In California, 33% of students are designated as ELs. The state excludes 2% of 4th grade students from the NAEP reading assessment due to their EL status. Because 2% of 33% is 6%, that means that 6% of California's EL population is excluded.

DATA: NATIONAL CENTER FOR EDUCATION STATISTICS (NCES)

EdSource 5/08

### Ranking issues: What comparisons among states are valid?

Some NAEP reporting ranks states based on their performance. When that is done, California ranks near the bottom on almost every NAEP test: 49th (out of 52 total jurisdictions) on 4th grade reading and 48th on 4th grade math. However, when comparing NAEP performance among states, results are not as clear as they may seem.

Ranking states can result in a mischaracterization of how one state is performing relative to another. Instead, testing officials recommend looking at statistical differences in average scale scores to compare achievement among states. Although it does not drastically affect the outcomes of overall state-by-state comparisons, this method is more precise because it accounts for sampling and measurement error. In the case of 8th grade NAEP in math, for example, results show that, statistically:

- California has a higher scale score than four other jurisdictions (Alabama, Washington D.C., Mississippi, and New Mexico).
- California's scale score is not different from the scores of another four jurisdictions (Hawaii, Louisiana, Nevada, and West Virginia).
- California's scale score is lower than 44 other jurisdictions.

*NAEP's Data Explorer tool has a button to check whether differences in scores are statistically significant, meaning that a result is within a margin of error that would be expected when you test a sample of students instead of the whole population. For more information, go to:*

**<http://nces.ed.gov/nationsreportcard/nde/criteria.asp>**

profound disabilities and English learners with very limited English proficiency. Sometimes NAEP regulations and state policy regarding specific accommodations may differ. For example, many states permit tests to be read aloud or translated, as needed. NAEP does not allow this accommodation so many students who take tests under these conditions would not take NAEP. California,

however, does not allow that particular accommodation, so the state's English learners have higher rates of participation on NAEP. In California, 2% of sampled English learners were eligible for testing accommodations on the grade 4 reading test, and 3% were accommodated on grade 4 math.

Excluded students are those who are not able to meaningfully participate in the

assessment, even with state-approved testing accommodations. In California, 2% of 4th grade students were excluded from the NAEP reading assessment because of their English learner status. And 1% of English learners were excluded from the math assessment.

Currently, reported NAEP results do not control for the variation in states' exclusion policies.<sup>1</sup> California's policy requires almost every student to be tested. Generally, the only students excluded from taking the exams are those whose parents have opted to not have their child take the test. Only 2% of California's sampled English learners were excluded from the 4th grade reading test compared with 5% of Texas' sampled English learners.

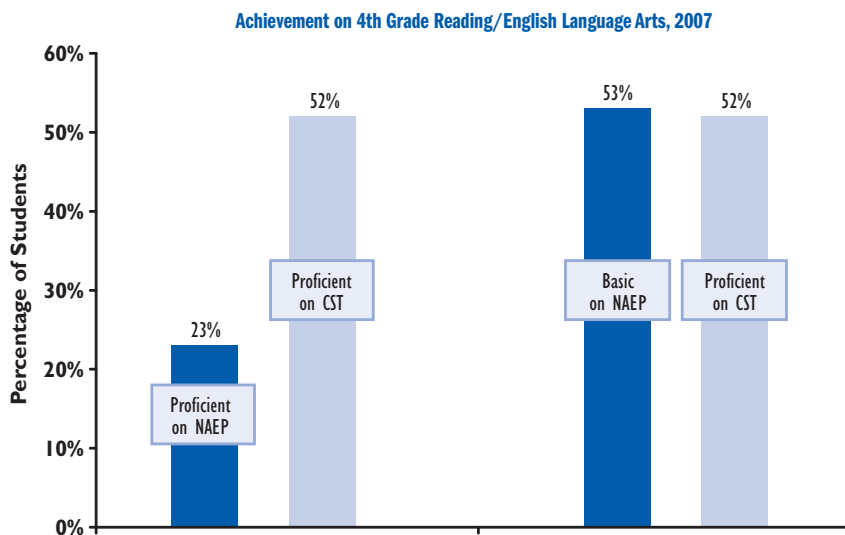
As Figure 3 illustrates, although California's overall rate of exclusion for English learners was no different from the national average (2% of all students tested), far more English learners are assessed in California than elsewhere. In other states, fewer students were identified as English learners for NAEP, so a greater percentage of those students were excluded from the assessment than in California.

These accommodation and exclusion issues have particular relevance for California. Given that more English learners are being tested, California's overall mean score could logically be expected to be lower than that of other states.<sup>2</sup> The data also show that there is a significant difference in the performance of California's English learners on NAEP compared with the national average. However, California's English learners cannot explain the state's poor performance on NAEP completely. Non-English learner students in California also score significantly lower than the national average for non-English learners on every NAEP assessment.

### Analyzing subgroup achievement yields more accurate comparisons

For the purpose of state-to-state rankings, comparing subgroups of students, rather than overall performance, leads to more meaningful interpretations. These comparisons begin to control for variables such as demographics and exclusion policies that can affect the accuracy when overall student achievement is compared.

**figure 4** | Some experts say that the CSTs' "proficient" achievement level may be more comparable to "basic" on NAEP



DATA: NATIONAL CENTER FOR EDUCATION STATISTICS (NCES)  
CALIFORNIA DEPARTMENT OF EDUCATION (CDE)

EdSource 5/08

The achievement of a few student subgroups in California is closer to the national average than the overall student population. On the 4th grade reading and math NAEP tests, for example, California's results for white students were not statistically different from the national average for white students. The same is true for Asian/Pacific Islander students on 4th and 8th grade reading and 4th grade math. African American students' achievement in California does not differ significantly from the national average for African American students on 4th grade reading.

These findings suggest that some of the overall difference between California's average scores and the national average are a reflection of demographic differences between California and the nation as a whole. However, subgroup-level results for every other 4th and 8th grade math and reading test not listed above show significantly lower performance than the national average. For example, California's white students scored below the national average for white students on the grade 8 reading and math tests.

### Is "proficient" on NAEP the same as "proficient" on the CSTs?

Education stakeholders, policymakers, and the media often focus on achievement levels rather than scale scores when interpreting student achievement on both NAEP and CSTs. Reporting that "50% of students scored at the proficient level" tends to be easier to understand than such metrics as "an average scale score of 219." However, looking more closely at the assumptions that underlie those achievement levels can help explain how it is possible for 52% of California's 4th graders to score at or above proficient on the English language arts CST while only 23% score at or above proficient on the NAEP reading assessment.

### Methods to set cut scores differ on NAEP and the CSTs

The way that officials establish the passing score on a test (defined as proficient on both the NAEP and the CSTs) determines how hard it will be for students to succeed. On NAEP and the CSTs, a student's score corresponds to an achievement level of basic, proficient, and so on. The cut score (the minimum score that a student must earn in

order to pass a test or earn a specific achievement level such as proficient) varies for each test administered. To draw a comparison, a teacher may set the passing score at 80% for one test and at 60% for another. Or a teacher may set the passing score at 70% for both an easy test and a hard test. Differences in how many students pass these two tests can be explained, in part, by the variation in how passing is defined.

The cut scores for NAEP and the CSTs were developed independently of one another, using different methods. Federal officials set NAEP's achievement levels for the 4th and 8th grade reading and math tests using the Angoff method. In summary, the Angoff method uses a panel of judges who estimate the probability of a student answering a question correctly if the student performs *just barely* at the basic, proficient, or advanced achievement levels. The panelists' item-level judgments are then combined to determine a cutoff point defining each achievement level.

To set CST cut scores, California used the Bookmark method in which panelists are shown the test items ordered from easiest to hardest. Panelists are asked to "bookmark" the locations on the continuum where a *just barely* basic, proficient, or advanced student should and should not be able to answer. (For more information on how achievement levels are set, see Andrew J. Rotherham's 2006 report, *Making the Cut: How States Set Passing Scores on Standardized Tests*.)

### The tests define achievement levels differently

Although the difference in how performance levels are set may seem technical, a great deal of controversy surrounds this issue. These details determine the rigor of state tests relative to NAEP. In California, where the percentage of students scoring proficient or above on the CSTs is higher than on NAEP, some critics say that the state's test is too easy or the cut point for proficiency is set too low. The National Center for Research, Evaluation, Standards, and Student Testing (CRESST) has stated that "California's achievement levels are somewhat more lenient than NAEP."



## How standardized tests are scored

Scoring methods for standardized tests are not as simple as the number of right answers. The following example of how the state scores its California Standards Tests (CSTs) illustrates this point and defines some common terms.

**1) Raw score:** Raw scores identify the number of questions answered correctly on a test or subtest. Because half of the questions change from year to year, raw scores on the CSTs should not be compared.

In 2007, the mean raw score on the 8th grade English language arts CST was 43.96 (out of 75 test items).

**2) Scale score:** A formula is used to convert raw scores into scale scores ranging from 150 to 600, taking into account differences in the difficulty of test items. New raw-score to scale-score conversions are calculated annually so that scale scores have the same meaning year to year, within each grade and subject area.

In 2007, the mean scale score on the 8th grade English language arts CST was 339. This did not change from the 2006 assessment.

**3) Performance levels:** Performance levels are established using a judgmental process so that for each subject, the percentages of students in each category change smoothly from grade to grade. The cut scores for “proficient” and “basic” stay the same every year; “advanced,” “below basic,” and “far below basic” may change slightly from year to year.

The cut point for basic is 300 for all CSTs. The cut point for proficient is 350. Cut points for other achievement levels may vary slightly from year to year.

Determining cut scores is as much art as it is science because the process depends on the subjective judgment of a panel of judges. Even if the same standard-setting methods were used for NAEP and CSTs, results may not be the same for the two tests because the definitions for the achievement levels are different:

- For NAEP, panelists were asked to use a standard that defines proficient in part as representing “solid academic performance at each grade assessed. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real-world situations, and analytical skills appropriate to the subject matter.”
- For the CSTs, California’s State Board of Education defines proficient as “a competent and adequate understanding

of the knowledge and skills measured by this assessment.”

Another explanation for fewer California students reaching proficient or above on NAEP is that the cut scores for achievement levels on NAEP are set relatively high, a concern of the National Academy of Sciences (NAS), which has evaluated NAEP. Testing officials have characterized NAEP achievement levels as “aspirational,” setting a very high standard for proficient that is substantially above students’ current performance. Critics say that NAEP results would be easier for the public to understand if the achievement level standards were not so far ahead of observed performance.<sup>3</sup>

Some researchers, along with U.S. Secretary of Education Margaret Spellings, have said that the proficient achievement level, as defined by many states, may be more comparable to the basic achievement level on

NAEP. California’s 2007 results from the 4th grade NAEP in reading and the California Standards Test in English provide an example. When student achievement is compared using the measure of percent of students proficient or above on both NAEP and the CST, the performance gap is dramatic. But when performance is compared using basic or above on NAEP and proficient or above on the CST, the performance gap disappears. (See Figure 4.)

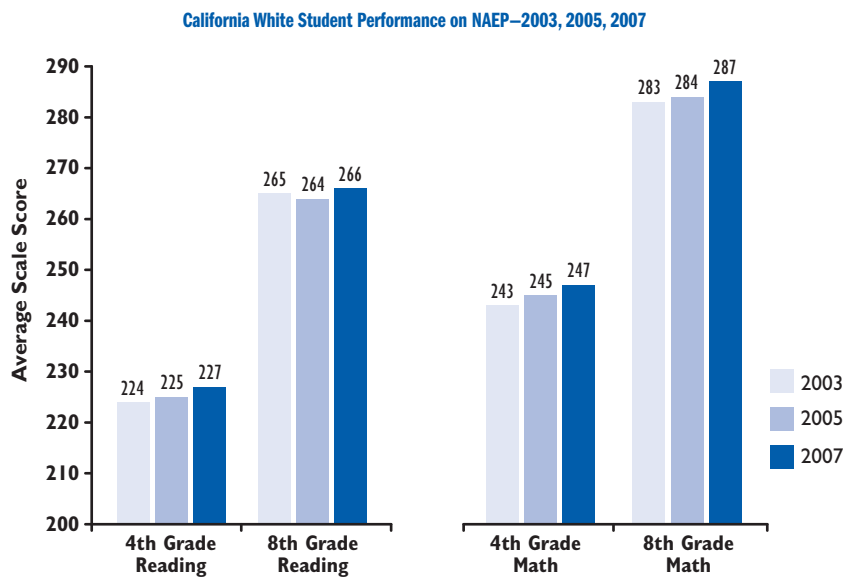
Achievement levels for the main NAEP were set in the early 1990s as a way to judge student performance against a standard (proficient) instead of simply reporting a score, which had been NAEP’s approach with the long-term trend assessment for 20 years. This shift was controversial, and the achievement levels are still being used on a trial basis until the NCEC commissioner, based on an independent evaluation, deems them to be “reasonable, valid, and informative to the public.”

Evaluation studies of NAEP’s achievement levels have been mixed, with some studies lending support to NAEP’s definitions and others raising concerns about what basic, proficient, and advanced really mean. The 1998 NAS evaluation found that “NAEP’s current achievement level setting procedures remain fundamentally flawed.” Questions about the way achievement levels are defined have also prompted NCEC to issue cautions about the use and interpretation of NAEP student performance data using achievement levels. Stanford University Professor of Education Edward Haertel notes that the bulk of the NAS criticisms, though aimed at NAEP, would also apply to the standard-setting methods used for the CSTs and the high-stakes assessments in other states.

## Interpreting NAEP results

Although NAEP is considered by many to be a very strong assessment, it is also nuanced; and interpreting the results can be complex. The challenges involved in interpreting NAEP data have led to misunderstandings of what NAEP results reveal about student achievement in the United States.

**figure 5** California's white students have made gains on each of the four key NAEP tests since 2003



DATA: NATIONAL CENTER FOR EDUCATION STATISTICS (NCES)

EdSource 5/08

### Points to keep in mind

Because of NAEP's sophisticated test design, NCES and the California Department of Education urge caution when interpreting NAEP results. Researchers, parents, reporters, state policymakers, and other parties interested in interpreting NAEP data should keep the following "rules" in mind:

- NAEP scores should not be compared across test subjects or across test grades, but they may be compared over time or between states.
- Average scale scores represent how students performed on a test. Achievement-level results indicate how that performance measured up against set expectations for achievement.
- NAEP assesses student performance based on the knowledge and skills included in the NAEP frameworks.
- NAEP results are estimates based on a representative sample of students.
- California's student demographics have an effect on the state's overall performance on NAEP. For greater accuracy, it is important to compare subgroup results with those of similar students in other states and the nation.

- Observed differences among years, subgroups, or states may not be statistically significant.
- NAEP and the CSTs are not designed to be comparable tests. They are apples and oranges, often assessing different content and skills using different test formats.
- NAEP and the CSTs use different methods to determine cut scores.
- Achievement levels for the two tests are defined differently, as are the cut points required to reach them.

### Performance results shed light on California's academic progress and future challenges

With these guidelines for interpreting NAEP results in mind, the achievement of a subgroup—white students—helps shed light on how California's students perform on this national assessment.

California's white students have made gains on each of the four key tests (grades 4 and 8 math and reading) since 2003. (See Figure 5.) Certain gains have been more noteworthy than others. Both 4th and 8th grade math have seen the greatest and most

consistent improvements. Less improvement has been made in reading in grade 4. White students' performance in 8th grade reading has been almost flat since 2003, worsening in 2005, and then improving in 2007.

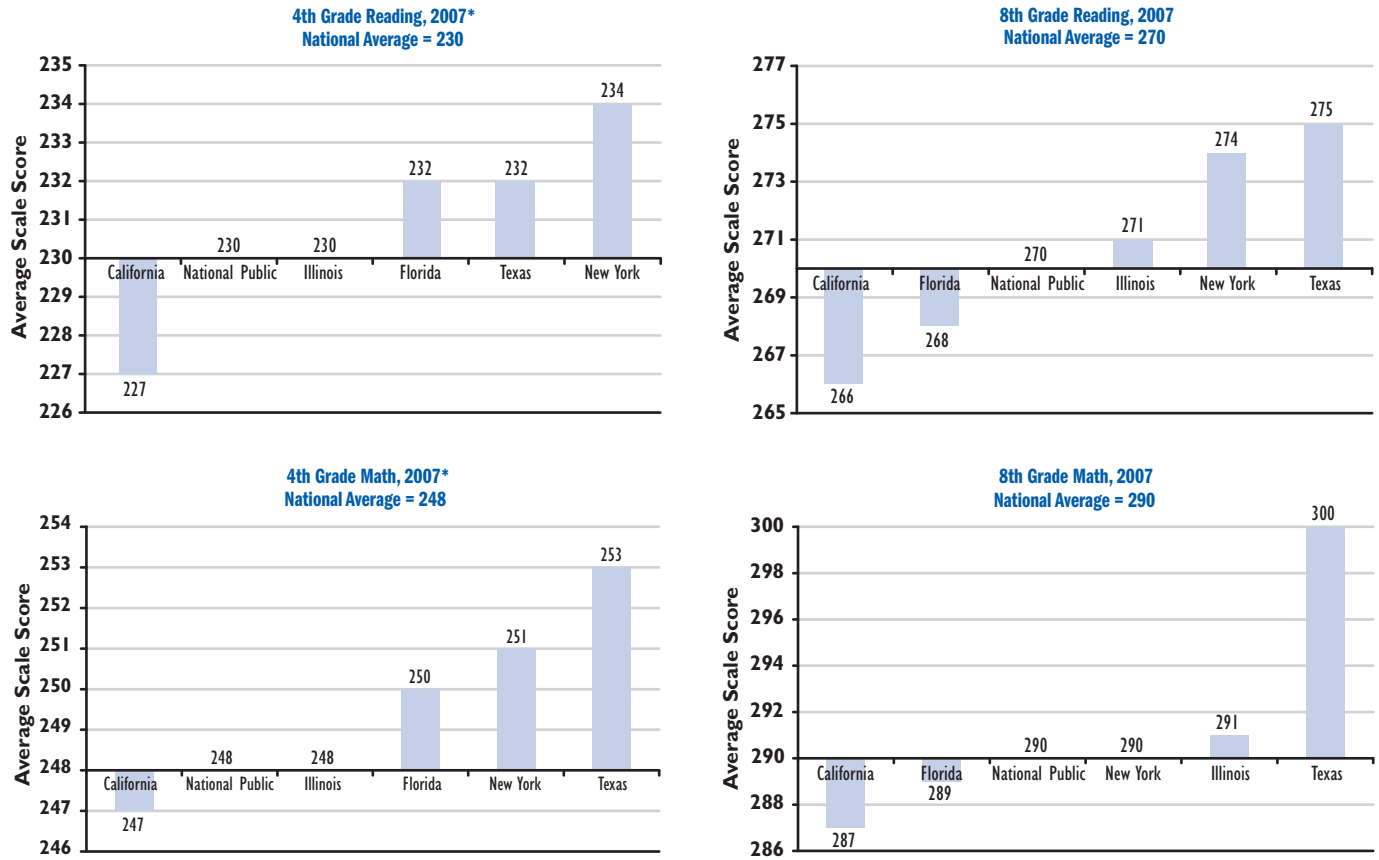
Compared to the four other largest states and to the nation as a whole, California's white students continue to struggle on the 8th grade NAEP assessments in reading and math. Even when controlling for California's demographic differences, the state's white 8th graders score significantly lower than the national average for white students. In contrast, however, California's results for white students are not significantly different on 4th grade reading and 4th grade math than the national average for this subgroup. (See Figure 6.)

### Both tests—NAEP and the CSTs—help illuminate a complex subject

Many factors contribute to how the story of student achievement is told in California. Differences in the way NAEP and the California Standards Tests are designed and administered can affect results, as can the process by which cut scores and achievement levels are envisioned and set. The content and format of each test differs, and further study is necessary to determine the extent to which the NAEP frameworks are aligned to California's academic content standards. The demographics of California's students also clearly affect the state's overall performance on NAEP, making comparisons between comparable subgroups more apples to apples. Comparing student achievement among states relies on using the correct measures and being cognizant of the sophisticated statistical factors that may be at play.

When policymakers and advocates cite California's test results just to prove a point—NAEP results to show how students are failing in the state, CST results to show how students are improving—it can divert attention from efforts to improve public schools in California. Student performance on these two separate assessment systems can answer different questions, all of which are important.

figure 6 California's white students scored at about the national average on 4th grade tests in 2007, while 8th graders scored below the national average



\*No statistically significant difference between California's performance and the national average.

DATA: NATIONAL CENTER FOR EDUCATION STATISTICS (NCES)

EdSource 5/08

### Using the data wisely

Some politicians, policymakers, and reporters have used NAEP results as proof of the effectiveness of No Child Left Behind (NCLB) on improving students' academic achievement. However, the National Assessment Governing Board warns that NAEP should not be used to prove cause-and-effect relationships between any factor and student performance on NAEP. Because NAEP is not an experimental or longitudinal study, its data cannot be used to make claims such as "NCLB is effective because 4th grade reading scores have been on the rise." NAEP results are strictly descriptive. For example, although students enrolled in private schools perform better on NAEP than students enrolled in public schools, it cannot be inferred that enrollment in the private school is what caused students to score better on NAEP.

Instead, NAEP was designed to provide data and information to help inform inquiry, policymaking, and public discussion on such topics. It provides a snapshot of student performance as well as trend data that can illuminate important questions that warrant more in-depth research.

The CSTs are closely aligned with California's academic content standards, the material state leaders deem to be most important. As a result, they provide rich

information regarding how well students, schools, districts, and the state are mastering California's content standards.

NAEP provides California with evidence of how well the state's students are learning compared to other states and the nation as a whole. Test content is based on a national consensus regarding what students need to learn and thus provides a valuable comparison. California's performance on NAEP surfaces policy issues for state leaders to consider related to the state's strategies for instruction, assessment, and accountability.

Improving the academic performance of the state's 6 million students is not just a matter of raising test scores, but NAEP and the CSTs each provide important evidence of the progress and success of California's education system. [11](#)

## NAEP Resources

California Department of Education. [www.cde.ca.gov/ta/tg/nr](http://www.cde.ca.gov/ta/tg/nr)

Frequently Asked Questions About NAEP. [www.nces.ed.gov/nationsreportcard/faq.asp](http://www.nces.ed.gov/nationsreportcard/faq.asp)

National Assessment of Educational Progress (NAEP), National Center for Education Statistics (NCES), U.S. Department of Education, Institute of Education Sciences. [www.nationsreportcard.gov/reading\\_2007](http://www.nationsreportcard.gov/reading_2007) and [www.nationsreportcard.gov/math\\_2007](http://www.nationsreportcard.gov/math_2007)

No Child Left Behind Act (NCLB), Public Law 107-110, 115 Stat. 1425 (2002). [www.ed.gov/policy/elsec/leg/esea02/107-110.pdf](http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf)

## Works Cited

*Answering the Question That Matters Most: Has Student Achievement Increased Since No Child Left Behind?* Center on Education Policy (2007). [www.cep-dc.org](http://www.cep-dc.org)

Daro, P., et al. *Validity Study of the NAEP Mathematics Assessment: Grades 4 and 8*. Conducted by the NAEP Validity Studies Panel (September 2007). [www.air.org/publications/documents/NAEP\\_Math\\_Validity\\_Study.pdf](http://www.air.org/publications/documents/NAEP_Math_Validity_Study.pdf)

Feuer, M.J., et al., Eds. *Uncommon Measures: Equivalence and Linkage Among Educational Tests*. National Research Council (1999). [www.nap.edu/catalog.php?record\\_id=6332](http://www.nap.edu/catalog.php?record_id=6332)

Jacob, B.A. *Test-Based Accountability and Student Achievement: An Investigation of Differential Performance on NAEP and State Assessments*. NBER Working Paper (2007). [www.nber.org/papers/w12817](http://www.nber.org/papers/w12817)

Kiplinger, V.L., and Linn, R.L. *Raising the Stakes of Test Administration: The Impact on Student Performance on the National Assessment of Educational Progress*. *Educational Assessment*, 3(2), 111-133 (1995/1996).

[http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/13/8e/ed.pdf](http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/13/8e/ed.pdf)

Linn, R.L. *Fixing the NCLB Accountability System*. CRESST Policy Brief 8 (2005). [www.cse.ucla.edu/products/policy/cresst\\_policy8.pdf](http://www.cse.ucla.edu/products/policy/cresst_policy8.pdf)

McLaughlin, D.H. *Properties of NAEP Full Population Estimates* (Report to NCES). Palo Alto, CA: American Institutes for Research (2005).

*National Assessment of Educational Progress & California Achievement Test*. Curriculum, Instruction and Evaluation Council, Association of California School Administrators (ACSA) (2004). [www.acsa.org/doc\\_files/NAEPboardpaper%5B1%5D.pdf](http://www.acsa.org/doc_files/NAEPboardpaper%5B1%5D.pdf)

O'Connell, J. *State Schools Chief Jack O'Connell Comments on Release of 2007 National Assessment of Educational Progress*. California Department of Education news release (September 2007). [www.cde.ca.gov/nr/ne/yr07/yr07rel122.asp](http://www.cde.ca.gov/nr/ne/yr07/yr07rel122.asp)

Pellegrino, J.W.; Jones, L.R.; and Mitchell, K.J., Eds. *Grading the Nation's Report Card: Evaluating NAEP and Transforming the Assessment of Educational Progress*. Committee on the Evaluation of National Assessments of Educational Progress, Board on Testing and Assessment, Commission on Behavioral and Social Sciences and Education, National Research Council. Washington, D.C.: National Academy Press (1998). [www.nap.edu/catalog.php?record\\_id=6296](http://www.nap.edu/catalog.php?record_id=6296)

Pellegrino, J.W. *Should NAEP Performance Standards Be Used for Setting Standards for State Assessment?* Phi Delta Kappan (March 2007). [www.pdkintl.org/kappan/kappan.htm](http://www.pdkintl.org/kappan/kappan.htm)

Rotherham, A. *Making the Cut: How States Set Passing Scores on Standardized Tests*. Education Sector (2007). [www.educationsector.org/usr\\_doc/EXPCutScores.pdf](http://www.educationsector.org/usr_doc/EXPCutScores.pdf)

*Using the National Assessment of Educational Progress To Confirm State Test Results: A Report of the Ad Hoc Committee on Confirming Test Results*. National Assessment Governing Board, U.S. Department of Education (2002).

[www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/1a/70/76.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1a/70/76.pdf)

Wei, S.; Lukoff, H.; and Haertel, E. *Using Test Content to Address Trend Discrepancies between NAEP and California State Tests*. Presented at AERA, San Francisco (2006). [www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/29/00/1f.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/29/00/1f.pdf)

Yeager, M. *Understanding NAEP: Inside the Nation's Education Report Card*. Education Sector (2007). [www.educationsector.org/analysis/analysis\\_show.htm?doc\\_id=560606](http://www.educationsector.org/analysis/analysis_show.htm?doc_id=560606)

## Endnotes

1 McLaughlin, D.H. *Properties of NAEP Full Population Estimates* (Report to NCES). Palo Alto, CA: American Institutes for Research (2005).

2 Wei, X.; Shen, X.; Lukoff, B.; Ho, A.D.; and Haertel, E.H. *Using Test Content to Address Trend Discrepancies Between NAEP and California State Tests*. Presented at an American Educational Research Association meeting, San Francisco (2006). [www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content\\_storage\\_01/0000019b/80/29/00/1f.pdf](http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/29/00/1f.pdf)

3 Pellegrino, J.W. *Should NAEP Performance Standards Be Used for Setting Standards for State Assessment?* Phi Delta Kappan (March 2007). [www.pdkintl.org/kappan/kappan.htm](http://www.pdkintl.org/kappan/kappan.htm)



**Trish Williams**

*EdSource Executive Director*

## 2007-08 EdSource Board of Directors

**Davis Campbell, President**

*President, California School Boards Association Governance Institute*

**Lawrence O. Picus, Vice President**

*Professor, Rossier School of Education, University of Southern California*

**Martha Kanter, Fiscal Officer**

*Chancellor, Foothill-De Anza Community College District*

**John B. Mockler, Secretary**

*President, John Mockler & Associates, Inc.*

**Susan K. Burr**

*Executive Director, California County Superintendents Educational Services Association*

**Carl A. Cohn**

*Distinguished Leader in Residence, College of Education, San Diego State University*

**Christopher Cross**

*Chair and CEO, Cross & Joftus, LLC*

**Kenneth F. Hall**

*Executive in Residence, University of Southern California*

**Gerald C. Hayward**

*Co-director (Retired), Policy Analysis for California Education*

**Janis R. Hirohama**

*President, League of Women Voters of California*

**Santiago Jackson**

*Director, Office of Legislative and Governmental Affairs, Los Angeles Unified School District*

**Kelvin K. Lee**

*Superintendent (Retired), Dry Creek Joint Elementary School District*

**Donna Lilly**

*Co-president, American Association of University Women-California*

**Jo A.S. Loss**

*President-elect, California State PTA*

**Paul J. Markowitz**

*Teacher, Las Virgenes Unified School District*

**Amado M. Padilla**

*Professor, School of Education, Stanford University*

**Don Shalvey**

*CEO and Co-founder, Aspire Public Schools*

**Krys Wulff**

*Immediate Past Director-at-large, American Association of University Women*