

## Student Assessment

The California Federation of Teachers supports high academic standards and information curriculum frameworks. The CFT believes that the three major goals of schooling are: democratic equality — the preparation of good citizens, social efficiency - the need to prepare individuals to join the workforce to expand the economy, and social mobility — the means by which individuals gain a competitive advantage in the workplace (Labaree, 1997 and Callahan).

The CFT believes that assessment should be used to support student learning and best practices. Assessment should be a means of fostering growth toward maximizing student potential. It should support the highest levels of student learning possible. Thoughtful and meaningful assessments provide important information that, when combined with information from other sources, can lead to decisions that promote student learning and equality of opportunity.

To help create a common language and understanding, the most commonly used terms and assessments are defined below.

1. **Norm-referenced tests (NRTs)** should be used to compare students, schools, district, and states with each other. NRTs give us some insight into how students in California, for example, compared to students in New York. These tests do not tell us how well any of these students did in relation to a standard. Instead, students are scored based on how well they did compared to their peers. These results are typically reported as percentiles and are reported as a “bell-shaped” curve where half of students will fall below the 50th percentile and half will fall above.

2. **Criterion-referenced tests (CRTs)** should be used to compare individual student performance against a specified standard. CRTs give us information about whether students met the standards. The results are typically reported as performance levels (basic, proficient, advanced). Student scores are based on how well they knew the content and could answer the questions, and not on how well their peers performed on the same questions. Data from CRTs should be used to inform programmatic/instructional decisions.

3. **Formative assessments** should be used to guide instruction. These assessments occur during teaching and are embedded in instruction. Results are received instantly, which allows teachers to adjust their instruction immediately. Most are teacher-developed, and all should be implemented based on teacher judgment.

4. **Summative assessments** should be used to give a snapshot on whether students mastered the standards by a particular point in time. These assessments occur at the end of a unit of instruction and tell us whether students “got it.” Results are also received anywhere from two weeks to two months later. As a result, these tests cannot guide instruction in the short term. However, results can provide some information regarding programmatic/instructional decisions and guide the future delivery of material covered during the unit if, for example, all students failed to comprehend a specific set of concepts and thus all failed to perform on certain questions.

5. **Benchmark/formative assessments** should not interrupt classroom instruction and should reflect the standards/curriculum being taught. Benchmark/formative tests that are used as a predictor of future success are typically not aligned to the curriculum currently being taught and interrupt classroom instruction rather than complement it. Benchmark/formative tests should reflect the content being taught in the classroom and should serve to supplement and provide another piece of information to teachers about their instruction and where each student is in relation to the content they are learning. These assessments should not be created or implemented by out of classroom personnel or consultants. The best assessments, including benchmark/formative assessments, are created and implemented by educators working in collaboration at the grade, department, or learning community level.

6. **Diagnostic assessments** must cover a few concepts in depth. In order for an assessment to provide educators with diagnostic information about a student it must include enough questions about a topic and must include easy and difficult questions (called “outliers”) to make a valid judgment. Most tests, including high-stakes tests, cover numerous topics which mean they can only have a few questions per topic. In addition, these tests are designed to eliminate “outliers” which could skew the data. As a result, they should not be used to make diagnostic decisions. Additional resources are needed to support diagnostic assessment because many of these assessments require a one to one student to teacher ratio and take significant instructional time to complete.

7. **Adaptive testing** should be used to identify the appropriate level at which students are performing for a particular subject or concept. Adaptive testing is done by computer and

asks students more difficult or less difficult questions based on their answers to previous questions. Sometimes called “off-grade” testing, this approach allows teachers to better focus instruction on each child’s strengths and weaknesses by helping to identify the specific concept or process where their learning has broken down.

8. **Educational Growth assessments** should only be used by teachers to estimate their students’ educational growth over time. These assessments can assist classroom teachers in making data informed decisions regarding the effectiveness of instructional strategies and programs for individuals and groups of students. Since such assessments are estimation tools, they should not be used to make high-stakes decisions about students, teachers, or other school staff. Effective, educational growth assessments must be of high quality and must be closely aligned with classroom instruction.

## Current assessment programs

California currently has implemented the Standardized Testing and Reporting (STAR) Program. At present this consists of:

- the California Standards Test (CST) grades 2–11. The CST assesses student mastery of California’s content standards. It covers three subject areas only: language arts and mathematics in grades 2 through 11 and science in grades 5 and 8–11. In language arts, a writing assessment is administered in grades 4 and 7. This assessment complements the multiple-choice test and helps to ensure students can demonstrate specific standards-based skills in writing. Social science is not addressed in grades 2–7, but is assessed in grades 8–11.
- the California Alternate Performance Assessment (CAPA) for special education students grades 2–11.
- the California Modified Assessment (CMA) for special education students grades 3–5 with more academic skills than are assessed in the CAPA, but are not yet ready for the CST.
- the Standards-based Tests in Spanish (STS) grades 2–7.
- the Apenda 3 (Spanish-speaking learners grades 8–11).

the Early Assessment Program (EAP) grade 11 developed in conjunction with the California State University system to show readiness for college level English and mathematics.

Other parts of the California Assessment System include:

- the California High School Exit Exam (CAHSEE) grades 10–12. The CAHSEE has been developed to ensure that students who graduate from high school can demonstrate competency in California’s academic standards. Students graduating from high school have to pass all sections of the exit exam to receive a diploma.
- the California English Language Development Test (CELDT) grades K–12. The CELDT is being administered in English to all students designated English language learners to ensure growth in English language skills. It consists of three parts, an individually administered test of oral skills, a multiple-choice test, and a constructed response portion.

- the Physical Fitness Test (PFT) grades 5, 7, and 9.
- the California High School Proficiency Examination (CHSPE) for students who wish to leave school after grade 10.
- the General Educational Development test (GED).
- the National Assessment of Educational Progress (NAEP).

## CFT concerns with the current assessment programs

CFT supports classrooms that are dynamic and interactive and does not believe that students are passive and static receptacles of information. We are preparing children to be critical thinkers, who have the skills to access, assess, and understand the social, economic, political and personal realms in which they live and work.

Current assessment programs have a multiplicity of problems, including:

- reliability and validity issues,
- the inappropriate high-stakes consequences of testing, equity,
- the narrowing of curricular focus,
- the number of tests and the loss of valuable instructional time,
- the harmful effects of stress on students, and
- a lack of teacher input regarding content and scope
- expenditures and redirection of hundreds of millions of dollars statewide from classroom instruction to administration, collection, and scoring of tests.

## Reliability and Validity

**Reliability** is the measure of consistency for an assessment instrument. The instrument should yield similar results over time with similar circumstances and populations.

A test is **valid** if it measures the learning that it is intended to measure. Tests are only valid if they are representative of the range and major aspects of the standards.

In drawing conclusions from data, it is essential to understand what the test was designed to measure and the margin of error in scores. A test that is reliable and valid for one purpose may not be reliable and valid for another purpose. A single test will not tell us what we need to know about an individual student or an individual school.

As presently constituted, California’s wide variety of tests have **not** been proven to be **either valid or reliable**.

## High Stakes Testing And Its Consequences

The CFT believes that the use of any single test in making important decisions—such as graduation, promotion, and school funding **is never justified**. A single test can’t adequately measure

all the information needed to make such important decisions. In addition, high stakes testing has its own negative affects: 1) it distorts the goals of a broad-based liberal education and displaces thoughtful and creative curriculum with a teach-to-the-test approach; and, 2) it has social justice consequences in that it emphasizes a culturally biased and limiting curriculum that is disengaging for all students, but particularly impacts children of poverty and color. In short, when the use of tests is inappropriate, especially in making high-stakes decisions about a child's future, it undermines the quality of education and equality of opportunity.

CFT believes that the inappropriate use of tests in making critical life decisions is wrong for students and is likewise wrong when used for employment or compensation decisions for teachers. Teaching involves much more than improving test scores on multiple-choice tests. It involves building self esteem, exciting students about learning, teaching students how to work with others, and other similarly unquantifiable contributions. The use of student test scores to judge teacher effectiveness in schemes like "value added" or "pay for performance" will inevitably lead to a decline in teacher morale and may lead to increased teacher attrition, or may be a disincentive for prospective teachers.

## Equity

Children don't enter school equally ready to succeed. Second language learners are at a great disadvantage receiving instruction in a language in which they are not proficient. Studies have shown that the socioeconomic status of a student can also have a significant effect in his/her school performance. Special needs students are also at a disadvantage and are not always given the accommodations to which they are entitled.

Furthermore, not all individuals perform well on standardized tests even when they understand the material. Such individuals are stigmatized unfairly merely because of their inability to perform on these tests under pressure.

The implementation of CAHSEE has caused numerous problems for those students unable to pass it. Those problems range from an increase in the dropout rate to denying students the ability to use a diploma earned through the completion of required course work to gain access to further education including apprenticeships, and limiting the types of employment available to them.

## Narrow Curriculum

The content and format of current student tests tend to narrow the curriculum and limit instructional approaches. Many California elementary and middle schools have eliminated science, social science, art, music, career technical education, PE, and field trips and focus only on tested content areas. Teaching narrowly to the objectives of a particular test runs counter to the genuine goals of education-creative critical thinkers able to work collaboratively to solve today's complex challenges.

## Time Lost to Testing

Teachers and students are devoting increasing amounts of time and other resources to the preparation for and the administration of multiple standardized tests. This leaves little time for teaching and re-teaching to ensure mastery of content standards and the development of critical thinking skills

## Stress on Students

Health professionals report that children suffer harmful levels of stress due to excessive testing.

## Lack of Teacher Input

There has not been sufficient teacher input in testing and assessment reforms. Teachers have not been appointed in sufficient number to commissions, committees, and panels. The state does not reimburse districts for their expenses if they release teachers to do this work, so it is difficult to recruit classroom teachers to these bodies. It is a mistake to leave out the very people who are best able to inform the process.

## Recommendations

California's assessment system must be reliable, consistent, efficient and measure what is important. California should develop a comprehensive standards-based assessment system that relies on multiple measures to evaluate student and school success. The system needs to be fair to all participants and teachers need to be involved in each step of the process.

## Reliability and Validity

It must be shown that scores reported for individuals or for schools are sufficiently accurate to support each intended interpretation. Accuracy should be examined for the scores actually used. It is impossible to draw valid conclusions about the quality of a school's academic program using standardized test scores when those scores can be distorted by the inclusion or exclusion of small subgroups. This problem is exacerbated by the lack of student accountability. Furthermore, it has been the case since the implementation of the No Child Left Behind Act of 2001 (NCLB) that a small subgroup's perceived negative test scores can render meaningless seemingly positive test scores in every other subgroup at a school.

When testing programs use specific scores to determine "passing" or to define reporting categories like "proficient," the validity of these specific scores must be established with regard to the objectives sought. The purpose and meaning of passing scores or achievement levels must be clearly and accurately determined. Sound and appropriate procedures must be followed in setting passing scores or proficiency levels.

With any high-stakes testing program, ongoing evaluation of both intended and unintended consequences is essential. The governmental body that mandates the test should also provide resources for a continuing program of research and for dissemi-

nation of research findings concerning both the positive and the negative effects of the testing program.

## Multiple Measures

Multiple measures of assessment should be used to evaluate student achievement and provide meaningful information to teachers, parents, and students. This information should include student work, coursework grades, and research-based district and statewide assessment tools. Far-reaching and critical educational decisions should be made only on the basis of multiple measures of assessment.

Furthermore, local schools should be empowered to implement an array of assessments to holistically assess student progress.

## Equity

Fairness demands that all children have the opportunity to learn what is being assessed. Adequate resources need to be allocated to support the needs of each student. English learners should be tested in their own language until they are English language proficient. Research has shown that it takes at least five to seven years for an English learner to gain academic English language proficiency.

Special accommodations for English language learners may be necessary to obtain valid scores. If a student lacks mastery of the language in which a test is given, then that test becomes, in part, a test of language proficiency. Unless a primary purpose of a test is to evaluate language proficiency, it should not be used with students who cannot understand the instructions or the language of the test itself. If English language learners are tested in English, their performance should be interpreted in the light of their language proficiency.

Assessments should be used to increase the opportunities for students, rather than deny opportunities, such as employability after high school and college admissions.

In testing individuals with disabilities, steps should be taken to ensure that the test scores accurately reflect what is intended to be measured rather than any disabilities. These students should only be assessed using the tests identified in their Individualized Education Program (IEP). The school should not be penalized for adhering to the IEP as written. The state should allow all available test modifications and accommodations for students receiving special education services.

When schools, districts, or other administrative units are compared to one another or when changes in scores are tracked over time, there must be explicit policies specifying which students are to be tested and under what circumstances students may be exempted from testing. Such policies must be uniformly enforced to assure the validity of score comparisons. In addition, reporting of test score results should accurately portray the percentage of students exempted.

Students who fail a high-stakes test should be provided meaningful opportunities for remediation. Remediation should focus on the knowledge and skills the test is intended to address, not just the test performance itself. Sufficient time must be allowed

before retaking the test to provide students the opportunity for remediation.

## Curriculum

An assessment instrument that is fully aligned to curriculum standards is critical to implementing a true standards-based education system. Full implementation is needed to define the foundation of a meaningful plan for K-12 education where the state sets realistic standards, monitors school achievement, provides sufficient resources, and provides real help when required.

The state needs to create more time for instruction by combining and/or eliminating redundancy in testing.

Both the content of the test and the cognitive processes engaged in taking the test should adequately represent the curriculum. Assessments should not be limited to that portion of the curriculum that is easiest to measure. When testing is for school accountability or to influence the curriculum, the test should be aligned with the curriculum as set forth in standards documents representing intended goals of instruction.

When content standards and associated tests are introduced as a reform, opportunities to access appropriate materials and professional development consistent with the intended changes should be provided. In particular, when testing is used for individual student accountability, students must have had a meaningful opportunity to learn the tested content and cognitive processes. This includes being taught in an environment which is safe, which has optimal class size, and which has the necessary resources. The assessed content must be incorporated into the curriculum, materials, and instruction students are provided.

## Teachers

Teachers need to be in the majority on advisory and decision-making panels and committees. Teachers are the experts in the education field; therefore they should take the lead in developing assessment instruments. The perspective of educational professionals provides needed balance to non-practitioners. The state must fund release time so a larger number of teachers can take part in this process.

Teachers should have the opportunity to share their system of assessment to help outside observers, such as parents, community members, and policy makers, gain confidence in the process they use to assess student progress.

## Conclusion

Assessment should advance student learning and inform teachers as they make instructional decisions. Multiple sources of assessment information should be used when making decisions about the tracking, promotion, or graduation of individual children. Methods of assessment must be appropriate for their purposes and should assess a broad range of subject matter knowledge. Instruction and curriculum should be considered equally in judging the quality of a program. Assessment should be an open process with all stakeholders knowing what is expected, what will be measured, and what the results imply for next steps.



Tests must be valid and reliable for the purposes for which they are used. Tests must measure what the student was taught. Tests must provide students with multiple opportunities to demonstrate proficiency. Tests must provide appropriate accommodations for students with special needs or limited English proficiency. All standardized assessments should be aligned with the state curriculum standards.

The current high-stakes tests in California are a political solution to the educational challenge to close the achievement gap and to provide opportunity for all students. We call on legislators and policymakers to repeal laws and policies that tie significant consequences and rewards to scores on single assessments.

While this paper is focused on assessment, it must be noted that California has developed “world class” educational standards without matching “world class” funding for schools, health care, or other social service support systems necessary for students’ learning and achievement.

Our thanks to the AFT, Mary Alice Callahan, the Parent Teacher Association, the International Reading Association, the National Council of Teachers of Mathematics, the American Psychological Association, and the American Educational Research Association for their policies and ideas, some of which we have included in this paper.

## Bibliography

- “Appropriate Use of High-Stakes Testing in Our Nation’s Schools.” American Psychological Association, May 2001.
- “Assessment That Serves Instruction.” *Education Update* 39, 4 (1997).
- Feldman, S. “Standards are Working.” *American Educator* Fall 2000: 5–7.
- “Finding New Ways to Assess Students’ Learning.” *Education Update* 40, 8 (1998)
- Heubert, J.P. and Hauser, R.M. (editors). *High Stakes Testing for Tracking, Promotion, and Graduation*. Washington, D.C.: National Academy Press, 1999.
- Hoffman, J.V., L.c. Assof, and S.G. Paris. “High-Stakes Testing in Reading: Today in Texas, Tomorrow?” *The Reading Teacher* 54, 5 (2001): 482–492.
- Labaree, David. *How to Succeed in School Without Really Learning*. 1997
- Linn, R.L. “Assessments and Accountability,” *ER Online* 29, 2 (2000).
- Making Standards Work *Educational Leadership* 59, 1 (2001).
- Marzano, R.J. and A.L. Costa. “Question: Do Standardized Tests Measure General Cognitive Skills? Answer: No.” *Educational Leadership* 45, 8 (1988): 66–71.
- “Matching Assessment with Curriculum.” *Education Update* 40, 4 (1998).
- Natriello, G. and A.M. Pallas. “The Development and Impact of High Stakes Testing.” *The Civil Rights Project: Harvard University* (1999).
- “On Literacy Assessment.” *International Reading Association*, resolution, 1991, revised May, 2003.
- Popham, W. J. “Why Standardized Tests Don’t Measure Educational Quality.” *Educational Leadership* 56, 6 (1999): 8–15.
- Popham, W.J. “Burned at the High Stakes.” *American Educator* Winter 2000-2001: 35–39).
- Popham, W. J. “Teaching to the Test?” *Educational Leadership* 58, 6 (2001): 16–21.
- “Shifting the Focus to Learning: California’s Accountability Debates.” *EdSource* June 1998.
- Teaching for Authentic Student Performance. *Educational Leadership* 54, 4 (1996–1997).
- “Trying Too Hard? How Accountability and Testing are Affecting Constructivist Teaching” *Education Update* 43, 3 (2001).
- “Using Assessment to Motivate Students.” *Education Update* 39, 8 (1997).
- Using Standards and Assessments. *Educational Leadership* 56, 6 (1999).
- “The Accountability Question. *Education Update* 4 1, 7 (1999)
- What Do We Mean by Results? *Educational Leadership* 57. 5 (2000).

*Revised 1/31/09*