Report from Members of the Task Force on Alternative Assessments
January 27, 2003

Submitted By:

Alex Caputo-Pearl, Coalition for Educational Justice, Crenshaw HS
Ilya Clark, Coalition for Educational Justice, Muir MS
Mike Dreebin, UTLA Elementary Vice President
Michele Foster, Claremont Graduate School
Linda Guthrie, UTLA Secondary Vice President
Reynaldo Macias, UCLA Chicano Studies Department
Jeannie Oakes, UCLA Institute for Democracy, Education, and Access
Becki Robinson, Beyond the Bell
John Rogers, UCLA Institute for Democracy, Education, and Access
Brian Stecher, RAND Institute
Joyce Germaine Watts, Achievement Council
Susan Way-Smith, Los Angeles Educational Partnership

In May 2002, the LAUSD School Board – along with grassroots groups of parents, students, and teachers – expressed concerns with the negative impact of California’s high stakes tests, such as the California High School Exit Exam (CAHSEE) and the Stanford 9, on teaching and learning in Los Angeles schools. The Board called for the formation of a Task Force to study how a system of alternative assessments might “benefit [the district’s] students and schools.” LAUSD’s PERB convened a diverse group of researchers, educators, and educational advocates in July 2002. The group met roughly once a month between July and December 2002 – at the same time that the San Francisco School Board, in a significant development, convened its own Task Force on the same matter.

The LAUSD Task Force members all have years of experience working with educational assessment. They represent a wide variety of backgrounds, perspectives, and fields of expertise. Befitting the stature of LAUSD, several Task Force members are national leaders in educational assessment and evaluation. Other members have well-established reputations across Los Angeles schools and communities for their efforts to promote high quality learning for all students – through their work within the many diverse organizations listed above.

Through a series of monthly meetings, the Task Force identified a set of research-based understandings about educational assessment generally and high stakes assessment in particular. Task Force members also developed a set of policy recommendations for LAUSD that flow directly out of the research-based understandings of assessment.

As mentioned above, the charge called upon the Task Force to study how a system of alternative assessments could benefit students and schools. This report examines aspects of the environment into which a system of alternative assessments would be introduced. From this contextualized view, there are two existing systems that need to be examined – (1) the system of standards-based education, and (2) the system of assessments, particularly high-stakes testing, that operates within standards-based education.
Thus, the objective of this report is to provide a complement to that of PERB. Within a standards-based system, the theory is that standards are set, equitable opportunity to learn is guaranteed, and assessments are administered – to gain valuable feedback on individuals, groups, sub-groups, etc. The PERB report addresses assessment, the third step in this progression. This report seeks to address assessment within its larger context – that of opportunity to learn and equity; the basic professional, technical, and legal standards for assessment; and the testing burden and its impacts on the overall educational program.

A complete study of how a system of alternative assessments might “benefit [the district’s] students and schools” must address the questions of how such a system would differ from current testing practices; whether such a system would be compatible with current testing practices; whether such a system could be effectively implemented under current conditions regarding opportunity to learn and equity; and whether such a system could simply be layered onto current practices, given the existing testing burden.

By taking on these contextual questions, and by addressing both what the district can do on its own and how the district can approach external constraints, this report works with that of PERB to provide a full picture of assessment.

**Research-Based Understandings of Assessment and High Stakes**

The following understandings are based on the literature on assessment and high-stakes testing, as well as the discussions of the Task Force.

1) Assessment is a process of gathering, describing, or quantifying information about performance. Quality assessment systems include a combination of multiple assessments into a comprehensive reporting format. The purpose of student assessment is to support and enhance learning and teaching by identifying students’ academic strengths and weaknesses; informing instruction and the provision of supplemental intervention and support; and tracking students’ learning over time, at individual and sub-group levels. Assessment can serve multiple purposes – however, one of the problems is that, by trying to make one system serve all purposes, none are served optimally. Increasingly, student assessment is used for accountability – that is, as the basis for important administrative decisions about students, incentives, schools, districts, or states.

2) To serve any of these purposes accurately and fairly, assessments and assessment systems must produce comprehensive, credible, and dependable information. In addition, systems should not lead to negative educational effects, such as disproportionately doling out punitive measures to those students who have not had an equal opportunity to learn, etc. This requires meeting a set of basic professional, technical, and legal standards, such as those set by the American Psychological Association, American Educational Researchers Association, National Council of Measurement in Education, and the US Department of Education Office of Civil Rights. Notably, researchers stress that adhering to these standards is particularly important when assessments are used to make “high stakes” decisions about students. Using a single test for high stakes decisions is particularly dangerous because test scores have a significant margin of error.
3) The State of California’s STAR testing program addresses some of these purposes, specifically some forms of monitoring and accountability, but does not address diagnosis or instructional improvement. Overall, the State of California’s assessment policies fail to meet basic professional, technical, and legal standards in the following ways: Single tests, including the STAR program and the CAHSEE, are used to make “high stakes” decisions such as grade-to-grade promotion, high school diplomas, and distribution of money; the impacts of these high-stakes consequences on minority and low income students is not monitored systematically; and the validity of the scores obtained by limited English proficient students is not ensured. Moreover, students across California have unequal access to qualified teachers, unequal access to textbooks and other learning materials, unequal access to classes, uncrowded buildings with adequate facilities, adequately equipped schools, and unequal access to comparable days of instruction. All of these features influence student learning and test scores, but are outside students’ control.

4) The negative impacts of high-stakes tests, rooted in the aforementioned failures of the State’s assessment policies, are felt most strongly by low-income students of color, and, thus, are of particular concern to districts like LAUSD.3

5) It is not possible to redress these problems by layering alternative assessments on top of existing testing practices. Such additive strategies do not address the failure of existing assessments to meet standards, and the failure to ensure equity. Further, they layer additional assessment requirements onto already overloaded teaching schedules – losing more instructional time, creating more disruptions to the educational program, etc. For example, there are already 17 assessments used in LAUSD.4 These assessments include those that are externally-mandated, those that are chosen for administration by the district, and those that are voluntary opportunity tests, such as the SAT and Advanced Placement tests. If the problems of meeting standards, ensuring equity, and reducing testing burden are to be addressed, and the above criteria for fair and useful assessment are to be met, a comprehensive alternative system is needed to replace the state system.

6) Testing alone can not turn around low-achieving schools – in fact, poor assessments, or good assessments that are administered under poor conditions, can widen existing achievement gaps. Turning around low-achieving schools will require investments in capacity building – that is, improving students’ access to resources, and improving teachers’ knowledge (including knowledge of assessment), pedagogical skills, and understandings of students. This capacity building goes hand in hand with assuring an equitable distribution of resources and learning conditions. Thus, assessments should be used to complement, rather than to supplant, efforts to improve curriculum and teaching – efforts such as those to increase teaching and on-task time in schools, to enhance diversity and critical thinking in the curriculum, to direct resources where they are most needed, etc.

7) LAUSD, working with other districts and with the state, can develop a comprehensive alternative system of assessment that meets the demands of the above research-based understandings.5 By combining on-demand testing with locally-constructed alternative measures and systems to ensure equity, and by combining standardized tests with qualitative, ongoing, classroom-based, and criterion-referenced tools, assessment can better serve instructional improvement and monitoring functions.
Recommendations

The following recommendations are based on the aforementioned research-based understandings and the discussions of the Task Force. They are based on the larger context of assessment with regards to opportunity to learn and equity; the basic professional, technical, and legal standards for assessment; the testing burden and its impacts on the overall educational program; and, assessments’ compatibility with current testing practices and a coherent standards-based approach.

1) Members of the Task Force recommend that the Board initiate further district study on:

- The gap between the standards established for quality assessment and the present conditions of assessment. Special consideration should be given to the problem of testing English Language Learners in English, a language they do not know.
- Lessening the testing burden. For example, many members of the Task Force are very interested in the Learning Record, but it could not be implemented on top of existing tests. When analyzing testing burden, the district should take into consideration that classroom-embedded alternative assessments (that are supported by professional development which enhances capacity-building for teachers) can often be administered efficiently without significant increases in burden, and can provide evidence, gathered over time, that is useful in helping teaching and learning, and in raising student achievement.
- The costs of standardized tests. Costs include loss of teaching days during testing; expenditures for salaries during testing; hiring of consultants; purchase of test preparation packages; administrative overhead; retaining of legal experts, etc.
- The legal and financial implications of a district declining to administer high-stakes tests, or declining to follow through with high-stakes penalties associated with those tests until the equity of those tests is confirmed. This would also include investigation of opportunities for waivers or exemptions under state and federal law.
- Comprehensive systems of alternative assessment which meet the criteria outlined by the State’s Student Learning Working Group (see Endnote #5); meet the basic professional, technical, and legal standards outlined above (see Research-Based Understanding #2); use research from the Coalition for Authentic Reform in Education as one starting point (see Endnote #5); and complement efforts to improve curriculum and teaching, such as those to increase teaching and on-task time in schools, to enhance diversity and critical thinking in the curriculum, to direct resources where they are most needed, etc.
- The availability of single assessments, such as the Learning Record, that could be used as part of a more comprehensive system.

2) Members of the Task Force recommend that:

- The Board immediately initiate work with other districts – such as San Francisco, and other districts where Board Members have expressed interest – towards a moratorium on the high stakes penalty associated with the CAHSEE, until the state can confirm that all students are provided with adequate and equitable educational opportunities and that the test, and all others related to high-stakes decision-making, complies with the basic professional, technical, and legal standards, as discussed in the research-based understandings. During
such a moratorium, the CAHSEE test may continue to be administered and results may continue to be reported.

• The Board direct this work towards the State Board of Education, the State legislature, and the Governor’s office.

• The Board urge that the State produce an Educational Impact Report before reinstating the high stakes consequences of CAHSEE or any other state tests. This report would include, but not be limited to, the material costs, educational and social consequences of a test or assessment procedure, the impacts on curriculum and instructional priorities, and would include effects on students, schools, and community, including academic performance, distribution of physical and human resources and equitable learning opportunities, drop-out rates, etc., each disaggregated according to race, gender and socio-economic level.

3) Members of the Task Force recommend that the Board immediately initiate work with community and research groups towards the development and adoption of an Opportunity to Learn Index for LAUSD that would provide public information on students’ equitable access to critical tools, resources, and practices that are vital to both assessing schools, and evaluating the efficiency of tests and assessments. For example, such an index might include access to uncrowded facilities, uncrowded classrooms, modernized facilities, credentialed teachers, A-G classes, AP classes, up-to-date textbooks, desks, technology/computers, instructional materials, clean bathrooms, appropriate language programs and materials, and counselors, as well as measurements of suspension rates, student achievement at multi-tracked schools, etc. Such an index would provide valuable information on the relationship between educational inputs and outputs. The index should also include a companion process for remedying problems that come to light as a result of the collection of data.

4) Members of the Task Force recommend that the Board use the results of further study, in Recommendation #1, to both guide its own assessment practices and to initiate work, in a timely fashion, with other districts and state policymakers to develop a comprehensive approach to assessment that balances the state’s need for accountability as well as schools’ needs for diagnosis and instructional planning – and, that remedies the problems with the current system.
The Standards for Educational and Psychological Testing, created by the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education, present a number of principles that are designed to promote fairness in testing and avoid unintended consequences. They include:

- Any decision about a student’s continued education, such as retention, tracking, or graduation, should not be based on the results of a single test, but should include other relevant and valid information.
- When test results substantially contribute to decisions made about student promotion or graduation, there should be evidence that the test addresses only the specific or generalized content and skills that students have had an opportunity to learn. For tests that will determine a student’s eligibility for promotion to the next grade or for high school graduation, students should be granted, if needed, multiple opportunities to demonstrate mastery of materials through equivalent testing procedures.
- When a school district, state, or some other authority mandates a test, the ways in which the test results are intended to be used should be clearly described. It is also the responsibility of those who mandate the test to monitor its impact, particularly on racial and ethnic-minority students or students of lower socioeconomic status, and to identify and minimize potential negative consequences of such testing.
- In some cases, special accommodations for students with limited English proficiency may be necessary to obtain valid test scores. If students with limited English skills are to be tested in English, their test scores should be interpreted in light of their limited English skills.
- Likewise, special accommodations may be needed to ensure that test scores are valid for students with disabilities. Not enough is currently known about how particular test modifications may affect the test scores of students with disabilities; more research is needed. As a first step, test developers should include students with disabilities in field testing of pilot tests and document the impact of particular modifications (if any) for test users.


Further, according to the APA, “School officials using such tests must ensure that students are tested on a curriculum they have had a fair opportunity to learn, so that certain subgroups of students, such as racial and ethnic minority students or students with a disability or limited English proficiency, are not systematically excluded or disadvantaged by the test or the test-taking conditions. Furthermore, high-stakes decisions should not be made on the basis of a single test score, because a single test can only provide a ‘snapshot’ of student achievement and may not accurately reflect an entire year’s worth of student progress and achievement.” (http://www.apa.org/pubinfo/testing.html)

*The following are examples of the uneven access of California students to learning resources and conditions related to their opportunities to learn the material on which they are tested:*

- In California, schools with the highest number of minority students have more than 20% of teachers who are underprepared, while in schools with the lowest number of minority students less than 5% of teachers are underprepared (Center for the Future of Teaching and Learning, 2002).
- In California, 42% of teachers in schools with the largest concentrations of low-income children don’t have enough books for their students to take home, and 21% use books that don’t cover the state standards (Louis Harris poll sponsored by Public Advocates, 2002).
- 33% of California students attend an overcrowded school or one in need of modernization (UCLA Institute for Democracy, Education, and Access, 2002)
Most of the students in California who are on multi-track schedules (which are in the most overcrowded schools) are in LAUSD and are minority students. In 2002, 31% of students in these schools passed math on the CAHSEE, while normal schedule passing rates were at 58% (UCLA Institute for Democracy, Education, and Access, 2002).

For example:

- In California, 72% of African-Americans and 70% of Latinos who took the CAHSEE in spring 2002 received failing scores. Only 48% of the class of 2004 have passed both sections of the CAHSEE. Almost 2/3 of whites have passed both portions, while less than 1/3 of African-Americans and Latinos have passed (UCLA Institute for Democracy, Education, and Access, 2002; Los Angeles Times).
- In California, at high schools where fewer than 25% of the students have passed the math portion of the CAHSEE 21% of the teachers lack full certification. Only 9% of teachers lack credentials at schools where 50% or more of the students have passed (Center for the Future of Teaching and Learning, 2002).
- Of the 25 lowest-performing high schools in the state on the 2002 CAHSEE, 8 are in LAUSD serving predominantly Latino and African-American students (UCLA Institute for Democracy, Education, and Access, 2002).
- In other states, exit exams have had a sobering impact on students of color and low-income students. For example, the MCAS test in Massachusetts determines what students receive a high school diploma, beginning with the class of 2003. On the 2000 administration of the MCAS, 61% of African-American students scored at the failing level in English/Language Arts and 77% in the Mathematics test. For Latinos, 66% were failed in English/Language Arts, and 80% in Mathematics. In Texas, in 1998, eight years after Texas began implementing its TAAS test to determine what students receive diplomas, MALDEF charged that only 40% of Mexican-American and African-American students passed the tests, while white students passed at a rate of about 70%. Further, Mexican-American and African-American students represented 85% of the 7,650 students who failed the final administration of the TAAS each year (Linda Mizell, Tufts University, 2001; ERASE Initiative, Applied Research Center).
- Many teachers have reported that the focus on high-stakes testing has contributed to narrowing the curriculum, too much emphasis on teaching low-level skills amenable to multiple-choice formats, and an undue focus on test preparation (United Teachers-Los Angeles, California Teachers Association).

For example, last year, many fourth grade students in LAUSD experienced 5 different major assessments in the span of 2 months (Achievement Council).

Much work has been done on comprehensive alternative systems of assessment. For example, the following recommendations are from Challenging Goals, Guaranteed Opportunities to Learn, Fair and Useful Assessment, Systemic Accountability: A Coherent and Integrated System of High Quality and Equitable Education for California. This February 2002 report was done by the Student Learning Working Group, co-chaired by Jeannie Oakes and Sonia Hernandez, and presented to the California Legislature’s Joint Committee to Develop a Master Plan for Education, Kindergarten through University.

- Recommendation 7 of this report says that the state must develop an integrated and coherent assessment system that monitors programs as well as student learning and guides the provision of additional learning support.
- The report states that the Working Group recommendations generally favor local assessments when it comes to making important decisions that affect individual students, such as retention and graduation: “These local assessments are not only more accurate and fair to students, but they can also be profoundly powerful in affecting the overall understanding of local schooling – both within schools and in communities. Many of the recommendations bring teachers, school district officials, and the public into much more immediate contact with the relevance and appropriateness of local curriculum, pedagogy, and the assessment itself. As such, assessment can be seen as an adult-learning activity in its own right – as well as adults learning about students.”
- The report states that “reports of student performance should describe the programmatic context in which student outcomes are achieved, should emphasize descriptive information about school practices,
instrumental programs, staffing, and other aspect of students’ learning opportunities. This report should be
built upon the evidence developed in the school quality review process. Reporting should also include
analysis of funding, resources, and allocations of expenditures, between schools and districts as well as
among expenditure categories.”

- The report outlines the following criteria for a fair and useful assessment system:
  a) measure progress toward the full range of goals and standards;
  b) be as authentic as possible (i.e. representing real performance tasks and situations);
  c) measure higher order skills and abilities (e.g. ability to analyze, synthesize, apply knowledge in new
     situations, produce, create);
  d) emphasize depth and power rather than breadth and surface knowledge;
  e) be criterion-referenced, measuring and reporting what students have learned, rather than how they stand in
     relation to norms or other representations of the “bell curve.” If norm-referenced tests remain part of the
     assessment system, they should comprise only a small part of the system, consist of a constricted test,
     and be restricted to no more than three grade levels;
  f) report progress beyond minimums toward higher levels of proficiency;
  g) provide multiple ways for students to demonstrate their skills and knowledge, including different kinds
     of performances and multiple assessment strategies and measures that accommodate the diversity of
     communities and learners including students with disabilities;
  h) provide multiple opportunities and occasions for assessment that allow students to demonstrate their
     proficiency and allow teachers to evaluate student growth in a longitudinal, cumulative fashion using
     several kinds of evidence (e.g. samples of work, observations, performance on tasks);
  i) be as open as possible, with publicly known standards and rubrics rather than secret test items. The results
     and the test items themselves should be made available immediately and reported in ways that enable
     teachers to guide students and design further learning opportunities that allow students and families to
     take a more active role in directing their own learning;
  j) allow for the determination of student accomplishment by exhibition of performance rather than course
     credits or seat time;
  k) prohibit the use of a single measure to make high stakes decisions about students; important decisions,
     including course placement, grade retention, graduation, state scholarship funds, and college eligibility,
     must not be made solely on the basis of a test score.

Additionally, examples of comprehensive alternative assessment systems exist. Maine and Nebraska have
systems in place (http://www.fairtest.org/examarts/Spring%2002/Maine_and_Nebraska.html). Also, the
Coalition for Authentic Reform in Education (CARE) in Massachusetts has developed a four-pronged model:
- Local Assessments. Each school would submit its accountability plan for review and approval to a
  regional board, established by the state Department of Education. The plan would outline how the
  school will assess progress toward a broadly stated set of competencies.
- External Quality Reviews. Every couple of years, schools would undertake a self study. External
  auditors would review the self study, visit the school, report on progress toward academic excellence,
  and report on progress towards equitable distribution of quality resources and learning opportunities.
- Standardized tests. These would be limited to literacy and numeracy and would not have high stakes
  decisions attached.
- Annual Reports. Each school and school district would annually report to ‘stakeholders’ on a set of
  ‘indicators’ developed by the state. These would include, but not be limited to, academic performance
  reported in terms of race, gender, low income status, special needs and limited English proficiency, and
  would include access to quality resources and learning opportunities (http://www.fairtest.org).

Almost ten years ago, researchers George Madaus and Walter Haney at the Center for Evaluation and
Policy Research at Boston College estimated these costs at approximately $20 billion annually for the nation
(Haney, Walter, George Madaus, and Robert Lyons, The Fractured Marketplace for Standardized Testing,
There are several organizations that have done path-breaking work on tools that would support such an index – the Applied Research Center’s Racial Justice Report Card, the work of the Institute for Democracy, Education, and Access, the Achievement Council, and the Coalition for Authentic Reform in Education. Further, LAUSD could draw on the assessment expertise of UCLA’s CRESST and the research on accountability conducted by RAND. Most of these resources are accessible to LAUSD through the membership of the Task Force (http://www.arc.org).