

## **POLICY BRIEF**

# **The Achievement Gap in California: Context, Status, and Approaches for Improvement**

*Eva L. Baker, EdD - University of California, Los Angeles, Center for Research on Evaluation, Standards, and Student Testing (CRESST)*  
*Noelle C. Griffin, PhD - UCLA, CRESST*  
*Kilchan Choi, PhD - UCLA, CRESST*

### **Overview**

By way of introduction, we begin with a definition of the term achievement gap, and address potential misconceptions in its interpretation. We then trace briefly the policy context surrounding the accountability systems now in place, including the history and logic that led ultimately to the No Child Left Behind legislation [1].

To place the California achievement gaps in context, we focus on one high profile international comparison by the Organisation for Economic Co-Operation and Development (OECD; The Programme for International Student Assessment – PISA). This comparison details U.S. performance in mathematics in terms of place within the international distribution in 2003 and 2006. Of particular relevance is the finding that, while the U.S. is in the middle of the pack when considering raw scores, our performance drops when adjusted for economic indicators. This finding means that the U.S. is underperforming relative to its overall advantages. Moreover, the U.S. data show considerable spread (or gap) among student performers. A proxy for ethnicity in PISA is

a comparison of students by their generation status (i.e., first vs. second generation) in each country, an analysis where the U.S. is once again among the least effective in closing gaps. The conclusion from this analysis is that the U.S. has not made major progress since the publication of *A Nation at Risk* (1983), a document that identified the challenge to our educational system to keep up with the rest of the world or chance the potential future loss of global economic leadership.

Moving from the international stage, our second set of analyses reports overall U.S. trends using U.S. data, again principally drawing on a single indicator, the National Assessment of Educational Progress (NAEP) in mathematics and language arts achievement. We report California's poor standing, or overall gap compared to other states, on NAEP and its documented inability to catch up to the levels of other states. We report data for student subgroups with persistent achievement gaps, including African Americans and Hispanics, compared with white populations and with each other. We also briefly discuss data for English learner (EL) students, students eligible for

free lunches, and gender differences. In particular, we use other states as benchmarks for both level of attainment and changes for under-performing groups, with the inevitable conclusion that California is unlikely to catch up to other states nor to close its achievement gaps at its present rate of growth.

With the data establishing but a middling position for the U.S. in international achievement comparisons and a poor showing in NAEP results, we turn our attention to the California scene, using descriptive, archival data on educational achievement available through the state web sites. We consider the Academic Performance Index (API), the California Standards Tests (CST), and the California High School Exit Examination (CAHSEE), among other indicators. Our analyses attend to differences found among subgroups of students, and status and rates of change are described. Here again, we find largely implacable differences among subgroups with almost no suggestion that the rates of change for under-performing subgroups will rise rapidly to “close” the gap. In addition, we consider, although not in detail, a handful of postsecondary and preschool reports, which largely support the same finding.

In order to assess the research base to augment our descriptive analysis of California achievement, we follow this section with a review of salient research studies regarding the closing of the achievement gap. Relevant findings related to school practices and achievement are summarized. For instance, some studies found high performance for typically lower performing groups when efforts have been made in generating social capital, that is, in team building, raising degrees of trust within the school, developing a sense of collective efficacy, the availability of shared and

transparent support systems, among other features. Necessarily these studies are for the most part limited to existing comparisons of programs or micro-treatments, rather than studies where interventions have been randomly assigned to students or schools. Therefore, the inferences drawn from them are weaker than we would like, but should be considered as sources of potential useful strategies.

We next consider assumptions, purposes, and roles of assessment, and the centrality of validity evidence in general and as they relate to our charge. Here we analyze the relative attention given to the intent of the California standards framework and the tests associated with it (the CSTs), the limitations of survey measures as outcome measures when they are intended to detect changes in learning, and interpretation options for such measures. We do not believe that the form of the measures is necessarily a limiting factor, although they are less well received by teachers. Rather, we are concerned with whether the consequences of assessments and the strictures to use data for decision making are based on fairly understood test content, adequate content sampling, and front end analyses to support teacher diagnosis and teaching. We discuss formative assessment, its potential and some data in its support, but note that teachers and institutions of higher education don't consistently subscribe to its effective use. We describe the importance of thinking of interventions as the method to reduce the correlations between pre- and post-instructional performance. Addressing concerns for validity interpretations relative to their consequences is important but largely absent with the exception of the CAHSEE report [2] because of the urgency of schedule, politics, sanctions, as well as time and cost of conducting appropriate studies.

## Conclusions

The major conclusions of the report are stark. A first important conclusion from the data is that we are not succeeding now, nor are there harbingers of improvement on the horizon. California students fall well below national expectations and averages for achievement, which means they are in the lower reaches of world performance levels. In California, as nationally, membership in identifiable subgroups remains a strong predictor of level of success on examinations—the gap endures. California is near the bottom ranks of the other states in this regard, and while we can argue for the uniqueness of our circumstances, they alone do not forgive our responsibility to educate all students well.

A second, inescapable conclusion is that, to the extent it occurs in the early grades, learning is neither cumulative, well sustained in middle and upper grades, nor supported, repaired or relevant enough to our youth to change their performance patterns. While potentially a test artifact, California displays the national trend that academic performance decreases as students rise in the grade structure of schools, a fact persisting in the face of the differential dropout of lower performing students.

Third, we conclude from analyses of California policies, of those of selected international educational systems, and of

states that are outperforming California, that our educational policies have not been consistent or sufficiently coherent, and that approaches to support coherence have been inadequately sustained, occasionally ill-chosen, and too remote from instruction and learning.

In summary, it is the negative conclusion of this report that, based on the available evidence, California has not done enough to change the day-to-day learning experiences of students to close gaps in achievement. We have not designed nor applied widely effective instructional systems in which teachers have flexibility and sufficient content and pedagogical knowledge to address each student's need. We have not given a coherent and sustained approach to instruction, and we have not been effective in our approaches for underperforming groups. If we have spent much time and energy on structural and governance issues over the last two decades, it is now essential to focus right on the classroom. We conclude an overhaul is needed of our thinking and acting on the matter of the achievement gap. We need to change our attention to the performance patterns of individual students, within groups, if necessary, but the individual must override thinking in great swaths about remedies for groups or institutions.

## Counter Arguments to Our Conclusions

Counter arguments may be made based on different interpretations, sources of research and observation of practice, and analyses of feasibility. They include: 1) that the effects of changes in California are below the surface and will emerge soon; 2) that we have fine standards and tests and don't want to undo the processes that melded political

and programmatic elements; 3) that immigration and the myriad languages and cultures represented in California by first, second, and third generation learners must be respected and English learning is sufficiently successful; 4) that the state policy can only deal with top-down mandates for global recommendations; 5)

that to understand coherence we need trusted ways to measure instructional practice; 6) that we should not add or reorganize outcomes in the middle of the stream, especially for students who are currently struggling; 7) that our individual-level student monitoring system will help despite the fact it will likely rely on limited test formats; and 8) that policy and practice has

been and continues to be focused on responding to NCLB.

We naturally find these assertions wanting (as well we might, as we wrote them) and counter that we cannot make minor changes in policy and practice given the dismal level of California educational effectiveness as measured by existing survey instruments.

## Obstacles to Policy Implementation

Obstacles to revised policy implementation in the areas of measurement, accountability, assessment and coherence derive from both political and technical factors. On the political side, the players, including the State Board (with its naturally changing membership and leadership), numerous *ad hoc* task forces, the State Department of Education, the Office of the Secretary of Education, institutions of higher education, and professional organizations, have only fitfully worked together, and on more than one occasion have operated at distinct cross-purposes.

In other ways we are not efficient or sufficiently focused on programs and incentives to assure appropriate outcomes. In California, because of our recent history and growing diversity, educators and policymakers spend considerable time (for instance, for innovative major projects) negotiating with group representatives about decision and program components. As a result, the interval for innovation start-up is protracted compared with more homogeneous settings, and leaves less time, energy and mobilization for making deliberative programmatic choices intended to focus on student learning.

A view widely shared by the public, policymakers, and educators is that performance on the “test”—whether it is CAHSEE or CST—is understood as the key

outcome of education. It swamps all other options, including unsought but potentially compelling evidence about students’ ability to retain, sustain, apply, and transfer their learning as measured on the external examination. One consequence of this reality is the ever-growing focus of instruction on just the item types on the test, rolling over content options outside of the measured domains, and ignoring interesting subtopics within the standards topics. Our lower-performing students are, as a consequence, given little choice, and school for some means a focus on the test, rather than the various ways the standards could be applied and made useful, especially at the secondary level.

With regard to validity questions, including the urgent need to document the impact of instruction on measures used to fundamentally evaluate the effectiveness of education, the state has not pushed vendors of examinations to provide strong evidence on information on the range of purposes intended for the examinations: accountability, accuracy, representativeness of standards, instructional sensitivity, diagnostic utility by teachers, and consequences overall for the educational standing and health of the schools. This is not at all unusual and a function of the speed and cost of implementing assessments.

In addition, there remain great differences in interpreting—beyond the external tests—common expectations of performance for all groups. In secondary schools, these differences in expectations and instruction can be documented by looking at the proportion of non-credit bearing remedial coursework required of entering freshmen in some sectors of the higher educational system, where freshmen with reported high performance (e.g., grade point average or grade in relevant courses), find themselves

in remedial settings. This assignment leads to college dropout and disaffection as well as greater time and cost to complete a course of study. Thus, the inhibitions between widespread and serious connections between secondary school and institutions of higher education must be overcome by rapid and intense engagement of the P-16 community if a serious understanding of the standards and their importance for achieving social justice and intellectual parity is to occur.

## **Courses of Action: Our Recommendations**

Based on our review of existing data, research base, and the state of the art regarding assessment policy and practice, we formulated a series of assessment-focused policy recommendations directed at addressing the achievement gaps in California. For any of these recommendations to succeed will require not only expertise, collaboration, and trust between a wide range of stakeholders and resources, but also long-term administrative and policy support for the changes to take hold. Given the time-sensitive nature of the work ahead, for each recommendation we provide a suggested timeline for completion and approval.

### **Standards**

Standards need to be reorganized, reduced, and include elements of cognitive readiness integrated with content to support their coherence and cumulative impact.

Action: Begin immediately, convening experts in cognition, content, and pedagogy with guidance from states that have already initiated the process and supported by new technology. Suggested completion and approval: September 2008.

### **Curriculum Frameworks**

Curriculum frameworks should be resurrected or refined for all subject areas at a level of specificity to guide subtask instruction, to provide options for teachers in addressing shortfalls, and to support the coherence of the system and fairness of the system.

Action: Begin immediately convening subject matter, cognitive and pedagogical experts from across sectors. Review available frameworks and California documents. Create and test formats, and by coordinating with Standards group, lag only somewhat in the creation, trial and fielding of these documents. Completion and approval: June 2009.

### **Measurement System—Design**

A blended measurement system is immediately needed that includes uniform external measures of a limited set of standards with clear and sufficient content and cognitive sampling. This endeavor would also include a formative assessment system driven by teachers to assess the acquisition of subskills and prerequisite knowledge outlined in the curriculum frameworks. At the secondary level end-of-course examinations should be acquired, and

implemented, and made a part of accountability systems. These examinations should offer additional opportunity for student choice and motivation. To the extent possible, the state should plan for high school exit criteria that also include acquisition of desired skills certified by businesses, non-profit organizations, or other providers. It should be expected that California will move rapidly to assessment situations that capitalize on technology as a means of determining deeper understanding and more complex learning. As these elements come online, the emphasis on CAHSEE as a graduation requirement might diminish. In addition, organizational effectiveness measures related to the fair distribution of social capital for all members of the learning community should be undertaken and included in assessments of educational progress and level. These measures, along with measures of instructional practices will move the system away from test-focused learning to standards-based accomplishments. Actions: Begin immediately on end-of-course examinations for secondary school, examining options in the U.S. and abroad. As standards are clarified, design, with vendors assistance, try out and use preliminary quality criteria (e.g., of instructional sensitivity, accuracy, standards fidelity, and fairness). Following standards approval, revise CSTs accordingly with technical approaches to maintain trend. Completion and approval: January 2010.

### ***The Measurement System–Validation***

The purposes of measures, indices, and other external indicators of effectiveness must be clarified, and evidence obtained to document or to guide improvement on the extent to which the test results interpretations support the purposes [3]. Key among these purposes is instructional efficacy, and evidence of the value of instruction for preparing students

for required examinations is a minimum criterion. Evidence with respect to their acquisition of the content intended by standards also needs to be collected through additional measures of retention, application and transfer. These measures may well be incorporated at key points in the emerging measurement system to provide guidance and incentives for more coherent, integrated instruction, including attention to social capital. All validity studies must consider the performance of identifiable subgroups and linguistic attributes of the examination. The studies should be conducted as soon as possible, referencing existing measures already on the drawing board. Actions: Identify critical validity studies for measures of learning and school effectiveness budget and schedule for needed validity studies. Develop mechanism and fund studies by June 2010.

### ***Interaction Among Sectors***

Collaboration among the sectors of the P-16 group, from governance, structural, political, standards, training, and interaction perspectives should be accelerated by general agreements and detailed connections to promote coherence, fairness, understanding, common expectations, and support for under-performing students. These agreements must include professional organizations and other powerful players on the educational scene. A clear set of goals and short-term milestones should be adopted and followed. Actions: Select and convene membership, identify subcommittees relevant to standards, underperforming groups, frameworks, implementation, professional development, and validation. Ongoing.

### ***Students First***

The system must shift its attention to the classroom and in particular to the growth and development of each student,

independent of background, but focused on instructional needs. Teachers will in many situations need to be provided with new incentives that focus on individual rather than group progress. Gaps should diminish and new skills should emerge if our aforementioned recommendations related to standards, frameworks, valid assessments, technology, student choice, and rights to social capital are taken together, taken seriously, and implemented with care and

speed. Action: Start immediately with a campaign to model and illustrate what a focus on individual learning would operationally mean. Create or adopt approaches from other successful venues for use. Focus on micro-interventions rather than complex, unwieldy or burdensome packages. Encourage exploration with results. Create an incentive system that rewards attention to individual growth. Ongoing.

## References

1. No Child Left Behind Act of 2001, Pub. L. No. 107-110, § 115 Stat. 1425 (2002).
2. Becker, D. E. & Watters, C. (2007, October 24). *Independent evaluation of the California High School Exit Examination (CAHSEE): 2007 evaluation report*. Alexandria, VA: Human Resources Research Organization (HumRRO).
3. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.