1-1-2009

Unraveling the Myths of Accountability: A Case Study of the California High School Exit Exam

Kerri Ullucci
Roger Williams University, kullucci@rwu.edu

Joi Spencer
University of San Diego

Recommended Citation
Ullucci, Kerri and J. Spencer. 2009. "Unraveling the Myths of Accountability: A Case Study of the California High School Exit Exam." *The Urban Review* 41 (2): 161-173.

Urban Rev (2009) 41:161–173
DOI 10.1007/s11256-008-0105-y

Unraveling the Myths of Accountability: A Case Study of the California High School Exit Exam

Kerri Ullucci · Joi Spencer

Published online: 24 September 2008

Abstract Believing that accountability could be a vehicle for change, the California Department of Education (CDE) requires all high school students to pass the Calfornia High School Exit Exam (CAHSEE) in order to graduate. In doing so, California joins many others states in mandating a high school exit exam as a current or future requirement for graduation. In this essay, the authors will argue that this testing approach to school change is based on myths about the role of assessment, the information testing can provide and the impact high stakes testing has on urban schools. Although California is the focus of this analysis, these issues are salient across the county. Testing as a solution to poor student achievement is based on faulty assumptions. It is these assumptions this piece seeks to address.

Keywords High-stakes testing · Standardized tests · Accountability

Like many states across the country, California is struggling to revamp its educational system. As of 2004, 50% of African American males failed to graduate from high school (Education Week 2007). Just 22% of fourth graders attained proficiency on the NAEP reading assessment in 2005 (National Center for Education Statistics 2006). A full 38% of schools did not make adequate yearly progress (AYP) during the 2004/2005 school year (Education Week 2006). Faced with a daunting array of shortcomings, California has been searching for an

K. Ullucci (✉)
Education Department, University of Massachusetts, Dartmouth, 285 Old Westport Rd., North Dartmouth, MA 02747, USA
e-mail: kullucci@umassd.edu

J. Spencer
School of Leadership and Education Sciences, University of San Diego, 5998 Alcala Park, San Diego, CA 92110, USA
e-mail: joi.spencer@SanDiego.edu

educational cure-all. With the passage of the California High School Exit Exam (CAHSEE) in 1999, California appeared to anchor its hopes for educational reform to standardized tests. Believing that accountability could be a vehicle for change, the California Department of Education (CDE) required all high school students to pass the CAHSEE in order to graduate. California, along with 22 other states, now mandates a high school exit exam as a current requirement for graduation. In doing so, the CDE tied the reinvigoration of schools to a common, if not well examined, premise: testing can cure all that ails us.

Like many states, California has equated school success and improved student outcomes with students' abilities to perform well on tests. High stakes teaching-testing which attaches serious consequences to passing or failing-is not novel. States including Alaska, Massachusetts and Texas have instituted high school exits exams which fix graduation to the passing of tests. Despite seemingly benevolent intentions, high stakes testing does not provide a quick-fix remedy to the issues that burden US schools. In this article, we argue that a testing approach to school change is based on misunderstandings about the role of assessment and the information testing can provide. In particular, we are concerned with the significant negative impact high stakes tests have on urban schools. Although the CAHSEE serves as a case study for this analysis, these issues are salient across the educational landscape as more states join the high stakes testing trend. The CAHSEE's shortcomings are not unique. It is these myths and assumptions that this piece seeks to unwind.

The Background

The California legislature passed the California Public School Accountability Act in 1999 as the final prong of standards-based reform. As previous reform measures already standardized content and curriculum throughout the state, the CAHSEE became the instrument which ensured these standards were being met. Public school students take the CAHSEE the spring of their sophomore year. The test contains both language arts and math sections. Students who fail one or both sections of the test have up to five additional opportunities to pass. The first sitting of the CAHSEE was in Spring of 2001 for volunteer freshman. After a brief period of recalibration, the CAHSEE began to serve as a graduation requirement for the class of 2006.

According to the California Department of Education, the CAHSEE has three stated goals:

- to ensure that pupils who graduate from public high schools have grade appropriate skills in reading, writing and mathematics
- to significantly improve pupil achievement in public high schools
- to help identify students who are not developing skills that are essential for life after high school and to encourage districts to give these students the attention and resources needed (California Department of Education 2006).

It can be argued that the CAHSEE is based on laudable aims. First, the state wanted to ensure that a California diploma had value. Several conservative organizations have critiqued the academic rigor of high schools, including the

American Diploma Project (2004), The Business Roundtable (1998), and California Business for Education Excellence (2005). The California Department of Education itself recognized that local proficiency standards were set below high school level (California Department of Education 2006). By enacting an exit exam, the state wished to "ensure that pupils who graduate from public high schools can demonstrate grade level competency" (California Department of Education 2006, p. 1). Secondly, the state wanted to improve overall pupil achievement in high schools. Through the test, the state believed they could monitor that state curriculum frameworks were actually being reached. Finally, the state voiced equity concerns. Through the state-wide implementation of the CAHSEE, the CDE attempted to ensure that all schools were being held to similar expectations, regardless of their location or demographics.

CAHSEE as Assessment and Accountability

The CAHSEE is a high stakes system of *assessment* and *accountability*. While these terms are often used interchangeably, their meanings are not the same. Well created assessments allow educators to gauge what students know. Teachers use formal and informal assessments to collect insights about students' understandings, to look for gaps in knowledge, and to better understand the effectiveness of their own teaching. Assessments can take traditional forms, such as tests and quizzes, or more novel forms including projects, performances, portfolios, reports, interviews and reflections. Good assessment provides immediate feedback as to potential next steps a teacher can take.

High stakes tests employ standardized examinations to *assess* particular sets of skills and knowledge. Not all standardized examinations are high stakes. When performance on a standardized examination becomes linked to consequential rewards and sanctions, it becomes a high stakes test. Successful performance on the CAHSEE provides high schoolers with a significant reward- a high school diploma. Conversely, for those who fail to pass, it sanctions them with a life-altering penalty: the denial of a diploma. This approach, the CDE posits, will make California high school students *accountable* for their own learning. In this context, "accountable" can be replaced with "responsible". Accountability is about making parties responsible for a certain outcome. Herman (2007) defines accountability as the "idea that individuals, organizations and the community not only are responsible for their actions, but must also answer for their performance to an outside authority that, in turn, may impose a penalty for failure." (p. 3). Accountability has a vastly different purpose than assessment. We concur with Fusarelli's claim that accountability is a political construction (2001). Heubert and Hauser's National Research Council study (1999) cautions that standardized tests are often used as policy instruments, designed for political ends. Accountability is also inherently a marketplace ideal. It provides a mechanism for guaranteeing the input (the state or district's financial investment) is being used wisely to increase output (students ready to contribute to the workforce) (Engel 2000). Accountability systems attach "important administrative decisions about students, incentives, schools, or districts" (Caputo-Pearl et al. 2003), to some

external measure. In the current climate, accountability systems do not simply address academic goals. Instead, accountability systems include political, economic, personal and social goals for its various stakeholders (Laitsch 2006).

The CAHSEE follows a long tradition of the use of standardized assessments in American society in general and in K-12 education specifically. Formal standardized assessments emanate from behaviorist philosophy which, "places great emphasis on what can be measured quantitatively" (Gunzenhauser 2003, p. 53). Behaviorist theories began to dominate American schools in the late 1800s, (Oakes and Lipton 1999) manifesting themselves in the form of IQ exams.[1] Some of the earliest IQ tests served to filter out mentally "inferior" children from regular classes using criteria that were clearly subjective (Sacks 1999). Standardized testing exploded in the 1900s. Exams became employed as sorting mechanisms for a wide range of populations, from Army recruits to elementary school students. From their inception, such exams favored the privileged classes. Tests were culturally biased to favor well-off, White subjects (Sacks 1999) while immigrants, English language learners and the poor were cast to the bottom of the intellectual hierarchy.

The science behind behaviorism has been challenged on different fronts. First, the influence of theories of cognition challenged long-held ideas about the nature of intelligence. Cognitive theorists contended that learning was more than a series of stimuli and responses. Rather, learning was the active engagement between the learner and his or her environment. IQ tests specifically can be critiqued based on their truth claims. The belief that all phenomena can be measured quantitatively, "no longer inform(s) the work of today's psychometricians" (Gunzenhauser 2003, p. 53). Furthermore, he writes,

"…philosophers of science, have taken a more probabilistic and fallibilistic view of knowledge. Psychometricians and other statisticians believe that what we know scientifically is only known with a certain probability … all knowledge is built on foundations but is fallible; it is our best approximation of the truth until we are proven false."

According to this line of thinking, any test is simply a *measure* of student ability and no one can know for sure what a student's true ability is. Today's formal assessments-including the Scholastic Aptitude Test (SAT), Advanced Placement (AP) examinations and high school exit examinations such as the CAHSEE-are descendants of those first IQ examinations. In considering the limitations of the CAHSEE, it is important to be mindful of the beliefs and assumptions which accompany these approaches in order to question their validity and explanatory power for California.

High Stakes Testing as a Change Agent: Unraveling the Myths

Community members are bombarded by the rhetoric of testing. It is sold as a necessary part of schooling, a tradition that might not be well loved, but one that is needed nonetheless. Since most adults remember sharpening # 2 pencils and filling

[1] There were many other manifestations of behaviorism in classroom, including drill and practice and positive reinforcement schemes used for discipline purposes.

in bubbles themselves, pro-testing advocates have an advantage. Few adults will remember testing as traumatic, so how bad can it be for their kids? The first assumption relates to this shared experience. The assumption is that testing is a normal and necessary part of schooling, one that is rather harmless. While the practice of testing has been a common feature in schools for quite a while, the *ramifications of the testing are a new phenomenon*. High stakes consequences, including withholding monies for low performing schools and denying diplomas, are a recent development. Previously, if a student performed poorly on a test, the consequences ended with a low score. The low score did not translate into her inability to graduate, or in state directed sanctions against her school. Now, scores have tangible, real consequences for students and schools. They are not the tests of the past; now they have teeth.

Several other assumptions flood the media and are often sold as commonsense arguments to the public. They include:

1. Higher standards and rigorous tests will lead to better student learning.
2. High stakes testing provides a valid measure as to who deserves a high school diploma.
3. High-stakes testing will act as a motivator. It will encourage children to work harder and be more accountable for their learning.
4. High stakes testing is a fair and neutral tool. As all children are required to learn the same standards, all children have the same chances of success.

Assumption 1: Higher Standards and Rigorous Tests Will Lead to Better Student Learning

Higher standards would seem to be a logical precursor to better student achievement. By focusing the expectations teachers have for students, it would naturally follow that student outcomes will improve. Unfortunately, this is often not the case. The limitations are two fold. First, the tests tend to limit the curriculum that children experience. Thus, unless specific content is tested on the state exams, chances are the material will not be covered in class. In a study completed by RAND, researchers found that the curriculum in Texas was curtailed after the passage of high stakes testing, and content areas not on the test received little attention (Klein et al. 2000). A second study found similar results of curriculum narrowing and teaching to the test in New York (Mathison and Freeman 2003). Recent work by Au (2007) investigated 49 studies on the impact of high stakes testing on the curriculum. His findings suggest that "tests have the predominate effect of narrowing curriculum content to those subjects included in the tests, resulting in the increased fragmentation of knowledge forms into bits and pieces" (p. 264). Examples from abroad are also illustrative. Over the course of three years, Boaler (2002) studied math classrooms in two different schools in England.[2] In one

[2] On indicators of SES, racial composition, and mathematics achievement there was no significant difference between the students at these two working-class schools.

school, Amber Hill, mathematics instruction centered on preparing students for standardized tests. In another school, Phoenix Park, mathematics instruction was more concerned with open-ended learning. At the end of year 10, all students in England were required to take standardized assessments (the General Certificate of Secondary Education, GCSE). Phoenix Park teachers devoted very little time to preparation for this test. Despite such a focus, "significantly more of the Phoenix Park students passed the GCSE examination than the Amber Hill students (p. 99)."

Secondly, there is little evidence to show that higher test scores correlate with higher achievement in other measures, including SAT scores, NAEP scores, or even classroom grades. Amrein and Berliner (2003) argue that "If statewide high-stakes testing policies actually improve student learning, we should see that improvement reflected not just in the states' own test scores but also in independent measures" (p. 5). Haney's (2000) often cited work on the "Texas Miracle" found that gains on the Texas Assessment of Academic Skills (TAAS) did not translate to other academic gains. Haney looked at other indicators of student performance such as the NAEP, the SAT and the Texas Academic Skills Program (TASP). His research revealed that in the same years that Texas had reported gains on the TAAS, student performance on these other indicators had either no significant change (as in the case of the NAEP) or significantly decreased (as with the SAT and the TASP). Such decreased student performance may be an indication of narrowed teaching goals where students are exposed solely to information necessary for passing their states' high-stakes exam.

In an examination of 18 states' high-stakes outcomes, a study by Amrein and Berliner (2002b) found that "Sixty-seven percent of states that use high school graduation exams posted decreases in ACT performance after high school graduation exams were implemented… Fifty-six percent of states that use high-stakes high school graduation exams posted decreases in SAT performance" (p. 48). Moreover, they also found no consistent improvement in NAEP scores among schools with high stakes testing. The myth that higher standards coupled with high stakes testing will lead to better prepared children is not automatic. What we are perhaps seeing is that students are becoming better test takers. There are few indications that they are becoming stronger students when we look at other indicators.

Assumption 2: High Stakes Testing Provides a Valid Measure as to Who Deserves a High School Diploma

By far, the most pivotal question facing the CAHSEE is whether a high-stakes test is an appropriate and valid measure of student achievement. An important body of work challenges the claim that test scores capture real achievement (Hilliard 2000; Haney 2000; Amrein and Berliner 2002a; Cuban 2001). Linn and Baker (2002) assert that assessment must provide *valid* information about students and schools. At its most basic, validity requires that a test instrument measures what it claims to measure (Jaeger 1993). Without validity, a test can not be used as a decision-making tool. In the case of the CAHSEE, one goal of the test was to ensure a high school

education; hence, a valid test must reflect this goal. At this general level of validity, the CDE contradicts itself. The test is given in the 10th grade. Therefore, students can only be tested up to a tenth grade level. At this point, the first adjustment must be made; the state is actually measuring whether or not a child has a 10th grade education. Moreover, the math section assesses 6th grade, 7th grade and Algebra 1 standards (often taken in the 8th grade). In this case, the state is ensuring an 8th grade education. Finally, the pass rates for these tests are particularly low; students pass with a score of 60% correct in English and 55% correct in math (California Department of Education 2007). In essence, the CDE is requiring that students can correctly answer approximately half the test at an eighth or tenth grade level. Answering approximately half of a test correctly in a traditional classroom setting would be failing.

Valid usage is an additional concern in assessment validity. Not all tests measure all things. For example, a norm-referenced achievement test might be valid in providing information about a student's content knowledge in reference to her peers; however, it may not be valid as a factor in retention (Smith and Fey 2000). Moreover, many organizations which *specialize* in educational measurement caution against the use of high-stakes testing as a single form of assessment, including the American Psychological Association (2001), the American Educational Research Association (2000) and the American Evaluation Association (2002). These experts in educational measurement have been largely ignored by governmental stakeholders regarding the responsible use of high-stakes testing. Here the myth seems almost insurmountable; despite all that is known about the correct use of high stakes testing, the zeal for separating the deserving from the undeserving has blinded us. We continue to ignore the fact that one single test can not possibly measure the complete competence of a 16 year old student. That is, after all, what high stakes tests are doing. Test now matter more than classroom grades, more than teacher recommendations or portfolio reviews. Tests trump all.

Assumption 3: High-Stakes Testing Will Act as a Motivator. It Will Encourage Children to Work Harder and be More Accountable for their Learning

Following a market model of motivation, policymakers believe that rewards and sanctions will provide the needed catalyst for change. Just as a stock broker is encouraged to work harder to receive her bonus, children will work harder to receive their diploma. This is a simplistic explanation, one that forgets we are working with *children* who are trying to *learn* over the *long term*. Amrein and Berliner assert that "the assumption that high-stakes tests motivate students appears to be seriously flawed. In fact, such tests often decrease student motivation and lead to high student retention and drop out rates" (2003, p. 33). Research by Sheldon and Biddle (1998) also noted that high-stakes testing often decreased student motivation and limited intrinsic interest in learning. Students tend to be less intrinsically motivated when rewards are linked to a task (Amrein and Berliner 2003). Moreover, this assumption relies on students being motivated by the "reward" provided—a diploma. This is not always the case. Madaus and Clarke (2001), drawing on

research from Europe, argue that students who are from marginalized populations may see a diploma as meaningless if the student sees no personal or professional benefit in having one.

We can assume that if testing did increase motivation, and students were willing to work harder than they did before tests were instituted, then more students should be graduating. After all, this is part of the logic: if students know that the test could bar them from graduating, they will redouble their efforts to study hard and commit to learning. In fact, the very opposite is happening. Some research shows that schools with high stakes testing have higher drop out rates than those without (Amrein and Berliner 2003). A quantitative analysis of data from 25 states addressed the relationship between the pressure of high stakes testing and student achievement (Nichols and Glass 2005). Their research found that high stakes testing is negatively associated with students going on to 12th grade. Tying rewards to testing is not an instant cure.

Assumption 4: High Stakes Testing is a Fair and Neutral Tool. As All Children are Required to Learn the Same Standards, All Children Have the Same Chances of Success

At the time of writing, the CAHSEE had been used as a graduation requirement for just two years. However, from the very beginning, we can see the differential results of the test. Taking Los Angeles county as an example, it becomes clear that students are passing at very different rates. Data, compiled from the 2007 test sittings, are shown in Table 1.

Troubling gaps already appear. It is easy to believe that because all students are covering the same content via state standards, all children have an equal chance of doing well on the CAHSEE. Nothing could be further from the truth. Today, California's urban schools suffer from physical deterioration and inadequate instructional materials, a lack of qualified teachers, and an unequal distribution of rigorous academic courses (Harris 2002). For example, a poll of public school teachers (Harris 2002) found that students in schools with the largest concentrations of low-income children are 12 times as likely to be taught by unqualified teachers and almost four times as likely to have serious turnover problems with their staff. Schools with higher numbers of students of color also tend to receive less funding. A UCLA report (Rogers et al. 2006) found that schools with greater that 90%

Table 1 Scores from the CAHSEE 2007, combined school year tally, 10th graders

	Asian pass rate (%)	Caucasian pass rate	Hispanic/Latino pass rate	African American pass rate	Total pass rate
Math	95	88	63	54	70
English	89	90	66	65	73

All figures taken from online Dataquest sources (2007)

students of color spent $6,634 per pupil, while schools with less than 49% students of color spent $7,268 per pupil. The authors argue that while all California schools are impacted by educational challenges, "all communities don't suffer equally. Schools with high concentrations of students of color, many who are poor and learning the English language, report the highest rates of unqualified teachers and shortages of college preparatory courses in the state" (p. 1). Only 30% of schools with the highest student of color enrollments offer sufficient college preparatory courses (Oakes et al. 2006).[3] These are the courses that must be passed in order for students to apply to University of California and California State universities. Such inadequacies do matter.

Having uniform state standards is not the great equalizer. Without the total package of inputs being equal, including qualified teachers, updated curriculum materials, and safe and adequate facilities, measuring outputs only will continue to put children of color at a deficit. While we are judging all children by the same yardstick, schools are not providing all children with the same tools. For example while Lisa, a student from affluent Pacific Palisades, and Maribel, a student from center-city Lynwood, will both take the CAHSEE, they will approach it with unequal schooling histories. Lisa is in an honors track, has access to updated books, credentialed teachers and small classes. Maribel, although academically talented, is unable to be in honors courses because her year round school does not offer one when she is on track. There are 40 students in most of her classes, 50% of her teachers are not credentialed and there are not enough textbooks for each student. Is it fair to hold students to the same outcomes, when they are not treated to the same inputs?

Due to the newness of the test as a graduation requirement, it is impossible to determine long term consequences of this policy. However, using analogous examples from other large school districts using high-stakes exit exams we can see what California's future may hold.

Low pass rates on high stakes tests plague urban schools. Haney's work in Texas (2000) shows that exams such as the TAAS continue to have an adverse impact on students of color. His study shows an increase in the drop out rate and retention rate for African American and Latino students as a result of the TAAS. A study of immediate outcomes (retention, drop out rates) of high-stakes testing in New York, Minnesota and Texas found racial disparities in all three states (Natriello and Pallas 2001). In Texas, researchers found approximately 25% of African Americans and Latino students are kept back in the ninth grade, so as not to take the TAAS (and possibly fail) in the tenth grade (Amrein and Berliner 2003).

Initial results in California mirror these trends. Rogers et al. (2005) collected data from the 2005 testing which shows that while 12% of students overall did not pass the CAHSEE English portion, 35% of English language learners did not pass. Their report also found that "schools where there are large numbers of students who have not passed the CAHSEE are also schools with fewer qualified teachers, overcrowding and multi-track schedules that limit learning time" (p. 1). Schools with low pass rates on the CAHSEE are three times more likely to be overcrowded,

[3] These are schools with 90–100% African American and Latino student enrollment.

four times more likely to have shortages of qualified teachers and two times more likely to have at least 50% of math classes taught by instructors who are not credentialed in math (Rogers et al. 2005). Such tests continue to punish children from the poorest of communities because they fail to learn in inferior schools (Orfield 2000). This is where the fair and neutral myth disintegrates. Poor schools are not getting a fair shake at keeping up with the high stakes game. They are starting with inferior and inadequate inputs, and then are chastised when their outputs do not seem to measure up.

Conclusions

"As someone who has spent his entire career doing research, writing and thinking about educational testing and assessment issues, I would like to conclude by summarizing a compelling case showing the major uses of tests for student and school accountability during the past 50 years have improved education and student learning in drastic ways. Unfortunately, that is not my conclusion"

(Linn 2000, p. 14)

While respected scholars in the field of educational measurement and major associations which specialize in educational research caution the use of high stakes testing, how is it that policymakers still do not flinch? Our myths about merit, ranking and sorting run deep. Assumptions about the use and consequences of testing, coupled with desires for a silver bullet, have led schools to a flawed policy response.

When considering school reform, the baseline question guiding decision-making should be, "Does this policy improve student achievement for all children?" To revisit, the original goals of the CAHSEE were to:

1. To ensure that a California diploma is grade appropriate
2. To increase student achievement
3. To close the achievement gap

Across all three goals, there is scant evidence that would suggest we are or could reach these goals through the use of a high-stakes exit exam. The very construct of the test does not require students to have a 12th grade education. When analogies are drawn to other urban districts, including New York and Texas, we see that there is little in the way of academic gains that transfer to other measurement instruments. California's own state data already shows a racial gap in student performance. Other states' data warns us of potential increases in drop out rates, retentions and a narrowing of the curriculum. The very gaps that this response sought to close are beginning to be realized.

By responding with a test as a way to improve public education, the CDE is relying on standardization to foment change in education. This system is predicated on the belief that if all inputs are similar (a standard curriculum), then the outcomes should also be similar (higher performance on tests). Testing and standards also follow market ideology in their belief of "best practices"—the notion that there is one best way to get a task done (Oakes and Lipton 2002). Thus, rather than design

curriculum to meet the needs of diverse students and communities, one fixed set of standards is applied to all students, with one fixed test to ensure these goals are reached. However, applying such models to educational contexts is deeply flawed. Reducing educational equity to an input/output dichotomy undercuts the complexity of the issue. The playing field will not be leveled through a quick fix approach that treats equity as if it is a one dimensional issue.

The CAHSEE also raises hard questions about the role of equity, efficiency and excellence in education. Stout et al. (1994) explains that the value of equity is in contest with values of excellence and efficiency. These values seem to be mutually exclusive in the present educational climate. High-stakes tests are an easy way to assess large numbers of children. Achieving equity however, is not easy. Allowing tests to serve as a proxy for true educational changes, without providing the appropriate funding, materials and resources, especially for children of color and poor children, ensures that equity will never be reached (Hilliard 1998).

Myths about assessments and accountability cloud our collective vision. The public has been sold accountability schemes as just another form of assessment. But they are not. Assessment is the stuff of teachers. It is what educators do everyday, in small and painless ways, to figure our where their students are and where they want them to be. It guides and shapes the teaching of particular students, at a particular time.

Accountability is the stuff of politicians. It is about sorting, rating and placing blame. Its influence on teaching *these* students at *this* time is purely detrimental, hindering teachers, constraining the curriculum and pushing out those whom most need to be included. Important questions remain. To what extent is the CAHSEE a response to political pressure or needs? To what extent are these accountability schemes truly about universal student achievement? The silver bullet of accountability has missed its mark. Rather than eliminating that which gets in the way of urban learners, it inflicts further wounds on an already injured system.

References

American Diploma Project. (2004). *Ready or Not: Creating a high school diploma that counts*. (National Report). Washington, DC: American Diploma Project.

American Educational Research Association. (2000). *AERA position statement concerning high stakes testing in PreK-12 education*. Available: http://www.aera.net/policyandprograms/?id=378.

American Evaluation Association. (2002). *American Evaluation Association position statement on high stakes testing in PreK-12 education.* Available: http://www.eval.org/hst3.htm.

American Psychological Association. (2001). *Appropriate use of high stakes testing in our nation's school*. Available: http://www.apa.org/pubinfo/testing.html.

Amrein, A., & Berliner, D. (2002a). *An analysis of some unintended and negative consequences of high stakes testing*. (Report No. EPSL-0211–125-EPRU). Tempe, Arizona: Education Policy Studies Laboratory.

Amrein, A., & Berliner, D. (2002b). High-stakes testing, uncertainty, and student learning. *Educational Policy Analysis Archives*, *10*(18).

Amrein, A., & Berliner, D. (2003). The effects of high stakes testing on student motivation and learning. *Educational Leadership, 60*(5), 32–38.

Au, W. (2007). High stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258–267.

Boaler, J. (2002). *Experiencing school mathematics*. New Jersey: Lawrence Erlbaum Associates.

California Business for Education Excellence. (2005). *Closing achievement gaps: The next phase in improving CA's public schools*. Available: http://www.cbee.org/PDFs/CBEE%20Reform%20Plan%202005%20-%20FINAL.pdf.

California Department of Education. (2006). *California high school exit exam program overview* Available: http://www.cde.ca.gov/ta/tg/hs/overview.asp.

California Department of Education. (2007). Spotlight on the CAHSEE: What is the CAHSEE? Available: http://www.cde.ca.gov/ta/tg/hs/documents/ whatiscahsee07.doc.

Caputo-Pearl, A., Clark, I., Dreebin, M., Foster, M., Guthrie, L., Macias, R., et al. (2003). *Report from members of the Task Force on Alternative Assessments*. Los Angeles, CA: UCLA Teaching to Change LA report.

Cuban, L. (2001). Why bad reforms won't give us good schools. *The American Prospect, 12*(1), 46–48.

Dataquest. (2007). *CAHSEE results by gender and ethnic designation, (combined 2007, Los Angeles country)*. Available: http://data1.cde.ca.gov/dataquest/cahsee/ExitEth2.asp?cSelect=19,LOS,ANGELES&cYear=200607&RptType=ExitEth2&cAdmin=C&tDate=000000&cGrade=10&Pageno=1.

Education Week. (2006). *Diplomas count: An essential guide to graduation policy and rates*. Available: www.edweek.org/ew/toc/2006/06/22/index.html.

Education Week. (2007). *Graduation briefs: California State information*. Available: http://www.edweek.org/media/ew/dc/2007/ca_SGB07.pdf.

Engel, M. (2000). *The struggle for control of public education: Market ideology vs. democratic values*. Philadelphia: Temple University Press.

Fusarelli, L. (2001). The political construction of accountability: When rhetoric meets reality. *Education and Urban Society, 33*(2), 157–169.

Gunzenhauser, M. G. (2003). High stakes testing and the default philosophy of education. *Theory into Practice, 42*(1), 51–58.

Haney, W. (2000). The myth of the Texas miracle in education. *Educational Policy Analysis Archives*, *8*(41).

Harris, L. (2002). *A survey of the status of equality in public education in California: A survey of a cross section of public school teachers.* Available:http://www.edfordemocracy.org/TQI/Harris%20Poll%20-%20Equality%20in%20Schools.pdf.

Herman, J. (2007). *Accountability and assessment: Is public interest in K-12 education being served*? (CRESST Report # 728). Los Angeles, CA: National Center for Research on Evaluation, Standards and Student Testing.

Heubert, J., & Hauser, R. (1999). *Testing for tracking, promotion and graduation*. Washington, DC: National Academy Press.

Hilliard, A. (1998). Standards: Decoy or quality control? *Rethinking Schools, 12*(4).

Hilliard, A. (2000). Excellence in education versus high stakes standardized testing. *Journal of Teacher Education, 51*(4), 293–304.

Jaeger, R. (1993). *Statistics*. Thousand Oaks, CA: Sage Publications.

Klein, S., Hamilton, L., McCaffrey, D., & Stecher, B. (2000). What do test scores in Texas tell us? *Education Policy Analysis Archives, 8*(49).

Laitsch, D. (2006). *Assessment, high stakes and alternative visions: Appropriate use of the right tools for leverage improvement*. (Report No. EPSL-0611–222-EPRU). Tempe, Arizona: Education Policy Research Unit.

Linn, R. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4–16.

Linn, R., Baker, E., & Betebenner, D. (2002). Accountability systems: Implications of requirements of the No Child Left Behind Act of 2001. *Educational Researcher, 31*(6), 3–16.

Madaus, G., & Clarke, M. (2001). The adverse impact of high stakes testing on minority students: Evidence from one hundred years of tests. In G. Orfield & M. Kornhaber (Eds.), *Raising standards or raising barriers* (pp. 85–106). New York: The Century Foundation Press.

Mathison, S., & Freeman, M. (2003). Constraining elementary teachers' work: Dilemmas and paradoxes created by state mandated tests. *Education Policy Analysis Archives, 11*(34).

Natriello, G., & Pallas, A. (2001). The development and impact of high stakes testing. In G. Orfield & M. Kornhaber (Eds.), *Raising standards or raising barriers* (pp. 19–38). New York: The Century Foundation Press.

National Center for Education Statistics. (2006). *The National Report Card: State Reading, California*. Available: http://nces.ed.gov/nationsreportcard/pdf/stt2005/2006452CA4.pdf.

Nichols, S., Glass, G., & Berliner, D. (2005). *High-stakes testing and student achievement: Problems for the no child left behind act*. (Report No. EPSL-0509–105-EPRU). Tempe, Arizona: Education Policy Research Unit.

Oakes, J., & Lipton, M. (1999). *Teaching to change the world*. Boston: McGraw-Hill.

Oakes, J., & Lipton, M. (2002). *Teaching to change the world*. New York, NY: McGraw-Hill.

Oakes, J., Rogers, J., Silver, D., Valladares, S., Terriquez, V., McDonough, P., et al. (2006). *Removing the roadblocks: Fair college opportunities for all California students*. Los Angeles, CA: UC ACCORD/UCLA IDEA report.

Orfield, G. (2000). Policy and equity: Lessons of a third of a century of educational reform in the US. In F. Reimer (Ed.), *Unequal schools, unequal chances: The challenges to equal opportunity in the Americas* (pp. 400–429). Cambridge, MA: Harvard University Press.

Rogers, J., Holme, J., & Silver, D. (2005). *More questions than answers: CAHSEE results, opportunities to learn and the class of 2006*. Los Angeles, CA: IDEA report.

Rogers, J., Terriquez, V., Valladares, S., & Oakes, J. (2006). *California educational opportunity report 2006: Roadblocks to college*. Los Angeles, CA: UC ACCORD/UCLA IDEA Report.

Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do about it*. Cambridge, MA: Perseus.

Sheldon, K., & Biddle, B. (1998). Standards, accountability, and school reform: Perils and pitfalls. *Teachers College Record, 100*(1), 164–180.

Smith, M., & Fey, P. (2000). Validity and accountability in high stakes testing. *Journal of Teacher Education, 51*(5), 334–344.

Stout, R., Telluric, M., & Scribner, J. (1994). Values: The what of the politics of education. In D. Layton & J. Scribner (Eds.), *The study of educational politics: The 1994 commemorative yearbook of the politics of education association (1969–1994)*. Washington, DC: Falmer.

The Business Roundtable. (1998). *Building support for tests that count*. Washington, DC: The Business Roundtable.