

5-1-2005

## High-stakes Testing and Assessment: One is not the other

Enrique Murillo

*California State University San Bernardino*

Alayne Sullivan

*California State University San Bernardino*

Follow this and additional works at: <https://scholarworks.lib.csusb.edu/wie>



Part of the [Bilingual, Multilingual, and Multicultural Education Commons](#), [Curriculum and Instruction Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), [Educational Psychology Commons](#), [Elementary Education and Teaching Commons](#), [Junior High, Intermediate, Middle School Education and Teaching Commons](#), [Secondary Education and Teaching Commons](#), and the [Social and Philosophical Foundations of Education Commons](#)

---

### Recommended Citation

Murillo, Enrique and Sullivan, Alayne (2005) "High-stakes Testing and Assessment: One is not the other," *Wisdom in Education*: Vol. 1 : Iss. 1 , Article 4.

Available at: <https://scholarworks.lib.csusb.edu/wie/vol1/iss1/4>

This Article is brought to you for free and open access by CSUSB ScholarWorks. It has been accepted for inclusion in *Wisdom in Education* by an authorized editor of CSUSB ScholarWorks. For more information, please contact [scholarworks@csusb.edu](mailto:scholarworks@csusb.edu).

---

## High-stakes Testing and Assessment: One is not the other

### Abstract

Since the institution of the common school and the advent of universal education, Americans have placed tremendous faith in public schools. Public education cultivates an informed citizenry, one of the pillars of a liberal democracy. But more importantly, schools are a repository for our common dreams of human potential and individual self-actualization. Because they so thoroughly shape the lives and life-chances of our youth, school issues are freighted with an emotional charge. Education remains the last fully public American institution, one in which millions of students cast their common lot daily and strive to become better readers, better citizens, better people. With these premises as critical background, we have undertaken the project guiding this paper: to invite readers to think deeply about the impact of high-stakes testing on the school-aged citizens of this country *especially* in light of the core elements of what assessment ought to be about for students. We want to prompt readers to pause, to take note of some of the key threads in the professional conversation about assessment these days, and to then respond accordingly.

### Keywords

testing, curriculum, standardized tests, high stakes tests

### Author Statement

Dr Enrique Murillo and Dr Alayne Sullivan are professors in the College of Education, California State University San Bernardino.

## High-Stakes Testing and Assessment: One is Not the Other

Enrique Murillo and Alayne Sullivan

Department of Language, Literacy and Culture  
California State University, San Bernardino

### Introduction

Since the institution of the common school and the advent of universal education, Americans have placed tremendous faith in public schools. Public education cultivates an informed citizenry, one of the pillars of a liberal democracy (Delli Carpini and Keeter, 1996, and Nie, Junn and Stehlik-Barry, 1996). But more importantly, schools are a repository for our common dreams of human potential and individual self-actualization. Because they so thoroughly shape the lives and life-chances of our youth, school issues are freighted with an emotional charge. Education remains the last fully public American institution, one in which millions of students cast their common lot daily and strive to become better readers, better citizens, better people. With these premises as critical background, we have undertaken the project guiding this paper: to invite readers to think deeply about the impact of high-stakes testing on the school-aged citizens of this country *especially* in light of the core elements of what assessment ought to be about for students. We want to prompt readers to pause, to take note of some of the key threads in the professional conversation about assessment these days, and to then respond accordingly. We pose a brief range of questions for readers:

1. Is high-stakes testing and the reliance on standardized test scores the best way to assess student and teacher performance?
2. How are high-stakes test-scores being interpreted and used?
3. Is it possible that the enforcement of high-stakes testing on English Language Learners contributes to their systematic disenfranchisement from the public-school promise for a better life for all?
4. **AND/OR**, is the state-initiated urgency to raise test scores also raising our collective levels of awareness about the health of education, schooling, instruction and learning in California?

Ultimately, the premise guiding this article is that assessment and high-stakes testing are two very different phenomenon — that the underlying principles of sound assessment are badly contradicted by the practices and consequences of high-stakes testing. This position is developed in the pages that follow with careful regard for what literacy assessment ought to be about and how high-stakes testing practices and consequences stand in this light.

### Assessment Broadly Speaking: Describing and Improving Student Performance

Grant Wiggins writes that “assessment should be deliberately designed to improve and educate student performance, not merely to audit it as most school tests currently do” (1998, p. xi). Wiggins has written widely on school assessment. As the president and director of programs for CLASS (Center on Learning, Assessment, and School Structure), his stature in the field is well known. In *Educative Assessment: Designing Assessments to Inform and Improve Student Performance* (1998), Wiggins sets forth a comprehensive rationale for the importance of authentic versus “busy” tasks, useful assessment/feedback, and the guided expectation that assessment-feedback be implemented. Conventional, and particularly high-stakes tests, cannot provide the kind and degree of feedback demanded by authentic assessment. Wiggins and others suggest that educators must learn to ask “what is evidence of learning and how will I teach toward it?” rather than “What will I cover and have my students do?” as many usually ask themselves. These thoughts are consistent with those of Johnston (1997) who teaches us that the very word “assessment” derives from the Latin word *assidere*, meaning to sit alongside. The principle of feedback, both writers claim, goes hand-in-hand with accountability! A final thought from Wiggins puts the accountability factor in some perspective: We will have real, and not simply political accountability, when we are required to seek and use feedback as part of our daily practice. Teachers as well as students must come to recognize that seeking and receiving feedback on their performance is in their interest, and different from their previous experience with so-called supervision and evaluation, which has too commonly been counterproductive or useless (1998, p. xv).

Glanced at next, the work of W. J. Popham (2001; 1999) and J. H. McMillan (2001) extends that of Wiggins’, and is also widely referred to in the field of assessment. Popham’s (1999) text, *Classroom Assessment: What Teachers Need to Know*, and McMillan’s (2001), *Classroom Assessment: Principles and Practice for Effective Instruction*, both emphasize assessment as a systematic way to “get a fix on a student’s status” as Popham phrases it (1999, p. 3). In 1999, Popham, like many others, writes that teachers need to know how to construct and evaluate classroom assessments, interpret standardized tests, and guide instructional decisions. McMillan holds that assessment principles ought to be grounded on an instructional bedrock: “establishing credible performance standards, communicating the standards to students, and providing feedback to students on their progress (2001, p. xiii). Given the bent of this article, it is important to note that these authors, along with many others writing about assessment, explain the use and importance of standardized tests. Since high-stakes tests are almost always standardized, this point about their use and importance is highly relevant. Statistical details aside, McMillan offers a very helpful point of perspective about standardized tests when he writes the following: Standardized test results, when used appropriately, can provide helpful information for instructional planning. The key is being able to understand the scores that are reported as well as the limitations on how scores should be interpreted. As long as results from these tests are not used as the sole criterion, scores can be used to form conclusions about the ability or prior achievement of students.

[R]esults from these tests should always be used with other types of evidence, such as teacher observation and classroom assessments (2001, p. 84).

Three highly salient points from this just-cited comment lead us straight into the thick of the debate about high-stakes, standardized testing: (1) there are limitations on how high-stakes, standardized testing should be interpreted; (2) results from such tests should not be the sole criterion for educational decision-making; and (3) educational decisions should always be based on teacher observation as well as classroom assessments! It is both ironic and reassuring for those of us who have deep worries about standardized, high-stakes testing that one of the people who has been a trusted source of explanation about assessment benefits (W. James Popham) should also be among those most loudly warning us about danger on the horizon.

In 2001, Popham published *The Truth About Testing: An Educator's Call to Action* in which he begins point-blank by saying “I believe that today’s high-stakes tests, as they are used in most settings, are doing serious educational harm to children. Because of unsound high-stakes testing program, many students are receiving educational experiences that are far less effective than they would have been if such programs had never been born” (2001, p. 1). There is a great deal of kindly intended and wise advice to educators (across the board) in the Popham text. Sensibly, he cautions teachers, as well all who make decisions based on assessment outcomes, to make sure that a very modest number of classroom tests are those of the standardized and high-stakes variety and that they measure learner outcomes of indisputable importance. Further, he urges us to use diverse assessments and emphasize student responses to these as central to instructional decision-making. Even more, we should be, according to him, assessing student affect! He says, “[I]n a fevered quest for positive evidence about instructional quality, it’s easy to forget the reason we ought to be evaluating instruction in the first place: to help kids learn better” (2001, p. 146). *The Truth About Testing* contains specific tips on assessment in the wake of high-stakes tsunami (i.e. tidal wave). For example, literacy assessment programs need to be provided for teachers, administrators, policy-makers, and parents. Parent-action groups can be established as well as a public information campaign for all invested in our educational goals for students. Like Wiggins, Popham insists on an accountability factor — with a proviso: “I see nothing wrong with accountability-oriented programs for the evaluation of schooling if these evaluations incorporate appropriate kinds of evidence. It would be... senseless to reject standardized achievement testing... without replacing it with a new evaluative program that’s based on the right data” (2001, pp. 157 & 158). In the next section of the article, we discuss what high-stakes tests are more fully, contrasting them with the assessment elements just presented.

### **So... What is a High-Stakes Test?**

Around the nation, “fill-in-the-bubble” tests have typically come to dominate student and school assessment. These multiple-choice tests can either be norm-referenced (which deal with the relative relationships of an individual test-taker’s performance in relation to the performance of others) or criterion-referenced (which deal with a set standard or level of performance below which a test-taker

is not considered to have met the standard). Sometimes several tests (of both types) are used in combination, or are taken repeatedly, to assess performance. The development of items for norm-referenced and criterion-referenced tests is quite different. Although inaccurate to assume that these multiple-choice tests dominating assessment are strictly norm-referenced, most characteristically they are. The raw scores are standardized or “normed” against the scores of a population of test-takers who first took the test as it was being designed. The means, standard deviations, percentiles, and standardized scores of subsequent individuals taking the test are then evaluated and compared to those of this first group used as the norm even though the demographic and socio-cultural character of the norm group of test-takers may be completely unlike that of the group subsequently taking the test.

Individual test items are purposefully designed to have a level of difficulty that will be answered correctly by a certain percentage of respondents and not others. Test items yielding too high or low a percentage of correct answers on test items are eliminated or redesigned by test makers so that the overall effect of the instrument works as a screening device to sort students’ scores into a range of score groups. Although not always, most typically a normal distribution is divided into nine intervals, each called a stanine (*stan*-dard *nine*). Simply put, stanines are one-digit normalized scores. Most often students receive middle scores (4, 5, and 6), fewer receive more extreme scores (2, 3, 7, and 8), and even fewer receive the extreme scores (1 and 9). Students’ scores are interpreted as weak or strong given not only the raw score itself but which stanine group the score falls into (the 9th stanine is considered a very strong score, the 1st very low). Educators are pushing for higher stanine scores — that is for students’ scores to fall into a higher category — as well as to raise the raw scores themselves.

Standardized tests are problematic for several reasons. Norm-referenced tests cover as little as 50% of textbook content, much less the much broader curriculum content (Freeman and Hannan, 1983). Standardized tests measure only a very limited range of skills, in contrast to what is taught in the classroom. Multiple-choice questions treat learning as memorizing isolated pieces of information, rules, and procedures. That instruction in test taking skills (something as simple as practicing filling in the bubble sheets) raises scores reveals how many factors outside of content knowledge determine outcomes. Further, research on test bias clearly reflects the race, class, and gender advantage afforded some groups in such situations. In fact several educational historians have linked the underpinnings of standardized testing to the racist eugenics movement (see Sack, 1999).

Whether or not one views this claim as extreme or not, it is at the least recommended that we are all mindful Amrein’s and Berliner’s findings regarding pass rates for African Americans, Hispanics and children of poverty, in *High-Stakes Testing, Uncertainty, and Student Learning* (2002). They are clearly much lower in all states implementing high-stakes testing practices, high-stakes exit exams “disproportionately affect students from lower socioeconomic backgrounds” (Amrein & Berliner, 2002, p. 8), and high-stakes tests “will have greater consequences for America’s poorest children” (p. 8).

Relying on single-test (or a mix of tests), standardized scores, administrators and policy makers make critical decisions about students', teachers' and administrators' lives. When "high-stakes" are attached to these tests, such as making decisions about graduation or grade promotion, programmatic sanctions and financial rewards, or publicly reporting school scores (in a "reform by shame" manner), the results carry serious consequences for students, educators, and schools. Therefore, high-stakes testing refers to the use of assessment data to make decisions about enrollment, retention, promotion, incentives for children or teachers, or other tangible rewards or punishments. Decisions are also made about tracking (assigning students to schools, programs, or classes based on their achievement level), whether a student will be promoted to the next grade, which school a student may attend, and whether a student will receive a high-school diploma. In many cases, teachers' and principals' salaries, work load and job stability are tied to standardized test scores. These "high stakes" have forced teachers to waste instructional time on the transmission of limited, test-taking skills, or simply "teaching to the test." Yet each year the dedication of untold sums of money and amounts of time to these testing programs is growing.

High stakes in California include the Academic Performance Index (API) rankings on which public schools receive significant financial rewards or sanctions. The API rankings consist of the SAT-9 academic achievement test, which is norm-referenced, and the California Standards Test, which is criterion-referenced. High stakes also include the High School Exit Exam, since the award of a high school diploma may rest on a student's performance, as well as the Golden State Exam program, since scholarship monies are awarded on student performance. In California, there is not one single high stakes test, but a whole bevy of them. Oftentimes, many of these tests are taken multiple times. We next offer a brief historical background as well as recent work that considers high-stakes testing; the perspectives represent a broad range of the populace, media, scholarly, governmental, and professional educational organizations.

### **Assessment and Accountability in Context**

Over the past several years, the public has become convinced that there is a crisis in education and that a responsible or at least immediate response is to focus on high-stakes, standardized test results to tell us how students are doing but, far more importantly, to raise those test scores. That way, it can be demonstrated to the public and to politicians that students are doing what they need to be doing better (learning more) and teachers are doing what they need to be doing more (teaching students better). Or so it would seem. Once we have an even closer look at a range of specific concerns about high-stakes testing, we may begin to appreciate and enlarge upon a remark made earlier in this paper: that there really are limitations on how high-stakes tests should be used and interpreted, that they should never be the sole criterion for a big decision and that teacher's assessments should have far more to do with educational decisions.

Briefly, the prevailing sense that there's a crisis in education is believed to have first been "manufactured" from the Reagan White House's National Commission on Excellence (federally

appointed in 1981, and established by U.S. Secretary of Education Bell), when it issued its 1983 report, *A Nation at Risk: The Imperative for Educational Reform* (Berliner & Biddle, 1997). The wide circulation of the publication boosted the appearance of this crisis in the public schools. With this report, came others who similarly put into question the quality of schools, observing the decline of both test scores and enrollments in mathematics and science, and citing business leaders' dissatisfactions with students' poor skills and academic measures when compared to Japanese and European counterparts. The *Nation At Risk* report, relying on SAT scores (and more anecdotal evidence), argued that the United States' economic competitiveness was threatened by the public schools' "rising tide of mediocrity" (1983, p. 5).

The initial impact of this report along with the results of the 1994 NAEP (National Assessment of Educational Progress) tests amassed to produce our current rash of high-stakes accountability programs and the absolute over-reliance on standardized testing as the arbiter of both student and teacher performance. By 1982, thirty-six states have had some kind of student testing program (Odden and Dougherty, 1982; Amrein & Berliner, 2002).

### **Professional Stances and Critiques of High-Stakes Testing**

The *American Education Research Association* (AERA) has prepared a position statement on high-stakes testing based on the 1999 "Standards for Educational and Psychological Testing." Representing a professional consensus about sound and appropriate test use in education and psychology, as well as sponsored and endorsed by the *American Psychological Association* (APA) and the *National Council on Measurement in Education* (NCME), AERA says this:

[I]f high-stakes testing programs are implemented in circumstances where educational resources are inadequate or where tests lack sufficient reliability or validity *for their intended purposes* [emphasis added], there is potential for serious harm. Policy makers and the public may be misled by spurious test scores increases unrelated to any fundamental educational improvement; students may be placed at risk of educational failure and dropping out; teachers may be blamed or punished for inequitable resources over which they have no control; and curriculum and instruction may be severely distorted if high tests scores per se, rather than learning, become the overriding goal of classroom instruction (AERA, 2000).

Further, AERA sets forth a set of conditions, adopted in July of 2000, essential to sound implementation of high-stakes educational testing programs which include protection against high-stakes decisions based on a single test; that there be adequate resources and opportunities to learn; and that there be full disclosure of likely negative consequences of such testing programs. High-stakes tests must be representative of the curriculum; passing scores and achievement levels measured on such tests must be established; and meaningful remediation and consideration of students with language issues and special needs students must be considered. The reliability and validity for each intended use of test results must be separately established. Thus, cautions about the use and interpretation of high-stakes tests are made quite clear here.



In looking at high-stakes testing in light of high-school completion statistics, *The National Board on Educational Testing and Public Policy* (NBETPP) states that “[i]t’s a bull market for high-stakes testing that far surpasses the rush of the early 1970’s to test minimal competency” (Clarke, M. et al, 2000, p. 1). Their conclusion about high-stakes testing programs, based on correlational evidence, is that they are linked to decreased rates of high school completion. Of the twenty states included in a published study, “[t]he states with the highest dropout rates had MCT (Minimal Competency Test) programs with standards set at least in part by the state” (Clarke, M. et al, 2000, p. 2). In schools with proportionately more students of low socio-economic status, early dropout rates were 4 to 6 percentage points higher in schools that had a high-stakes test requirement. This study points to the effects of high-stakes testing on both dropout rates AND rates at which students are retained in various grades, which raises further worry about students’ sense of academic worth in such contexts. In commenting on standards, low achievement and high-stakes testing, *The National Education Association* (NEA) asserts that in using high-stakes tests to measure policymakers’ standards in schools designated as “low performing,” that uneven and limited understanding and participation by school staff, parents and community often follows. In regarding high-stakes sanctions they state, “[a]ccountability systems that focus primarily on the individual performance of school staff or other stakeholders, rather than the whole school and its goals, do not meet NEA’s criteria for fairness and equity” (NEA, 2001). The *Center for Research on Evaluation, Standards, and Student Testing* (CRESST) has also formulated a position on this topic that is consistent with all of the above-cited views (Phelps, 2001).

The now nationally recognized position against high-stakes testing by Alfie Kohn was most recently summarized on-line, which outlines “five fatal flaws” (Harris, 2001). These flaws include the outlook that the “rewards and punishment” effects of such testing tends to undermine long-term quality education; that the whole standards-movement link to testing confuses “harder” with “better”; and that the “back to basics” wing associated with high-stakes testing overlooks fundamental learning principles. Of greatest concern is the tendency for learning, curriculum, and assessment to become focused on finite rules, facts, and short-term memory processing and test-taking skills thereby curtailing emphasis on higher-order thinking and creativity.

Parent groups are listening: the *National Parent Information Network* (NPIN) worries about how high-stakes, standardized tests can shape and narrow students’ curriculum; how they affect students’ motivation; and how they can result in inappropriate remedial help (Patten, 2000). Citing the *American Federation of Teachers*’ (AFT) concerns, NPIN recently pointed to the high-stakes testing influence on students for whom English is not their first language and on curricular materials chosen to most align with test(s) content (Robertson, 2000).

The above-mentioned statements set forth a broad, credible and unequivocal note of caution about high-stakes testing. A critical element in this portfolio of concern warrants special attention. English Language Learners are affected by all of these factors in a way that is probably not fully appreciated even by the caring educators who design, pass out and enforce this testing franchise. Perhaps these

so-far-mentioned points are appreciable from the perspective of those of us who can and do voice our fears and expect to have them heard and responded to. Those who are most often silenced, most often disempowered, are the millions whose linguistic identity is not only left behind at the school door, but who are continually barraged by the rigidities of high-stakes tests. Those who bring to school a cultural and linguistic heritage that can be tacitly plumbed and aligned with the formatting, language, and literacy particulars of these tests have advantages that many others do not. English Language Learners, in fact, have a depth of life and language experiences that are shunned, if not actively negated, by the process, content, and emotional impact of high-stakes tests. In California, about 25% (close to 1.5 million) fall into this category. English proficiency is directly related to test scores, placing them at a substantial disadvantage to English proficient peers. We now set forth a summary of positions on high-stakes testing formulated by literacy organizations with an embedded regard for how the concerns raised impact English Language Learners.

The following statements require no summary or introduction: “The *International Reading Association* [IRA, italics added] strongly opposes high-stakes testing” and “[t]he Association believes that important conceptual, practical, and ethical issues must be considered by those who are responsible for designing and implementing testing programs” (IRA, 1999). IRA recommends that teachers implement rigorous classroom-based assessments that will inform parents, community and policy makers about various levels of literacy processes. They also suggest that parents question the effects of high-stakes testing on their children and that they lobby for ongoing, process-based classroom assessment that will reflect and improve literacy instruction over extensive periods of time. Finally, policy makers should respect multiple measures of assessments, that reflect the complexity of literacy while also avoiding the extrinsics of incentives, money, or publication of test scores that reward and/or punish schools and teachers.

Along related lines, *The National Council of Teachers of English* (NCTE) recently passed resolutions urging reconsideration of high-stakes testing and asking for the development of a test-taker’s bill of rights. The Council comments that the use of high-stakes tests “has continued to escalate and to cause measurable damage to teaching and learning in U.S. schools” (NCTE, 2000). NCTE claims that high-stakes tests “often fail to assess accurately students’ knowledge, understanding, and capability” and that this testing “harms students’ daily experience of learning, displaces more thoughtful and creative curriculum, diminishes the emotional well-being of educators and children, and unfairly damages the life-chances of members of vulnerable groups” (NCTE, 2000). Among the many NCTE resolutions urging caution about high-stakes testing are those that preserve the right to experience a challenging curriculum that is not constrained by any test; the right to display competencies through various means; and the right to know how the results of the test will be used. More broadly, the *Center for the Improvement of Early Reading Achievement’s* (CIERA) 1998 report, spurred by the 1989 National Education Summit which resolved that “all children in America will start school ready to learn,” focuses on comprehensive efforts to define and assess young children’s readiness for early school experiences. Of pivotal concern is the use of high-stakes readiness

assessments to track, label and retain children in kindergarten. Condemnatory reports about these assessments have been issued by the *National Association for the Education of Young Children* (NAEYC, 1988; 1990), the *National Association of Elementary School Principals* (NAESP, 1990), the *National Association of Early Childhood Specialists in State Departments of Education* (NAECS, 1987), the *National Association of State Boards of Education* (NASBE, 1988; 1991) and the *National Commission on Children* (1992). In combination with all of the work previously summarized, CIERA's synthesis of these positions seems worth quoting at length:

The principal message of these reports was that the methods, materials, and logic of educating older students should not be imposed on young children. The policies that were criticized were those that increased attention to academic outcomes at the expense of children's exploration, discovery and play; methods that focused on large group activities and completion of one-dimensional worksheets and workbooks in place of actual engagement with concrete objects and naturally occurring experiences of the world; and directives that emphasized the use of group-administered, computer-scored, multiple-choice achievement tests in order to determine a child's starting place in school rather than assessments that rely on active child engagement, teacher judgement, and clinical opinion (Meisels, 1998).

CIERA's summary position raises issues that are applicable to all involved with the school population at-large, not just young children. For example, in recognizing individual learning differences and developmental variability, it is stated that "[i]t will never be the case that all children will attain the same level of performance at a single culturally defined point in time" (Meisels, 1998). Questions are raised within this report about high-stakes readiness assessments: can readiness be assessed without the harm of label or stigma?; will early-school programs teach to the standardized test?; and if "non-readiness" is determined, will this be viewed as a "problem" in the child or the community?

While respectful of readiness definitions varying from the (a) idealist/nativist (with emphasis on the internal dynamics of the child and distancing broad environmental factors), to the (b) empiricist/environmental to the (c) social/constructivist (stressing sociocultural influences of community, family and school), to (d) an interactionist perspective (with focus on child's learning process and many environmental factors), the report makes three conclusions. First of all, "no single assessment will solve all of our educational needs or solve all of our educational problems" (Meisels, 1998). Secondly, the negative impact of early childhood high-stakes assessments includes retention, special education programs, and parents holding their children out of school — all of which have harmful consequences. Thirdly, the whole culture of high-stakes readiness tests encourages an inclination to consider it as a problem to be eradicated rather than a process that occurs over time in supportive contexts.

This slightly protracted discussion of high-stakes readiness assessment is important. It focuses a level of awareness about the whole relevance of the developmental and sociocultural context. This

issue is paramount if we really believe that education ought to foster constant evolution of learning for students of all ages across all subject areas.

The *National Council of Teachers of Mathematics* (NCTM) has recently published a statement about high-stakes testing that echoes many of the perspectives already summarized. Just as other groups' recommendations, their's focus on the need for multiple and appropriate sources of assessment; that it should be an open process with everyone knowing what is expected; and that all aspects of mathematical knowledge and its connections should be assessed.

With all of the above ideas in mind, we now pose final thoughts in glancing at some government-associated platforms, values, and activity. A November, 2000 statement published in *Education Week* by William L. Taylor raises a fair-minded comment about the link of civil rights and high-stakes testing. Serving as counsel in the drafting of comprehensive reforms to the Title 1 program (an \$8 billion federal program of aid to disadvantaged students), and vice-chair of the *Leadership Conference of Civil Rights* and the *Citizens' Commission on Civil Rights*, Taylor refers to the recognition by Congress of the rights of all Americans to a quality education. He further articulates the legislative link thereby forged in the form of Title 1. The resultant imperative was to set higher standards, and devise "accountability" systems measured through high-stakes testing. The ultimate consideration, he comments, ought to be whether or not the demand for accountability is accompanied by "tests that are fair, nondiscriminatory, and representative of what teachers cover in class" (Taylor, 2000).

The general hubbub over testing, Taylor suggests, overlooks the differences between good and bad tests. "Fill-in-the-bubble" tests tell us little about what children are learning while others are effective means of measuring higher standards. "Taking a blunderbuss to all testing would forfeit the opportunity to use good tests for diagnostic purposes, and as one means to improve instruction" (Taylor, 2000). Unfortunately the U.S. Department of Education (USDOE) currently has no plans to call upon states to verify that tests are of high quality and do not include questions outside the curriculum taught in schools). Perhaps we all ought to work toward voicing a conversation about how to forward the civil rights of all students by setting high standards in light of a sound curriculum that is soundly assessed.

A recent joint report of the *National Academy of Sciences* (NAC) and the USDOE, and approved by the Governing Board of the *National Research Council* (NRC) ought to be regarded in light of the position expressed by Taylor in the preceding paragraph. The report was undertaken to determine appropriate test use and offer recommendations about high-stakes testing issues. Three principal criteria are submitted in deciding on appropriate test use: (1) the **validity** for a particular purpose — whether it measures a test taker's knowledge in the content area being tested; (2) **attribution of cause** — whether a student's performance reflects knowledge & skill related to instruction, or language barriers or disabilities unrelated to skills being taught; and (3) **effectiveness of treatment or intervention** — whether test results lead to educationally beneficial responses.

Commenting how "lower achievement test scores of racial and ethnic minorities . and low income families reflect persistent inequalities in American society and not inalterable realities about these

groups of students,” various recommendations are offered about appropriate test use (Heubert and Hauser, 2001). The NAC/USDOE’s recommendations about appropriate test use include the condition that accountability ought to be shared by states, school districts, public officials, educators, parents and students, and that higher standards cannot be achieved merely by imposing them. Secondly, tests should be used for high-stakes decisions only after implementing changes in teaching and curriculum over a period of years AND that high-stakes testing decisions be considered only in combination with grades, developmental factors, attendance and teacher recommendations. Further, a caution is raised about test scores that are invalidated by teaching so narrowly to the objectives of particular tests that no accompanying improvement is made to a broader set of academic and social skills.

With all these thoughts in mind, we consider another reality that must be acknowledged according to the Fordham report published in January of 1999:

[I]n the real world, testing will continue. Testing experts have much to contribute to efforts to ensure that testing is done well. Unfortunately many of them share an ideological orientation that makes any type of standardized test impossible to swallow. Until these experts reexamine their most fundamental beliefs about teaching and learning, all the hard work of improving standardized tests will have to be done without them (Phelps, 1999).

The report presents almost all of the points that have been raised within this document, additionally noting that the U.S. is the most tested country in the world. It also casts a helpful historical perspective on how the *National Assessment of Educational Progress* (NAEP) — the only one national standardized test of general academic and literacy abilities — became associated with state-by-state, “standards-based” reporting of its scores into “basic”, “proficient”, and “advanced” categories in 1988. NAEP’S 1992/1994 scores comparison (with 1994 scores somewhat lower than 1992) has been associated with California’s embrace of high-stakes testing as a means of leveraging accountability for teachers’ work with basic literacy — primarily Kindergarten to grade 3 reading — skills.

What are now widely known as the “Reading Wars” (phonics vs. whole language debate) is linked to the low scores California received on the 1994 NAEP test relative to other states. Though the test had no substantive focus on phonics/whole language at all, and though many demographic, as well as test design and administration elements were far more likely to have explained California’s 1994 NAEP results, California has brought down many high-stakes testing and curricular realizations in the wake of those scores. “Ready or not, the High School Exit Exam will be administered to ninth-graders this spring” for example (Posnick-Goodwin, 2000). Along with the SAT-9 and Golden State Exams, this test is worrying teachers for all of the reasons already mentioned, according to recent issues of the *California Educator*.

A most recent exhaustive study released by the California Teachers Association (CTA), comparing various indicators of the lowest- and the highest-ranking schools under the state’s API, raises very serious questions about the effectiveness of current strategies from the Legislature at improving

student performance (California Educator, 2001). Several areas appear to correlate to a school's standing on the Index, including poverty, language barriers, teacher training, ethnicity, school calendar and school size.

If the Fordham Report is correct and high-stakes testing is here to stay and is supported by many more than oppose it, how are we to engage in productive dialogue about this contentiously accepted reality?

### Closing Remarks

Certainly we do need to look at how tests are designed, interpreted, and responded to along curricular and instruction lines. We need to keep designing alternatives to testing, as many of our ranks at the CSU already have. We need to better contribute to the public discourse and decision-making. We need to look at all the levels of complexity that prevail, keeping the educational interests of all students in mind. And lest we lose perspective entirely, we also need to be aware and respectful of the fact that many noted assessment experts have offered us sage counsel about the realistic benefits of standardized testing.

We have gone into considerable depth in presenting this discussion topic because we want to avoid a quick, even overly-reflexive response hoping for a level of discussion relative to the complexity of the topic. If there are dangers associated with high-stakes testing, how can we openly and productively talk about them? As teacher-educators and professionals associated with the educational realities of school-aged students, can we begin a dialogic forum that may lead to refinement of perspective on the high-stakes, standardized-testing reality? Clearly, there is widespread concern and organized response to the high-stakes testing movement; much of it has been documented within this article. How might we begin to responsibly address this topic with an eye to maintaining the benefits and advantages that standardized testing offers, while minimizing — no reducing dramatically — the dangers associated with high-stakes testing?

### References

- American Educational Research Association. (2000, July). *AERA Position Statement Concerning High-stakes Testing in PreK-12 Education* [Policies]. Washington, DC: Author. Retrieved December 6, 2000, from the World Wide Web:  
<http://www.area.net/about/policy/stakes.htm>
- Apple, Michael W. (1985). *Education and Power*. Boston, London & Henley: Ark.
- Apple, Michael W. (1989). American Realities: Poverty, Economy and Education. In L. Weis, E. Farrar, & H. Petrie (eds.) *Dropouts from School: Issues, Dilemmas, and Solutions*. Albany: State University of New York, pp. 205-225.
- Berliner, D.C. & Biddle, B.J. (1997). *The Manufactured Crisis: Myths, Fraud, and the Attack on America's Public Schools*. New York: Longman.
- Bracey, Gerald W. (1995). *Final Exam: A Study of the Perpetual Scrutiny of American Education*. Bloomington: Technos Press.

- California Educator* (2001). California Teachers Association. Volume 5, Issue 8.
- Center for Education Reform (1997). *School Reform in the United States: State by State Summary*. Washington, D.C.: Center for Education Reform.
- Clark, M., Haney, W., Madaus, G. (2000, January). High-stakes Testing and High School Completion. *National Board on Educational Testing and Public Policy Statements*, Vol. 1, number 3. Retrieved January 21, 2001, from the World Wide Web:  
<http://nbetpp.bc.edu/statements/V1N3.pdf>
- Delli Carpini, M.X. and Keeter, S. (1996). *What Americans Know About Politics and Why It Matters*. New Haven: Yale University Press.
- Elam, Stanley M. (September 1996). The 28th Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools. *Phi Delta Kappan*, vol. 78, no. 1, pp. 41-59.
- Finn, Chester & Rebarber, Theodor (1992). *Education Reform in the 90s*. New York: Macmillan Publishing.
- Firestone, William (1995). The States and Educational Reform. In Noblit, George & Pink, William (eds.) *Continuity and Contradiction: The Futures of the Sociology of Education*. Cresskill, NJ: Hampton Press, pp. 255\_278.
- Freeman, J. and Hannan, M.T (1983). *School District Demography and Governance*. Institute for Research on Educational Finance and Governance, School of Education, Stanford University.
- Gelberg, Denise (1997). *The "Business" of Reforming American Schools*. Albany: State University of New York Press.
- Harris, J. (2000). *Author Takes on High-Stakes Tests*. Columbus, OH: Eisenhower National Clearinghouse. Retrieved January 25, 2001, from the World Wide Web:  
<http://www.enc.org/print/topics/assessment/testing/documents/o,1946,FOC-001576-index,00.shtm>
- Heubert, J. P. and Hauser, R. M. eds. (1999) *High-stakes: Testing for Tracking, Promotion, and Graduation* [Committee Report]. Washington, DC: Committee on Appropriate Test Use, Washington, DC: National Research Council. Retrieved January 25, 2001, from the World Wide Web:  
<http://www.nap.edu/html/highstakes/>
- House, Ernest R.. (1998). *Schools for Sale: Why Free Market Policies Won't Improve America's Schools, and What Will*. NY: Teachers College Press.
- International Reading Association. (1999, August). *High-stakes Testing* [Position Statement]. Newark, DE: Author. Retrieved January 24, 2001, from the World Wide Web:  
[http://www.reading.org/advocacy/policies/high\\_stakes.html](http://www.reading.org/advocacy/policies/high_stakes.html)
- Lunenberg, Fred C. (November 1992). Introduction: The Current Educational Reform Movement — History, Progress To Date, and the Future. *Education and Urban Society*, vol. 25, no. 1, pp. 3-17.
- Meisels, S.J. (1998, November). *CIERA Report #3-002. CIERA Inquiry 3: Policy and Profession*. Ann Arbor, MI: Center for the Improvement of Early Reading Achievement Publications. Retrieved

January 25, 2001, from the World Wide Web:

<http://www.ciera.org/ciera/publications/report-series/inquiry-3/report32.html>

Murillo Jr., Enrique G.(1999). "Growing Pains: Cartographies of Change, Contestation and Social Division in North Carolina." Unpublished dissertation. University of North Carolina, Chapel Hill.

National Association for the Education of Young Children. (1988). NAEYC position statement on standardized testing of young children 3 through 8 years of age. *Young Children*, 43.

National Association for the Education of Young Children. (1990). NAEYC position statement on school readiness. *Young Children*, 46.

National Association of Early Childhood Specialists in State Departments of Education. (1987).

Unacceptable trends in kindergarten entry and placement. Springfield, IL: Author.

National Association of Elementary School Principals. (1990). Early childhood education: Standards for quality

programs for young children. Alexandria, VA: Author.

National Association of State Boards of Education. (1988). Right from the start. Report of the National Task Force on School Readiness. Alexandria, VA: Author.

National Association of State Boards of Education. (1991). Caring communities: Supporting young children and families. Report of the NASBE Task Force on Early Childhood Education, Alexandria, VA: Author.

National Commission on Children. (1992). Beyond rhetoric: A new American agenda for children and families. Washington, DC: Author.

National Commission on Excellence in Education (1983). *A Nation at Risk: The Imperative for Educational Reform*.

National Council of Teachers of English. (2000, November). *English Teachers Pass Resolutions on High-stakes Testing and the Rights of Test Takers* [Announcement]. Urbana, IL: Author.

Retrieved January 24, 2001, from the World Wide Web:

<http://www.ncte.org/news/2000resolutions2000November22.shtml>

National Council of Teachers of Mathematics. (2001). *High-Stakes Testing* [Position Statement]. Reston, VA: Author. Retrieved January 24, 2001, from the World Wide Web:

[http://www.nctm.org/about/position\\_statement/highstakes.htm](http://www.nctm.org/about/position_statement/highstakes.htm)

National Education Association. (2001). *Low Performing Schools* [Issues]. Washington, DC: Author. Retrieved January 25, 2001, from the World Wide Web:

<http://www.nea.org/issues/lowperf/>

Nie, N.H., Junn, J. & Stehlik-Barry, K. (1996). *Education and Democratic Citizenship in America*. Chicago and London: The University of Chicago Press.

Noblit, G.W., Dempsey, V.A., et al. (1996). *The Social Construction of Virtue: the Moral Life of Schools*. Albany: State University of New York Press.

Odden, A. and Dougherty, V. (1982). *State Programs of School Improvement: A 50-State Survey*. Denver, CO: Education Commission of the States.



- Patten, P. (2000, January-February). Standardized Testing in Schools. National Parent Information Network. Retrieved January 25, 2001, from the World Wide Web: <http://npin.org/pnews/2000/pnew100/feat100.html>
- Peterson, Paul and Noyes, Chad (1997). School Choice in Milwaukee. In Diane Ravitch and Joseph Vittereti (eds.) *New Schools for a New Century: The Redesign of Urban Education*. New Haven: Yale University Press, pp. 123-146.
- Phelps, R. P. (1999, January) *Fordham Report: Why Testing Experts Hate Testing*. Washington, DC: Thomas B. Fordham Foundation. Retrieved January 25, 2001, from the World Wide Web: <http://www.edexcellence.net/library/phelps.htm>
- Phelps, R. P. (2001). Test Bashing, Part 3: The Education Press's Cop-Out on Student Testing. *Education News*. Retrieved January 25, 2001, from the World Wide Web: [http://www.educationnews.org/test\\_bashingthe\\_education\\_press.htm](http://www.educationnews.org/test_bashingthe_education_press.htm)
- Posnick-Goodwin, S. (2000, November). Next spring's High School Exit Exam should concern educators at all levels. *California Educator*, Vol. 5, issue 3. Retrieved January 24, 2001, from the World Wide Web: [http://www.cta.org/cal\\_educator/v5i3/case\\_exam.html](http://www.cta.org/cal_educator/v5i3/case_exam.html)
- Ray, Carol Axtell & Mickelson, Roslyn Arlin (1993). Restructuring Students for Restructured Work: The Economy, School Reform, and Non\_College\_Bound Youths. *Sociology of Education* 66: pp. 1\_20.
- Robertson, A.S. (2000, November-December). "High-Stakes" Testing: New Guidelines Help Direct School Change. National Parent Information Network. Retrieved January 25, 2001, from the World Wide Web: <http://npin.org/pnews/2000/pnew1100/int1100b.html>
- Sacks, P. (1999). *Standardized minds: the high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus Books.
- Spring, Joel (2000). *The Universal Right to Education: Justification, Definition, and Guidelines*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Taylor, W.L. (2000, November). Standards, Tests, And Civil Rights. *Education Week*, Vol. 20, number 11. Retrieved January 25, 2001, from the World Wide Web: [http://www.edweek.org/ew/ew\\_printstory.cfm?slug=11taylor.h20](http://www.edweek.org/ew/ew_printstory.cfm?slug=11taylor.h20)
- Wiggins, G. P. (1993). *Assessing Student Performance: Exploring the Purpose and Limits of Testing*. San Francisco, CA: Jossey-Bass Publishers.

### Authors' Note

Alayne Sullivan and Enrique G. Murillo, Jr. are both teacher-educators in the Department of Language, Literacy and Culture at CSU San Bernardino. Dr. Sullivan coordinates and teaches in the Reading Program, and Dr. Murillo teaches in the area of Educational Foundations and Research Methods. The idea for this document came about while serving on the Educational Leadership Council for the College of Education. The authors would like to thank the members of the ELC for their support and contributions while writing this document, and are particularly grateful to both

Richard Ashcroft and Bob London for providing encouragement and insights. Also, many factors of the “contextual forces” were collaboratively analyzed by the second author with Lesley Bartlett. Correspondence may be addressed to: Enrique Murillo, Jr., College of Education, California State University, San Bernardino, 5500 University Parkway, San Bernardino, CA 92407-2397. Email: [alayne@csusb.edu](mailto:alayne@csusb.edu).