

The Testing Culture and the Persistence of High Stakes Testing Reforms

Michele S. Moses and Michael J. Nanna

Abstract

The purposes of this critical analysis are to clarify why high stakes testing reforms have become so prevalent in the United States and to explain the connection between current federal and state emphases on standardized testing reforms and educational opportunities. The article outlines the policy context for high stakes examinations, as well as the ideas of testing and accountability as major tenets of current education reform and policy. In partial explanation of the widespread acceptance and use of standardized tests in the United States, we argue that there is a pervasive testing culture, in addition to other contributing factors such as administrative utility, profit motives, and political ideology. Finally, we offer a critique of high stakes testing reforms in light of concerns about equality of educational opportunity.

Introduction

How is it that high stakes testing reforms persist despite evidence regarding a negative relationship with educational opportunity? In light of current testing and reform policies, our purposes are to: 1) highlight an important problem stemming from the misuse of large-scale testing as prescribed in educational reform initiatives such as No Child Left Behind (NCLB); 2) clarify why testing reforms have become so prevalent in the United States; and 3) explain the connection between current federal and state emphases on standardized testing reforms and educational opportunities, particularly among low-income students and students of color. In order to address these issues, we engage in critical analyses, following in the tradition of scholars such as Amy Gutmann (1987), who pointed out that “All significant policy prescriptions presuppose a theory. . . . When the theory remains implicit, we cannot adequately judge its principles or the political prescriptions that flow from them” (p. 6). In order to better understand the policy prescriptions surrounding current high stakes testing mandates, we first provide the political and research context for the debate surrounding their use. In an effort to identify some of the underlying principles and assumptions encompassing the ideas of testing and accountability,

we draw on diverse conceptual and empirical research (e.g., Amrein & Berliner, 2002; Evers & Walberg, 2004; Finn, 1995; Haney, 2000; Heubert & Hauser, 1999; Jones, Jones, & Hargrove, 2003; Koretz, 1995; Lemann, 1999; McNeil, 2000; McNeil & Valenzuela, 2001; Mehrens, 2004; Natriello & Pallas, 1999; Noddings, 2004; Orfield et al., 2004; Siegel, 2004; Valenzuela, 1999). Next, we explain three possible foundational reasons for the widespread acceptance and use of standardized tests in the United States, going beyond justifying their use as necessary ways of assessing student learning and academic progress. In addition, we develop the idea of the testing culture as an important reason why high stakes testing reforms are so prevalent. Finally, based on the philosophical work of Kenneth Howe (1997) and Ronald Dworkin (2000), we examine the concept of equality of educational opportunity in order to understand how it is connected to high stakes testing policy and access to educational opportunities. We argue that high stakes testing reforms, driven as they are by political and cultural ideology and concerns for efficiency and economic productivity, serve to impede the development of *real* equality of educational opportunity, particularly for the least advantaged students. As John Dewey wrote some 70 years ago: “[w]hat avail is it to win prescribed amounts of information about geography and history, to win ability to read and write, if in the process the individual loses his own soul” (1938, p. 49)?

The Policy Context for High Stakes Testing Reforms

Tests whose scores have a direct impact on a person’s life options and opportunities are considered to have high stakes. Some uses of high stakes tests include placing students in academic tracks, retaining students in grade, or deciding high school graduation eligibility. Viewed as a promising way to allot educational places and sort students based on individual merit (as based on test scores) rather than birth circumstances, intelligence and later achievement, testing began to be used widely in the middle of the twentieth century. The original intent was to increase equality of educational opportunity by providing a uniform and objective way to assess students’ abilities and aptitudes (Lemann, 1999). The use of high stakes tests has steadily increased since *A Nation at Risk* was published in 1983 (Amrein & Berliner, 2002; Dwyer, 2004). The No Child Left Behind Act of 2001 set in motion a flurry of education policy reforms in almost every state, directed primarily towards increasing measures of accountability and assessing educational outcomes by implementing large-scale standardized tests and, in so doing, galvanized the assessment movement into a national project.

Yet, some educational research has shown that high stakes tests serve to degrade educational aims, limit and constrict the curriculum, constrain teachers, stress students, and curtail access to postsecondary education (Amrein & Berliner, 2002; Booher-Jennings, 2005; Haney, 2000; Heubert & Hauser, 1999; Jones, Jones, & Hargrove, 2003; McNeil, 2000; McNeil & Valenzuela, 2001; Nichols, Glass, & Berliner, 2005; Orfield, et al., 2004; Siegel, 2004; Valenzuela, 1999). According to

Booher-Jennings (2005), relying on the data of student achievement test scores can serve to subvert both learning and equality of educational opportunity for far too many students. She found that some students are ignored because they have little hope of passing an exam, other students are placed dubiously in special education so as to avoid their being counted in a school's scores, while others are receiving more than their share of attention because they are on the cusp of passing the exam. (They even spend extra time in test preparation instead of attending music or gym class.) Other scholars tell a quite different story about how tests help improve schools by serving as diagnostic tools, measurements of curriculum impact, and assessments of both teachers' and students' efficacy and progress (Evers & Walberg, 2004; Finn, 1995; Mehrens, 2004). Mehrens (2004), for one, argued that multiple-choice tests are a better measure of students' knowledge and skills than alternative methods such as performance assessments. Koretz (1995), himself skeptical about the benefits of widespread standardized testing, pointed out: "tests . . . have redeeming social value. Assessment clearly has a great deal to offer educational reform if used prudently" (p. 156). Despite concerns, high stakes testing reforms have been proliferating in the name of providing a better education for, and fostering the academic achievement of, students who have been "left behind" in education. The idea is that high stakes motivate students and educators alike to raise academic achievement (as measured by test scores) and consequently help to close educational gaps of race, ethnicity, and class. As Nichols, Glass, & Berliner (2005) suggested, "When faced with large incentives and threatening punishments, administrators, teachers, and students, it is believed, will take schooling more seriously and work harder to obtain rewards and avoid humiliating punishments" (p. 1). Although there is some merit to this precept, at least in the professional world, and it is certainly appealing to many, this idea is not without controversy. One thing seems evident: there is widespread disagreement regarding the appropriate format and uses of testing, and often these disagreements follow political ideological lines as opposed to psychometric efficacy. In Lorraine McDonnell's (1997) words: "the politics of testing often revolves around conflicting values about the appropriate format and uses of testing" (p. 1).

The state of Texas is often seen as the leader in school reform based on testing. Beginning in 1991, students in Texas have had to pass the Texas Assessment of Academic Skills (TAAS)¹ in order to receive a high school diploma, even though such high stakes graduation exams have been shown to have a disproportionately negative effect on students of color (Natriello & Pallas, 1999; Nichols, Glass, and Berliner, 2005). Consider the case of student Natalie Martinez of San Antonio, Texas. Ms. Martinez, "an A or B student, except in math" and a talented singer, was awarded a full scholarship to attend the University of the Incarnate Word, but was unable to pass the mathematics portion of the TAAS and, as a result, did not receive her high school diploma (Schmidt, 2000, p. A28). The university subsequently revoked her admission. Though she eventually earned a diploma from a private high school that was not bound by the state test requirement and was readmitted to the

university two years later, the TAAS proved a formidable obstacle for an otherwise satisfactory student. Ms. Martinez commented: “The way I see it, if you want to go to college and be somebody, you are going to do what it takes. A test should not determine your future” (quoted in Schmidt, 2000, p. A28). Because one test *can* determine a student’s future, that student’s opportunities are seriously limited by the barriers imposed by many high stakes exams.

Supporters of high school exit exams argue that graduation tests may motivate students to be better prepared for college-level academic work. Thomas Vukovich, an associate provost at the University of Akron, credited Ohio’s graduation exam with raising incoming students’ level of preparedness (Schmidt, 2000). According to Vukovich, “I think it [Ohio’s exam] is sending a message. It is saying, ‘If you want to come to college, you have to prepare for it and be more serious about it in high school’” (quoted in Schmidt, 2000, p. A26). Indeed, according to the National Commission on the High-School Senior Year (2001), U.S. high schools are not succeeding in preparing all students for college-level academics. The study found that only 44 percent of first-year college students had taken a college preparatory curriculum in high school. Such findings lead some to champion difficult graduation exams as key to increasing the academic rigor in high school. As a result, there have been calls to use high school exit exams in college admissions and placement (Hebel, 2001). In this way, high school exit exams can be used to justify denying educational opportunities to students who do not pass their exams. In a court challenge to the TAAS, a United States District Court found the test constitutional despite plaintiffs’ complaints that it discriminates against African American and Latino students. The judge, Edward Prado, noted that the exit exam might help bridge achievement gaps between racial and ethnic groups because it serves to motivate both students and schools to do better (Schmidt, 2000). Yet, in the report *Betraying the College Dream* (2003), researchers from Stanford University’s Bridge Project found that confusing academic assessments interfere with students’ preparedness for college. High school academic requirements, especially high stakes state tests, often do not parallel expected college topics or workload. Indeed, the meaning of a high school diploma is at issue. Moreover, motivation, and in some cases, means, are being conflated with ability and achievement. Those who have adequate amounts of social and cultural capital most likely will perform better (Lareau, 2000).

An increasing number of states are now requiring that students pass an exam in order to graduate. As of mid-2004, these include Alabama, Alaska, Florida, Georgia, Indiana, Louisiana, Maryland, Massachusetts, Minnesota, Mississippi, Nevada, New Jersey, New Mexico, New York, North Carolina, Ohio, South Carolina, Tennessee, Texas, and Virginia, with Arizona, California, Idaho, Utah, and Washington planning to mandate such exams by 2009 (Center for Education Policy, 2004). Other states are gearing up to develop exams as well (Jones, Jones, & Hargrove, 2003). In light of the strong emphasis on testing reforms in current U.S. federal and state education policy, an important question arises: given the inconclusive

evidence and serious disagreement between scholars and researchers regarding the relationship between high stakes tests and inequalities, why have there been consistent calls for more testing reforms?

Understanding the Prevalence of Testing Reforms

Public schools long have been a target for politicians proclaiming that education is failing (Berliner & Biddle, 1995). The most recent federal criticism of America's public schools has come from President George W. Bush's *No Child Left Behind* (NCLB) policy, which, like *A Nation at Risk* before it, calls for more accountability with a commensurate increase in reliance on standardized test scores as evidence of educational efficacy. President Bush's call for more accountability and subsequent increases in standards in order to "leave no child behind" may actually contribute to increasing social and cultural inequalities and complicate an already complex situation (Booher-Jennings, 2005; Center on Education Policy, 2004; McNeil, 2000; Valenzuela, 1999).

Yet, the use of standardized tests in education in the United States can be traced back nearly 150 years, and in the latter half of the twentieth century became associated with a concern for fairness and educational opportunities so that students who would have been excluded from more prestigious higher education, for example, could compete on the basis of test scores rather than social status (Lemann, 1999). In addition to their meritocratic function, standardized tests can serve as a tool for measuring if and how well students are learning. Supporters argue that testing is fundamental to standards-based reform because results show whether students are mastering tougher coursework (Finn, 1995; Mehrens, 2004). The idea is, if educators set high standards and test them, students will learn.

Given the conflicting information regarding the consequences of standardized testing, why are high stakes standardized tests increasingly being used as measures of student learning and achievement, overall educational efficacy, and teacher effectiveness? We attempt to answer this question by describing three foundational factors that have fostered the widespread acceptance and use of standardized tests in the United States, beyond the most common explanation, which is based on the need to assess student learning. Few would disagree that we need to assess learning, but there is a great deal of disagreement as to how this should be, or even can be, effectively accomplished in an equitable manner. The three factors are: administrative utility, profit motives, and political ideology. We then highlight one new factor influencing the proliferation of high stakes testing reforms: the existence of a "testing culture" in the United States.

Factor #1: Administrative Utility

In addition to being seen as an efficient way to assess learning and progress, standardized tests have tremendous administrative appeal and can be cost-effective when processing large amounts of information. For example, college and standardized graduate school admissions test scores are often used in tandem with

other selection criteria to screen applicants for admission with the ultimate goal of selecting students who will *most likely* successfully complete a program of study. In this case, admission to college or graduate school is predicated on the two-fold assumption that 1) the desirability of a candidate is related to his/her scores on a standardized test, and 2) an individual's score on a standardized test points to his or her potential overall grade point average, or potential intelligence, graduation, and subsequent success (Thacker & Williams, 1974). Following this logic, standardized tests predict success in college or graduate school, with success in college or graduate school directly resulting in degree completion and professional success once out of school.

Whatever the outcome, standardized test scores seem to account for a considerable proportion of variance in admissions decisions (Ingram, 1983). Evaluating a large number of applications can be a tedious, difficult, and yet important task. Standardized test scores provide a relatively quick, convenient, and *prima facie* valid method for screening applications. For example, Millimet & Flume (1982) investigated the admission standards of numerous prominent graduate schools and found that a high standardized test score coupled with a high or moderate GPA was associated with the highest rate of acceptance. The interesting result in this study was that a low GPA coupled with a high standardized test score had a higher acceptance rate than a high GPA and low standardized test score, indicating the greater emphasis placed on standardized tests. Of course standardized test scores affect much more than admissions decisions. Test scores are used to determine educational progress, evidence of learning, teacher evaluations, school quality, and high school graduation, among other things. In each case, standardized test scores can provide a tidy solution to an administrative challenge. Furthermore, test scores are often used for the allocation of funds in the form of governmental bonuses to schools meeting state-set standards. As one example, “the initial [George W.] Bush plan called for more testing and the punishing and rewarding of states depending on pupil performance” (Spring, 2002, p. 74). The use of standardized test scores simply because they provide a great deal of efficiency and administrative utility provides an example of high stakes tests being used for reasons that have little, if anything, to do with actual educational value.

Factor #2: Testing and Profit

Money issues often play a part in policy decisions. Although the current increase in testing at all levels may not necessarily be the result of lobbying efforts on the part of the testing industry (which includes publishing companies, many of which have testing divisions, e.g., CTB/McGraw-Hill, Harcourt), the maintenance, expansion, and proliferation of testing efforts is of primary concern to the testing industry. Indeed, testing has become big business in the United States, and the testing industry, like many other industries, is frequently bottom-line profit driven and has benefited dramatically from privatization. Tests, as symbolic representations of an “official knowledge” (Apple, 1993)—one that is increasingly being set

with corporate and political interests in mind—are a highly marketable product and have tremendous appeal to a citizenry obsessed with numerical measures of outcome and well-being. With the demand for tests growing, the testing industry can expect to see a large increase in revenue; and with increased revenue come increases in marketing and lobbying power. Testing companies—many of which are publicly traded, for-profit organizations that must demonstrate annual growth and expansion in order to maintain acceptable stock ratings—thus will play their part in maintaining a thirst for testing that soon only they will be able to quench. This most likely will get worse, as the number of private high schools and tutoring services aimed at helping students prepare for high stakes tests (or providing an alternative educational pathway for those failing high stakes exams) is on the rise. Government endorsement of corporate ventures either through direct capital support or public policy is certainly nothing new, and the business is most likely welcome in the testing industry.

The financial success of the testing industry, however, should not be confused with the actual utility and validity of the tests to accomplish what it is they purport to accomplish. In theory, public and educational policies should be informed with sound evidence and the best of what is currently known, while also taking into account the needs of the market. Private industry, in theory, responds to the needs of the market as directed by the current policies in place in order to implement the resources needed for proper and effective implementation. Unfortunately, theory and reality do not always coincide and it is possible that the testing tail is now wagging the educational dog. Indeed, as we have mentioned, high stakes standardized tests have limits and may have negative consequences associated with their use (see, e.g., Booher-Jennings, 2005; McNeil, 2000; Nichols, Glass, and Berliner, 2005). However, the testing industry has much to gain both financially and politically by ensuring their widespread implementation and use, regardless of conflicting findings about the benefits of high stakes exams.

Factor #3: Testing and Political Ideology

Politics and education continue to be intertwined. In the case of high stakes tests, calls for increased testing and accountability are not limited to one political party. For example, liberal senators like Ted Kennedy co-sponsored the NCLB legislation. And during the 2004 presidential race, John Kerry and John Edwards's plan, although calling for additional accountability, proposed to use more "sophisticated tests that capture the full range of skills" that are needed for students to develop instead of relying on "fill-in-the-bubble" tests that "limit both teaching and learning" (www.johnkerry.com/issues/education). However, as Sacks (1999) observed, "political motivations and the exercise of political power by those in positions of authority, rather than sound educational reasons, have driven the nation's use of standardized tests in schools" (p. 70). In fact, political motivation and the exercise of political power have driven educational policy decisions on both sides of the ideological coin. Joel Spring (2002) provided a detailed historical outline of

the influence political ideology has played in shaping education policies in this country over the past 50 years, as well as the mechanisms and structures through which this influence has been, and still is, mediated. The issue at hand with regard to the use of standardized testing, then, results not from the tests themselves, but rather the political and ideological motivations that direct and institutionalize their use. Tests can be useful tools for measuring and monitoring numerous indicators. It is when high stakes and serious consequences are attached to student test scores that the indicators have the potential of becoming personally and socially harmful. There are an increasing number of studies that highlight primary and tertiary problems associated with high stakes testing. Though beyond the primary focus of this article, for a discussion of these and other issues see, e.g., Herman & Golan, 1993; Lomax, West, Harmon, Viator, & Madaus, 1995; Rothman, 1996; Anderson, 1998; Casas & Meaghan, 2001; Froese-Germain, 2001.

Understanding the current proliferation and use of high stakes tests requires an understanding of the context in which they were born. Much of the controversy surrounding three major federal reports and policies on public education, Ronald Reagan's *A Nation at Risk*, George Bush's *America 2000*, and George W. Bush's *No Child Left Behind*, stems from a fundamental disagreement between neo-conservative and liberal approaches. The current confrontation between neo-conservative and liberal positions is, in many ways, a reaction to the social policies of the 1960s and 1970s, which included the United States Supreme Court rulings that prohibited school prayer, upheld affirmative action, and established forced desegregation, as well as the War on Poverty, court-ordered busing, drug treatment programs, and political correctness movements on college campuses (Spring, 2002). According to some neo-conservatives, the "liberal elite" stands accused of trying to change social conditions through social policies instead of changing the moral character and values of individuals. Neo-conservative ideology tends to emphasize individuals rather than the material conditions surrounding the individuals. Stated briefly, neo-conservatism contends that society arises from and is influenced by the individual. From this perspective, standardized testing is an objective means towards organizing society based on ability and merit. The intent may have been benevolent in that early proponents of testing were attempting a move away from the class-based system to one that is based on ability. On the other side of the coin, liberal perspectives suggest that separating the individual from her or his cultural and socioeconomic surroundings is well-nigh impossible and that the objectivity of standardized testing is illusory.

Complicating these issues, political and business demands have led to dramatic shifts in educational focus. For example, markers of excellence such as educational equity in the 1960s and 1970s shifted towards excellence defined by social efficiency during the 1980s with the publication of *A Nation at Risk*. This shift towards social efficiency sought to address worker values and work habits, low worker productivity, as well as the declining economic edge of the United States in the world economy (Munro, 1998; Feinberg, 1993). It is manufactured crises like these (to use

the term coined by Berliner & Biddle, 1995) that suggest larger shifts in education policies, which only serve to “attack the ideals of justice” (Weiler, 1993, p. 216). The relationship between political ideology and standardized test use is related to the needs and demands of business, which stem back to fundamental differences in economic and social beliefs that play out in education policies and systematically (and negatively) affect those in poverty as well as people of color.

The issue, then, is that the use of high stakes testing is related to political ideology and the exercise of political will and power instead of being driven primarily by the best interests of students. Examining the testing culture in the United States may help us understand why this is the case.

The Testing Culture

Testing has become part of the very social fabric that comprises our current cultural blanket. As such, the longstanding effect that the early testing movement has had, and continues to have, on our collective cultural subconscious is difficult to discern. As Stephen Jay Gould (1981/1996) noted, “the twentieth century moved to the putatively more direct method of measuring the concept of brains by intelligence testing” (p. 23). Indeed, standardized testing has become commonplace in the United States and frequently begins so early in a child’s life that few people question its legitimacy (Oakes, 1985). Basic tests of intelligence have given way to achievement tests, which NCLB sanctions for use with students of all ages. These tests have high stakes when used as indicators of a student’s intellectual abilities with results that frequently set into motion a chain of events (educational and social) that establish a student’s position in the social hierarchy at an early age. In addition, there seems to be an ever-decreasing separation between notions of intelligence, personal self-worth, and the numbers that are taken as legitimate indicators of both. Anderson (1998) pointed out, “For many parents their child’s score does not measure just one day, or even a week, of their child’s work; it mirrors the family’s success, the parents’ affluence, and the child’s future” (p. 27). This trend is not limited to education, of course. Quantitative measures are increasingly being accepted as the only legitimate indicators of economic, political, mental, and physical well-being as well. The justification is often circular, however: if we can measure something, it is real, and the fact that we have measured it serves as evidence of its realness. Calls for increased standardized testing as measures of everything from student learning, teacher effectiveness, and school effectiveness underscore that Americans have bought into an overly simplistic view of the complex acts of teaching, learning, and even intelligence (Sacks, 1999) and are thus culturally predisposed to accepting a single number as an appropriate index of all three.

The underlying set of assumptions created through this complex process become ideological, and are known “only tacitly, remain unspoken, and are very difficult to formulate explicitly” (Apple, 1990, p. 126). These underlying assumptions eventually become so deeply entrenched in the culture that they need not be

explicitly expressed. Test results and subsequent policy decisions are taken at face value as legitimate.

There is a complex interplay between standardized testing and the culture in which it occurs and is accepted as legitimate. High stakes tests are used to lend legitimacy to existing cultural practices, particularly when the practice is validated by experts. Indeed, many of the technical aspects of testing are used as a justification for its legitimacy and, ultimately, widespread use. However, it is a complex interplay that runs both ways. For example, it is also the case that when standardized testing is accepted as a legitimate practice within a culture (by “legitimate practice” we mean one where test results are accepted as meaningful indicators of underlying constructs and/or content knowledge, i.e., where the symbolic numeric representations correspond to the underlying construct or content of interest), it becomes a powerful symbolic object of representation that is given unwarranted epistemic privilege. Moreover, the legitimizing forces of expert approval and validation, societal acceptance, institutionalized testing policies, as well as the testing mechanisms themselves, work together within a culture to perpetuate existing symbolic connections between testing and knowledge in ways that are neither justified nor sound. Testing, once accepted within a culture, is reproduced as a legitimate and meaningful representation like any other culturally specific tradition. Pierre Bourdieu (1977) stated that—in reference to reproductive forces within a culture—symbolic systems owe their practical coherence to the fact that they are the product of cultural practices that cannot perform their practical functions unless the principles they bring into play are not only coherent but also practical. From this perspective, the products of testing practices—such as test results and interpretations of the results—are justified only to the extent that they are connected to an underlying symbolic system of meaning that is seen as legitimate and having practical relevance within a culture.

Testing has become a powerful symbolic system within U.S. culture—one where the exchange of measured practices matches the anticipated practical functions for which they were intended. Standardized testing is often legitimated through a cultural (and political) apparatus that perpetuates existing class and race distinctions. Further complicating this problem is the fact that the language (i.e., underlying syntactical and symbolic representations) associated with testing has become deeply entrenched in the mainstream culture’s lexicon. As Deron Boyles (1998) posited, “consumer materialism is already ingrained in schools and is often represented in the language we use to talk about schooling. ‘Standards,’ ‘outcomes,’ ‘objectives,’ ‘goals’ bespeak consumer-materialist expectations for curriculum because they largely represent the intended ‘goods’ to be produced” (p. 24). Furthermore, the language associated with testing has become increasingly technical, “scientized,” and imbued with academic and scientific jargon that “experts” often need to decipher, leaving the general public to feed only on the bits of information that find their way into the media and political arena. Even though they may not like high stakes standardized tests, most Americans accept testing as

legitimate because it is a part of the culture, rather than accepting testing into the culture because it is legitimate.

With the logic of standardized testing firmly entrenched in the U.S. educational system, skeptical educators, scholars, and policy makers have a much harder time challenging the notions that “achievement” is measured best by tests or that high stakes tests support *learning* (Nichols, Glass, and Berliner, 2006). Given such a one-dimensional view of “achievement,” our primary concern is that high stakes exams too often may serve to erode students’ educational opportunities rather than enhance them (Booher-Jennings, 2005; Center on Education Policy, 2004).

Equality of Educational Opportunity

In order to clarify the conception of equal educational opportunity assumed by federal and state high stakes testing policy, we rely on Howe’s (1997) interpretation of equal educational opportunity and Dworkin’s (2000) notion of equality of resources.

Howe (1997) examined three different interpretations of the ideal of equality of educational opportunity: formalist, compensatory, and participatory. Howe attempted to advance a meaningful interpretation (i.e., participatory) built on defensible philosophical principles within the sea of misunderstanding surrounding this ideal. Under the formalist interpretation, equal educational opportunity merely entails that there be no formal, or legal, barriers. As long as there is no formal barrier (like, for example, the separate but equal doctrine) stopping someone from the opportunity, then it may be considered an equal one. Consider the monolingual Chinese-speaking students in the *Lau v. Nichols* (1974) case. Because they were allowed to attend school, formalists would argue that they had the same opportunity as other students to go to school and learn. Yet, with no bilingual education measures in place, that opportunity was an empty one because the students could not understand the language of instruction and thus could not learn. Howe dismissed the formalist conception as simplistic because it ignores the relationship between students and educational institutions, as well as the complex interplay between cultural capital and linguistic fluency. Moreover, having equal structures and policies in place says nothing about how well individuals are able to navigate through complex social systems; that is, equality is not just a structural matter, it is a cultural one as well.

Under the compensatory interpretation, equal educational opportunity is fulfilled when students’ disadvantages are remedied. The idea is that once individual deficiencies are corrected, students can better adapt to the dominant system and achieve equal educational opportunity. However, the problem with the compensatory interpretation is that it advocates compensating students who, because they have not been a part of forming the educational system, are considered to be at a deficit. They thus have to sacrifice themselves by having to make unreasonable changes in order to fit into the dominant structures, when it is those structures that ought to adapt to the needs of diverse students.

The conception of equality of educational opportunity assumed by high stakes testing reforms falls into a compensatory approach. High stakes tests are championed as vehicles for compensating for social and educational inequalities (NCLB, 2001). Such reforms purportedly are intended to motivate students and teachers to prepare better for tests, resulting in increased achievement as measured by the test scores. This way the testing reforms somehow are making up for the consistently poor resources, curriculum, and instruction received by low-income students and students of color. In the case of students of color, the high stakes testing reforms would serve to close the “achievement gap.” But the gaps may be widening (Center on Education Policy, 2004), and students may not be learning more under high stakes testing reforms (Amrein & Berliner, 2002; Booher-Jennings, 2005). Some students and teachers are faced with constricted curricula that do not adequately account for student diversity as well as high-pressure situations that serve to impede educational progress (Booher-Jennings, 2005; McNeil, 2000; Valenzuela, 1999). Moreover, in a recent study by Nichols, Glass, and Berliner (2005), “States with greater proportions of minority students implement accountability systems that exert greater pressure, [suggesting] that any problems associated with high-stakes testing will disproportionately affect America’s minority students” (p. ii).

In response to the formalist and compensatory notions, which Howe saw as creating opportunities that are not “worth wanting,” he fashioned the participatory ideal. The participatory ideal transcends the formalist and compensatory interpretations by focusing on the structural and institutional facets of education that cause oppression. In addition, it places inclusion and democratic deliberation at center stage and involves a renegotiation of the goals and procedures of education so that diverse perspectives can be included. This way, people who historically have been and currently are excluded from the educational conversation can have a voice and take part in the negotiation of educational opportunities that are actually worth having.

Dworkin (2000) conceptualized a comprehensive theory of justice centering on equality. For Dworkin, equality is *the* fundamental value within political theory. He grappled with the tension between liberty and equality as sovereign virtues, and acknowledged that the two ideals work together. “Can it really be more important,” he asked, “that the liberty of some people be protected, to improve the lives those people lead, than that other people, who are already worse off, have the various resources and other opportunities that *they* need to lead decent lives?” (p. 121). His major concern was that while equality may be a political ideal, it is not striven for in practice. Education is an important pathway for improving one’s quality of life; high stakes testing actually may impede this process for those most in need of continued education.

Dworkin’s conception of equality can serve to justify a governmental redistribution of resources. He argued for the idea of “equality of resources,” which emphasizes the resources and opportunities persons have, rather than equality of

welfare, or persons' well-being throughout life. He defined equality of resources as a fair distribution of "whatever resources are owned privately by individuals" (p. 65). A person's resources would include wealth and education together with personal resources such as health, strength, and talent, and these together with "legal and other opportunities" (p. 286). Because high stakes testing policies tend to limit access to educational opportunities, Dworkin would say they weaken the ideal of equality. For example, students are not treated with equal respect and concern when they are denied opportunities for higher education due to high stakes high school exit exams. Yet, students have the right to be treated as equals, that is, "with the same respect and concern as anyone else" (Dworkin, 1978, p. 227). One clear implication of the political theory on the concept of equality is that education policy or reform should not impede that right.

Advocates of high stakes testing would say that by holding all students to the same standards, policies like NCLB actually do serve to treat all students as equals (Finn, 1995). The problem here is that such a response illustrates a deep misunderstanding of the complex idea of equality and what equal actually means. To be treated the *same* is not necessarily to be treated as *equal* or to be treated equitably. Relevant differences among students, whether they are according to socioeconomic status, race, ethnicity, sex, ability, or religion, etc., require different treatments in order to have a chance at treatment as equals in a meaningful way (Gutmann, 1987).

Conclusion

Despite an emphasis on the ideal of equality within the American educational system, there exists substantial inequality, both between and within schools (Bowles & Gintis, 1976; Howe, 1997). High stakes testing remains the education reform of the day, touted as a solution to the problems of low educational achievement and attainment. As of this writing, every state except Iowa either has a high school graduation exam or is in some stage of developing one. Assumed under the testing mandates is that all students have an equal opportunity for passing high stakes graduation exams (i.e., they all take the same exam in their state), and likewise are afforded the same educational and life opportunities.

However, standardized tests—even in an ideal state—may reflect the inequities that already exist within schools rather than meaningful differences in intelligence, student learning, and teacher effectiveness that the tests are touted as measuring. In addition, standardized test scores that reflect the systematic differences between different groups may help to maintain existing educational and social inequalities and may have devastating affects on low-income students and students of color, or any group with unequal access to cultural and social capital. This flies in the face of what Howe (1997) and Dworkin (2000) envisioned for a participatory and equal educational system that enhances rather than erodes students' educational opportunities.

If we are concerned about enhancing opportunities for the least advantaged students in the United States, then the arguments presented herein carry significance for the debate over high stakes standardized testing. Several important questions call for more research. First, what are some other ways for states to hold students (and educators) accountable for learning and preparation for college? Do we need to have some standardized form of assessment at the end of high school? Do these assessments elevate the value of a high school diploma, or do they merely serve to lessen the strain placed on educational institutions, as well as solidifying already existing hierarchical and social inequalities?

Standardized tests are not inherently negative (Heubert & Hauser, 1999). Testing can serve to assess student learning and progress. However, when used in inappropriate ways, tests can be damaging to students both emotionally and academically and can lead to a restriction of educational opportunities, particularly for the least advantaged students in the United States. A major problem, as Nel Noddings (2004) has pointed out, is not that students have to take tests in school in order to meet demands for educational accountability; it is that students are *forced* to take tests with very high stakes. In 1916, Dewey warned about the dangers of educational aims being imposed on schools and students externally. His warnings, as well as more recent ones, have yet to be heeded within federal education policy. Numerous education groups have urged caution with regard to high stakes testing, including the National Education Association (NEA), the American Association of School Administrators (AASA), and the American Educational Research Association (AERA). The cautions put forward by these and other educational organizations generally center on how standardized tests are to be used and not whether they should be used. Indeed, test scores are important numbers that not only affect people's learning, but also their lives, their self-concept and self-worth. Test scores often continue to exist in our subconscious long after their intended outcome has passed.

Moreover, careful attention should be given to the influence that cultural and political forces have on the acceptance, use, and legitimacy of large-scale testing. To do otherwise would ignore the fact that these larger forces have a tremendous impact on how testing is viewed within a culture—an impact that transcends psychometric value. In this article, we have endeavored to provide a deeper understanding of why high stakes testing reforms are so prevalent and how they serve to constrain educational opportunity. The overall aim of these analyses is to promote education reform and policy that best serves the interests of students who have little voice in the discussion and formulation of the policies that directly affect their life options.

Notes

We would like to thank the anonymous reviewers for their very helpful and detailed feedback on this manuscript.

1. The state of Texas changed from the TAAS to a new, more comprehensive—and, some say, more difficult—exam called the Texas Assessment of Knowledge and Skills

(TAKS). Beginning with the 2004–2005 school year, the TAKS became the basis of the state’s accountability system.

References

- American Educational Research Association (1999). *Standards for educational and psychological testing*. Washington, D.C.: Author.
- Amrein, Audrey L. & Berliner, David C. (2002). High stakes testing uncertainty and student learning. *Educational Policy Analysis Archives* 10(18). Retrieved November 11, 2002 from <http://epaa.asu.edu/epaa/v10n18>.
- Anderson, S. R. (1998). The trouble with testing. *Young Children*, 53 (4), 25–29.
- Apple, Michael. (1990). *Ideology and curriculum*. New York: Routledge.
- Apple, Michael. (1993). *Official knowledge*. New York: Routledge.
- Berliner, David & Biddle, Bruce. (1995). *The manufactured crisis*. Reading, MA: Addison-Wesley.
- Booher-Jennings, Jennifer. (2005). Below the bubble: “Educational triage” and the Texas Accountability System. *American Educational Research Journal*, 43(2), 231–268.
- Bourdieu, Pierre. (1977). *Outline of a theory of practice*. Cambridge and New York: Cambridge University Press.
- Bowles, Samuel and Gintis, Herbert. (1976). *Schooling in capitalist America: Educational reform and the contradictions of economic life*. New York: Basic Books.
- Boyles, Deron. (1998). *American education and corporations: The free market goes to schools*. New York: Garland Publishing.
- The Bridge Project. (2003). *Betraying the college dream: How disconnected K-12 and postsecondary education systems undermine student aspiration*. Palo Alto, CA: The Stanford Institute for Higher Education Research.
- Casas, F.R. & Meaghan, D.E. (2001). Renewing the debate over the use of standardized testing in the evaluation of learning and teaching. *Interchange*, 32(2), 147–181.
- Center on Education Policy. (2004). *State high school exams: A maturing reform*. Washington, D.C.: Author.
- Dewey, John. (1938). *Experience and education*. New York: Collier.
- Dewey, John. (1916). *Democracy and education*. New York: Free Press.
- Dworkin, Ronald. (2000). *Sovereign virtue: The theory and practice of equality*. Cambridge, MA: Harvard University Press.
- Dworkin, Ronald. (1978). *Taking rights seriously* Cambridge, MA: Harvard University Press.
- Dwyer, James G. (2004). Introduction to symposium: School accountability and “high stakes” testing. *Theory and Research in Education*, 2(3), 211–217.
- Ehrenfeld, J. (2001). The trouble with testing. *Education Week*, 21(8), 43.
- Evers, Williamson M. & Walberg, Herbert J., Eds. (2004). *Testing student learning, evaluating teaching effectiveness*. Palo Alto, CA: The Hoover Institution Press.
- Feinberg, Walter. (1993). *Japan and the pursuit of a new American identity: Work and education in a multicultural age*. New York: Routledge.
- Finn, Chester E. (1995). Who’s afraid of the big, bad test? In Diane Ravitch (Ed.), *Debating the future of American education*, 120–144. Washington, D.C.: The Brookings Institution.

- Froese-Germain, B. (2001). Standardized testing + high-stakes decisions = Educational inequity. *Interchange*, 32(2), 111–130
- Gould, Stephen J. (1996). *The mismeasure of man*. New York: W.W. Norton and Company.
- Gutmann, Amy. (1987). *Democratic education*. Princeton, NJ: Princeton University Press.
- Haney, Walt. (2000, August). The myth of Texas miracle in education. *Education Policy Analysis Archives*, 8(41). Retrieved on December 16, 2000 from <http://epaa.asu.edu/epaa/v8n41/>.
- Hebel, S. (2001, February 9). Universities push to influence state tests for high-school students. *The Chronicle of Higher Education*, 47(22). Retrieved on April 22, 2003, from <http://chronicle.com/weekly/v47/i22/22a02301.htm>.
- Herman, J.L. & Golan, S. (1993). The effects of standardized testing on teaching and schools. *Educational Measurement, Issues and Practice*, 12, 20–25.
- Heubert, Jay P. & Hauser, R.M. (1999). *High stakes: Testing for tracking, promotion, and graduation*. A report of the National Research Council. Washington, DC: National Academy Press.
- Howe, Kenneth R. (1997). *Understanding equal educational opportunity: Social justice, democracy, and schooling*. New York: Teachers College Press.
- Ingram, R.E. (1983). The GRE in the graduate admission process: Is how it is used justified by the evidence of its validity? *Professional Psychology: Research and Practice*, 14, 711–714.
- Jones, M.G., Jones, B.D., & Hargrove, T.Y. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman and Littlefield.
- Koretz, Daniel M. (1995). Sometimes a cigar is only a cigar, and often a test is only a test. In Diane Ravitch (Ed.), *Debating the future of American education*, 154–166. Washington, D.C.: The Brookings Institution.
- Lareau, Annette. (2000). *Home advantage: Social class and parental intervention in elementary education*. Lanham, MD: Rowman and Littlefield.
- Lemann, Nicholas. (1999). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus, and Giroux.
- Lomax, R.G., West, M.M., Harmon, M.C., Viator, K.A., & Madaus, G.F. (1995). The impact of mandated standardized testing on minority students. *The Journal of Negro Education*, 64, 171–185.
- McDonnell, Lorraine M. (1997). The politics of state testing: Implementing new student assessments. *CSE Technical Report 424*. Los Angeles, CA: The Regents of the University of California.
- McNeil, Linda M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- McNeil, Linda & Valenzuela, Angela. (2001). The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric. In Mindy Kornhaber and Gary Orfield (Eds.), *Raising standards or raising barriers? Inequity and high stakes testing in public education*. New York: The Century Foundation, 127–150.
- Mehrens, William A. (2004). Using performance assessment for accountability purposes. In Evers, Williamson M. & Walberg, Herbert J. (Eds.), *Testing student learning, evaluating teaching effectiveness*. Palo Alto, CA: The Hoover Institution Press, 221–242.

- Millimet, C.R., & Flume, M.E. (1982). Estimating graduate admission standards in psychology. *Research in Higher Education*, 17, 125–137.
- Miner, Barbara. (2005). Testing companies mine for gold. *Rethinking Schools*, 19(2), 5–7.
- Munro, Petra. (1998). *Subject to fiction: Women teachers' life history narratives and the cultural politics of resistance*. Philadelphia: Open University Press.
- Murray, Charles. (1984). *Losing ground: American social policy 1950–1980*. New York: BasicBooks.
- National Commission on the High School Year. (2001, October). *Raising our sights: No high school senior left behind*. Final Report of the National Commission on the High School Year. Princeton, NJ: The Woodrow Wilson National Fellowship Foundation. Retrieved on July 15, 2003 from www.commissiononthesenioryear.org/Report/FINAL_PDF_REPORT.pdf.
- Natriello, Gary, & Pallas, Aaron M. (1999). *The development and impact of high stakes testing*. Sponsored by the Harvard Civil Rights Project. Harvard University and Teachers College, Columbia University and Columbia Law School. Presented at the High Stakes K-12 Testing conference. New York, NY, December 4, 1998. Revised November 1999. Retrieved on July 11, 2003 from www.law.harvard.edu/civilrights/conferences/testing98/drafts/natriello99html.
- Nichols, S. L., Glass, G. V, & Berliner, D. C. (2006). High-stakes testing and student achievement: Does accountability pressure increase student learning? *Education Policy Analysis Archives*, 14(1). Retrieved January 23, 2006 from <http://epaa.asu.edu/epaa/v14n1/>.
- Nichols, S.L., Glass, G.V., & Berliner, D.C. (2005). High-stakes testing and student achievement: Problems for the No Child Left Behind Act. Article retrieved on October 10, 2005 from: <http://www.asu.edu/educ/epsl/EPRU/documents/EPSSL-0509-105-EPRU.pdf>
- No Child Left Behind Act of 2001*. Public Law 107-110, 115 Stat. 1425 (2002).
- Noddings, Nel. (2004). High stakes testing: Why? *Theory and Research in Education*, 2(3), 263–269.
- Oakes, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.
- Orfield, Gary, Losen, D., & Wals, J. (2004). Losing our future: How minority youth are being left behind by the graduation rate crisis. Cambridge, MA: Harvard Civil Rights Project. Retrieved September 21, 2004 from <http://www.civilrightsproject.harvard.edu/research/dropouts/dropouts04.php>.
- Rothman, R. (1996). Taking aim at testing. *The American School Board Journal*, 183, 27–30.
- Sacks, Peter. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus Publishing.
- Schmidt, Peter. (2000, January 21). Colleges prepare for the fallout from state testing policies. *The Chronicle of Higher Education*, pp. A26–A28.
- Siegel, Harvey. (2004). High stakes testing, educational aims and ideals, and responsible assessment. *Theory and Research in Education*, 2(3), 219–233.
- Spring, Joel. (2002). *Political agendas for education* (2nd Ed.). London: Lawrence Erlbaum.
- Thacker, A.J., & Williams, R.E. (1974). The relationship of the Graduate Record

Examination to grade point average and success in graduate school. *Educational and Psychological Measurement*, 34, 939–944.

Valenzuela, Angela. (1999). *Subtractive schooling: U.S. Mexican youth and the politics of caring*. Albany, NY: State University of New York Press.

Weiler, Kathleen. (1993). Feminism and the struggle for a democratic education: A view from the United States. In M. Arnot and K. Weiler (Eds.), *Feminism and social justice in education*. London: Falmer Press.

William, R. (2003). Historical high-stakes policies relating to unintended consequences of high-stakes testing. *Educational Measurement*, 22(1), 33–35.

Michele S. Moses is an associate professor of educational foundations, policy and practice at the University of Colorado at Boulder.

Email: michele.moses@colorado.edu

Michael J. Nanna is director of Institutional Research and Planning at the College for Creative Studies.

Email: mnanna@ccsad.edu