Blake, J. E. (2012). High-Stakes Testing: A (Mis)construed, Normalizing Gaze *International Journal of Educational Policies, 6(1), 5-23.*

High-Stakes Testing: A (Mis)construed, Normalizing Gaze

Jan E. Blake[*]
University of South Florida, USA

Abstract
Today, under the federal mandates of No Child Left Behind (NCLB), high-stakes testing has become a central national policy issue where national and state policies within the reform agenda are both propelled and supported by mandated high-stakes tests. These high-stakes tests are outwardly held as a metaphor for standards of student learning, teacher effectiveness, and school success, where a number or score shapes our perceptions and our objectivity within the educational environment. With such sweeping expectation, meaning and consequence attached to a single test score one is left to question high-stakes reform agendas which, seemingly, are antithetical to the purpose of education. Although the term high-stakes testing is utilized and frequently referred to, a commonly held or specifically stated definition of high-stakes testing has not been shared amongst stakeholders. The purpose of this paper is to document how high-stakes tests are both recognized and framed within the current high-stakes, objectified educational terrain as the nation races into the Race to the Top initiatives.

Keywords: *High-stakes testing, education policy, reform, standard setting*

[*] Janice Blake is an Assistant Professor in the Reading and Childhood Education in the College of Education at the University of South Florida St. Petersburg.
E-mail: *jblake4@mail.usf.edu*

"We are entering the age of infinite examination and of compulsory objectification" (Foucault, 1995, p. 189).

Introduction

Testing tied to educational reform, as we know it today, had its inception in the mid-1930s. By the end of the 1940s almost every school in the United States was using some form of standardized test (Callahan, 1962). During the 1950s tracking and selection were at the forefront of educational goals under the implementation of the National Defence Education Act (NDEA). The 1960s saw a move to utilize test scores for program accountability of high-poverty schools with the implementation of the Elementary and Secondary Education Act (ESEA) and began a massive increase in testing. The 1970s marked the beginning of state mandated minimum competency testing. A sharp move in the 1980s saw randomly sampled test scores being flagged for purposes of identifying overall school system accountability under the initiatives of the National Assessment of Educational Progress (NAEP) report. As a result, the federal standards-based accountability movement became the hallmark of the 1990s as proficiency standards were mandated and then acted as the benchmark of testing and primary indicator of student knowledge, teacher proficiency and school effectiveness (Cross, 2004).

Under the federal mandates of No Child Left Behind (NCLB), high-stakes testing continues to hold a leading role in American education where test scores are being used for ways and means in which they were never designed, normed or intended (Linn, 2000, 2003; Messick 1995). In the half century since Sputnik (circa 1950), teachers have weathered a plethora of federal, state and local test-based educational reform. "Over the last 15 years, the movement for higher standards and accountability in our schools has led several states – and now the federal government, with the No Child Left Behind (NCLB) Act – to adopt test-based accountability policies" (Goldberg, 2004, p. 8). The last half century has seen consequential shifts in high-stakes testing educational policies and the implementation of far-reaching reform mandates resulting from the NCLB Act.

Today, the current press to finalize the development of the Common Core Standards along with competitive funding offered by Race to the Top initiatives for development of next generation high-stakes assessments should stand as cautionary moment to those who have weathered the erroneous efforts of the previous two decades of federal testing reforms. Concerns stemming from the positioning of these next federal initiatives are synonymous of yet another uncharted and misconstrued footrace, as evidenced by a recent review, conducted by The National Policy Center (2010), stating that research undergirding Race to the Top initiatives are not rigorous or scientifically based. This review further states, there was "extensive use of non- research and advocacy sources to justify policy recommendations"; and there was "an overwhelming reliance, with little or no research justification, on standardized test scores as a measure of student learning and school success" (National Education Policy Center, 2010).As high-stakes testing continues to morph as a construct and stand as a metaphor for teaching and learning, these new federal education initiatives frame high-stakes testing as a tool to name and characterize teaching and learning as a "product" (Blake, 2008; Cross, 2005; Gabriel & Lester, 2011) to be measured and manipulated in ways and means by a multitude of stakeholders. The purpose of this paper is to recognize and frame high-

stakes testing as a multi-faceted and misconstrued construct within the educational setting.

High-Stakes Testing in the Twenty-First Century

High-stakes testing is a central national policy issue (Allington, 2002; Cuban, 2007) where national and state educational policies within the reform agenda are both propelled and supported by mandated high-stakes tests. These high-stakes tests are seemingly held as a metaphor for standards of student learning, teacher effectiveness, and school success, where a number or score shapes our perceptions and our objectivity within the educational environment (Dorn, 1998). With such sweeping expectation, meaning and consequence attached to a single test score one is left to question high-stakes testing reform agendas which, distinctly, are antithetical to the purpose of education (Dewey, 1938).Recently, Nichols and Berliner (2007) asked us to consider how and why high-stakes testing has so seamlessly slipped into the culture of education in America. Stone (2002) cautions the overreliance on a single number, or test score, reminding policymakers that "to select one feature of something, assert a likeness on the basis of that feature, and ignore all other features" (p. 165) is to exclude or ignore what may the most essential to effective educational reform. Here, high-stakes tests present as a complex and variably recognized construct where the "varied nature, impact, and role of high-stakes testing" (Nichols, Glass & Berliner, 2006, p. 11) plays out, with serious consequence, across State education testing mandates (Perie, 2008).

Notably, the purposes and uses of high-stakes tests have become a source of concerned debate among stakeholders, who see the consequences of high-stakes testing as having significant effects within the larger educational reform known as No Child Left Behind (NCLB) (Amrein & Berliner, 2002b). Aspects of the NCLB Act form a core of accountability where the intent of the Act places an emphasis on standards of learning for all children and directs greater attention to those groups of children who have been largely ignored or marginalized in the past. One of the central policy objectives of NCLB was to increase testing to all students in the nation to improve student achievement by 2014. Consequently, the amount of testing in each state has increased dramatically and the resulting high-stakes climate of accountability has placed complex pressures on schools to increase student achievement. While the NCLB Act "stays the course of standards-based reform and encourages states to adopt ambitious subject-matter standards" (Linn, 2003, p. 4), exactly how states have mandated and enacted these reforms have played out and resulted in highly test responsive and test dominated educational settings.

Allington (2002) has stated that NCLB dramatically changed the testing story, making high-stakes tests one of the leading and central characters of current educational reform. Amidst persistent concerns, focusing on determining and reporting the quality of education and performance of schools and teachers, high-stakes testing is predominately viewed as a decidedly efficient way to obtain numbers and scores, which are then directly transferred to highly publicized standards measures (Kohn, 2000; Kozol, 2005; Nichols & Berliner, 2007) of educational reform. In contrast to a, somewhat commonly held, perspective that test scores are both reliable and valid measures of achievement, Heubert and Hauser (1999) contend that high-stakes standardized group achievement tests are neither. As a result, the compliance and reliance on high-stakes test data to guide the

educational reform agenda has facetiously co-opted the intent and promise of the NCLB Act.

Today, the over simplification of high-stake testing and accountability (Linn, 2003) surfaces as a consequential force behind much of the current educational purpose embedded within the Race to the Top initiatives. It is critical to note that federal law takes no position on how states and individual school districts enact or make use of high-stakes test data (Heubert, 2002). However, within the educational reform agenda high-stake test data is at the centre, pushing schools towards ill-conceived standards and accountability agendas (Spillane, 2004). Synonymously, Stone (2002) describes educational mandates resulting from high-stakes testing as those striving towards the lowest cost objectives where "[g]etting the most out of a given input or achieving an objective for the lowest cost are simple definitions of the goal of efficiency" (p. 61). As a result, the effects of high-stakes testing as an accountability tool has systematically influenced the standards reform, thereby resulting in a complex, misunderstood high-stakes testing environment which ultimately both influences and impacts the experience and success of children's schooling (Dutro & Selland, 2012). The predominant view of high-stakes tests as the panacea to both increase student achievement and to make schools more cost efficient is highly problematic

Historically, high-stakes tests have proven to be efficient tools in the production of numbers and scores, but conversely have presented as highly unstable instruments and indicators of effective teaching and student learning (Linn, 2000). However, Allington (2002) argues that teachers who are caught within the policy trap with "less and less professional autonomy paired with more and more accountability" (p. 33) find themselves losing more and more of their teacher professional autonomy as they work within a high-stakes test environment that places great value on the production of numbers and scores. While the initial intent of NCLB was to set educational standards, improve the educational learning opportunities for all students and thereby raise achievement scores of students, the current outcomes of NCLB appear to have cast a normalizing gaze, homogenizing and mandating a standardized and highly test based educational setting focused on achievement and conformity.

Today, as the NCLB education reform agenda nears 2014, where the mandates of NCLB call for 100 percent proficiency of the nation's children in reading and mathematics, it is vital that a commonly held understanding of high-stakes testing is at the forefront of all data-based reform. While there is a push to develop new standardized tests there is no agreement as to "what" these test actually are and "what" they commonly intend to measure. Now, as part of the Race to the Top competition, describing the next generation of high-stakes tests, statements such as these: "states … will be able to compare results with, and learn from, one another" (Duncan, 2012) should cause alarm. Are we asking - Is there now a standard high-stakes test to measure this homogeneous, normalized student who is on track to be ready for college and 21st-century careers? Are we concerned that without a clear and commonly held definition of high-stakes testing the Race to the Top next generation testing initiatives set the culture of schooling firmly upon highly consequential test-driven key elements:

1. Teachers will be evaluated in relation to their students' test scores.
2. Schools that continue to get low test scores will be closed or turned into charter schools or handed over to private management.

3. In low-performing schools, principals will be fired, and all or half of the staff will be fired.
4. States are encouraged to create many more privately managed charter schools.

Historical Context of High-Stakes Testing

Student achievement has become the prevailing concern in education. In 1983, A Nation at Risk: The Imperative for Educational Reform (NCEE) was published in the United States, declaring that students in American schools were failing and American education required a major repair (Tyack & Cuban, 1996). The result of this publication was the heightened scrutiny of curriculum, instruction, teachers, schools, and ultimately set in motion the current reform movement in which high-stakes testing today is at its centre. Under the Reagan administration, the federal government began to encourage states to "raise standards, increase testing, establish accountability, and strengthen requirements, and secure better teachers and better teaching" (Perkinson, 1995, p. 370). Subsequently, school reform was given its charge and the monitor and gate-keeper became high-stakes testing as we know it today.

This single report (NCEE, 1983) moved the education agenda from teaching the basics to monitoring school quality and setting standards of performance and proficiency (McGill-Franzen, 2000). Accountability, in the form of high-stakes tests at all levels, was seen as the key to raising individual student proficiencies. The pressure cooker of the high-stakes testing education environment was just beginning to come to a boil. Report cards of student test performance were now the standards to determine how and where schools and school systems placed on a national comparison. Education became a footrace where educators were now under considerable pressure of high-stakes consequences to raise test scores. Overwhelmingly, teachers responded to this high-stakes teaching environment by focusing on the skills being tested, teaching test-taking skills, using test format to guide and form their daily teaching content and practices (Allington, 2002; Amrein & Berliner, 2002b; Corbett & Wilson, 1991; McGill-Franzen & Allington, 2006). High-stakes testing became a polished machine reporting student achievement. However, to date, there has been no significant scientifically-based research evidence that high-stakes testing has come close to the intended effect of increasing learning and improving teaching (Nichols & Berliner, 2008). The serious issue within education research and practice has become establishing a common understanding of what high-stakes testing within the initiatives of reform means to all constituents - this is a critical element that continues to remain misconstrued and requires deconstruction.

Defining High-Stakes Testing: The Construct

Although the term *high-stakes* testing is utilized and frequently referred to in the current research, few educational researchers specifically state a definition of high-stakes testing in the presentation of their research. Yet current literature is found to be replete with references highlighting and debating the effects of high-stakes testing. A review of the literature found that few research studies specifically define the construct. In addition, this review determined that while educational researchers have made less of a specific

definition explicit within their writing, educational theorists, writers, and journalists have frequently made assertions referencing and defining high-stakes testing in the literature (Bracey, 2000; Kohn, 2000; Mehrens & Popham, 1992; Nichols & Berliner, 2007; Resnick & Resnick, 1985). These writers have defined high-stakes testing as being both problematic and spurious as well as being instruments of sound educational policy.

Researchers who have stated definitions of high-stakes testing in their scholarly writing generally define the construct in markedly similar ways. According to Spring (2004), high-stakes testing "refers to an examination that determines a person's future academic career and job opportunities" (p. 36). Crawford and Impara (2001) stated that "(W)henever assessments affect the lives of students, we may consider those to be high-stakes tests" (p. 140). Franzak (2004) described high-stake tests, stating "(S)tandardized assessments become high-stakes when educational or personnel decisions are based on the results" (p. 235). Drawing from their research, Paris, Lawton, Turner and Roth (1991) defined high-stakes tests as those tests where "the consequences are profound for the respondents" (p. 12). Jones and her colleagues (Jones, Jones, & Hargrove, 2003) contend that tests acting as a "way to measure student achievement and school quality and as a mechanism to hold students and educators accountable" (p. 1) are held as high-stakes test. Further, Guthrie (2002) stated that "(A) test or testing program is called high-stakes when it is used to make important decisions about individual students, teachers, or schools" (p. 370). Taken as a whole, high-stakes testing framed within the educational literature draws upon the historical and socio cultural context and describes the prevailing definitions of high-stakes testing, thereby representing a composite definition which is reflective of theory, research and practice.

Within the educational research forum Madaus' (1988a) highly referenced (Au, 2007; Cimbricz, 2002; Grant, 2000; Heubert & Hauser, 1999; Horn, 2003; Mathison & Freeman, 2003; Rex & Nelson, 2004) definition of high-stakes testing offers educational researchers a comprehensive description of the phenomenon. Madaus' definition describes a construct which many researchers, studying mandated testing, have embraced as an all-encompassing and definition of "high-stakes testing" as recognized in educational research:

> High-stakes tests include those used for the certification or recertification of teachers, promotion of students from one grade to the next, award of a high school diploma, assignment of a student to a remedial class, allocation of funds to a school or school district, award of merit pay to teachers on the basis of their students' test performance, certification or recertification of a school district, and placement of a school system in "educational receivership". (p. 30)

However, educational researchers who advocate that high-stakes testing holds great potential to both monitor and increase student achievement share a definition of high-stakes testing as one which has the ability and purpose to act as a lever of change within a contemporary educational reform movement (Grant, 2000). Consequently, the interpretations and a commonly held definition of high-stakes testing has become somewhat of a politically charged and divisive construct, where definitions are manipulated and understandings are blurred. With high-stake tests ultimately and profoundly influencing peoples' lives (Downing & Haladyna, 1996) outside of the stated

NCLB intent of accountability and responsibility it is essential that a common definition is shared among all stakeholders.

Defining High-Stakes Testing: Educational Research

While many scholars and researchers currently debate the definition and usefulness of high-stakes testing, the utilization of large-scale high-stakes tests have become a distinct piece of the educational landscape over the last 30 years (Heubert & Hauser, 1999). Importantly, Heubert and Hauser describe two fundamentally opposing ways of defining tests:

> A low-stakes test has no significant, tangible, or direct consequences attached to the results, with information alone assumed to be a sufficient incentive for people to act. The theory behind this policy is that a standardized test can reliably and validly measure student achievement; that politicians, educators, parents, and the public will then act on the information generated by the test; and that actions based on tests results will improve educational quality and student achievement. In contrast, high-stakes policies assume that information alone is insufficient to motivate educators to teach well and students to perform to high standards. Hence, it is assumed, the promise of rewards or the threat of sanctions is needed to ensure change. Rewards in the form of financial bonuses may be allocated to school or teachers; sanctions may be imposed through external oversight or takeover by higher-level authorities. (pp. 35-36)

Drawing upon research of state-wide testing programs in Pennsylvania and Maryland, Heubert and Hauser concluded that the perceived definition and the level of stakes associated with tests were less characteristics of the test, per se, but rather greater characteristics of the perceptions of test use. These researchers assert that the use of a single indicator of student learning to make high-stakes decisions about tracking, promotions and graduation was unethical. Based on their findings, Heubert and Hauser (1999) contend that "people may attach a level of stakes to a test that is out of character with the formal consequences associated with it" (p. 26) suggesting that the importance of a commonly held definition of high-stakes may be more readily understood and defined at the local level rather than framed within federal mandates.

However, educational researchers with opposing perspectives see this definition of high-stakes testing as problematic; in contrast, they define high-stakes testing as holding major consequences for students, teachers, and schools - calling for a clear understanding of the intent and outcome of President Bush's original initiative of NCLB. This politicization of assessment and accountability is described by Hillocks (2002) in his landmark study of how state assessments control learning. Hillocks defined high-stakes testing as assessments where "the fortunes of individual students, schools, and school districts rise or fall on the results" (p. 18). Similarly, Johnson and Johnson (2006) published an in-depth study of poverty, testing and failure, asserting that high-stakes tests are those which "base life-altering decisions of single test scores" (p. 202).

For these educational researchers high-stakes tests are those tests which critically impact programs, curriculum, and individual student achievement resulting in high-stakes consequences within the educational setting. While researchers may hold specific characteristics of the definition of high-stakes testing in contrast to others, it is essential that the commonly held definition is reflective of current educational research and educational practice, supporting a common construct and purpose of "high-stakes" testing.

Defining High-Stakes Testing: Professional Research Organizations

Professional educational research organizations, recognizing the need to define and state their organizational position regarding high-stakes testing, have issued position papers defining high-stakes testing. Drawing from the 1999 Standards for Educational Psychological Testing, the American Educational Research Association (AERA, 1999) stated, high-stakes test are those which:

> carry serious consequences for students or for educators. Schools may be judged according to the school-wide average scores of their students. High school-wide scores may bring public praise or financial rewards; low scores may bring public embarrassment or heavy sanctions. For individual students, high scores may bring a special diploma attesting to exceptional academic accomplishment; low scores may result in students being held back in grade or denied a high school diploma (p.1).

The National Reading Conference (NRC, 2004) published a comprehensive policy brief "focusing on the popularity of high-stakes tests, the uses and misuses of high-stakes tests and the consequences of high-stakes testing" (p.2). In this policy brief, Afferbach described that:

> [h]igh-stakes (reading) tests are those with highly consequential outcomes for students, teachers, and schools. These outcomes may include promotion or retention, student placement in (reading) groups, school funding decisions, labeling of schools as successful or failing and the of community support for a school. (p. 2)

Additionally, the International Reading Association (IRA, 1999) issued a position paper stating:

> [h]igh-stakes testing means that the consequences for good (high) or poor (low) performance on a test are substantial. In other words, some very important decisions, such as promotion or retention, entrance into an educational institution, teacher salary, or a school district's autonomy depend on a single test score. (p. 2)

Equally, these position statements from leading educational research organizations present a common, defining construct of high-stakes testing within the educational research community.

Defining High-Stakes Testing: Construct Definition

While high-stakes testing continues to be a highly complex and multi-faceted construct, high-stakes testing in educational research refers to the use of standardized testing measures as criteria for improving educational outcomes, determining grade promotion,

graduation, quality of schools, rewards or sanctions, ensuring equal educational opportunities, drawing in public support for schools, as well as many other highly attributable stakes and consequences (Heubert & Hauser, 1999). Although there may be no current agreement or common consensus that a specific test is named a high-stakes test, it is recognized, that high-stakes are not identifiable characteristics of the test itself, but rather the effects of the intended and unintended consequences of the test scores (Nichols & Berliner, 2007; Goertz & Duffy, 2003; Heubert & Hauser, 1999).

Whereas the No Child Left Behind (NCLB) Act of 2001 legitimized the role of high-stakes testing through federal legislation and fuelled the current debates over high-stakes testing, a salient factor contributing to the current reform debates may be that the term "high-stakes testing" does not appear in the hundreds of pages of the NCLB (2002) law (Johnson & Johnson, 2006). Instead this law simply states that responsibility in the form of standards and accountability is the main focus of the national educational goals. How we got from standards and accountability to mandated high-stakes tests across the nation may be at the heart of understanding how high-stakes are recognized and defined by others (Linn, 2003). This controversy has been playing out over the last decade largely in terms of purpose and use of high-stakes testing data. In particular, the lack of a common understanding of this complex construct stands as an opportunity for a common prolegomenon to undergird effective educational reform.

One of the most serious concerns related to an ambiguous and elusive definition of high-stakes testing is that, currently high-stakes testing forms the core curricular foundation that shapes American education policy (Nichols, Glass & Berliner, 2006, 2012; David, 2011). Within the press for accountability, high-stakes educational decisions are made based on a complex and arbitrarily understood construct. "With the growth of testing as a policy strategy, discussions about its use have moved more and more from the technical realm to the political world of electoral campaigns, interest groups, and public opinion" (National Research Council, 1999, p. 43).These critical issues of the high-stakes testing era have resulted in situated, interpretive and polarizing use and purpose of high-stakes test data. One thing is clear, however, high-stakes testing has become the primary instrument for data collection on which educational reform is enacted and reported on (Au, 2009). While the core tenet of any educational reform is to measure student success and effectiveness of schools, this last decade of test-driven accountability has been propelled forward but not broadly or commonly understood. As a result, Allington (2003) argues, it is imperative that with the term high-stakes testing punctuating almost every educational initiative or program, a commonly held definition of use and purpose must be recognized at the core of effective and successful school reform. More recently, Ravitch (2010) cautioned that with the current overreliance on high-stakes test scores and the push to raise the bar, the result has been the marginalization of other important goals of education and overall – "worse education" (p. 230).

Today, under current federal law, test-based accountability systems will very likely remain at the core of current and future educational reform as President Obama Race to the Top plan calls for "national Common Core Standards, national curriculum materials, and high-stakes national tests" (Onosko, 2011, p.1). Plainly, as we move forward, any further use and discussion of high-stakes testing requires a commonly held and understood definition. As the 1990s closed, standards based accountability in the form of high-stakes testing data had become the yardstick for all state educational reform mandates (Goertz, Duffy & Carlson- LeFloch, 2000). Looking back, over the last

decade, it is clear that high-stakes testing in its various forms has not met NCLB target outcomes. Recently, in a statement issued by the US Secretary of Education, current high-stakes test data projects that more than 80 percent of all US public schools will fail to reach their achievement targets in 2012 (Duncan, 2011), and almost every school in America will not meet target outcomes by 2014 (Berliner, 2013).This impoverished view of teaching and learning continues to frame schools as "failing"- giving further warrant and purpose to an ill-purposed used of testing in America's schools (Berliner, 2006; Berliner & Biddle, 1996).

As the nation's teachers and students ready themselves for the coming on slaught of increased high-stakes testing funded by the millions of dollars assigned to the Race to the Top incentives and given the complex relationships between high-stakes testing and teacher pedagogy and practice, high-stake tests warrant closer attention and characterization than they have received thus far (Guisbond & Schaeffer, 2012). While high-stake test scores are firmly entrenched within the press to evaluate student achievement and school success it is critical that the definition and utilization of test data are commonly understood by all within the educational reform movement (Linn, 2000). However, while high-stakes tests may at first glance appear to be a mode to determine common and measurable standards they do not measure within a common metric, as the notion of "stakes" may be defined by perception and therefore stand as highly unstable measure of student success and school effectiveness (Corbett & Wilson, 1999; Heubert, 1999).

Taken together, a composite, operational definition of high-stakes testing among researchers and professional organizations stands as an amalgam to frame education reform measures that aim to improve teaching and learning (NRC, 2004; AERA, 1999; IRA, 1999). Based on historical and current views of this complex construct, a composite, operational definition of "high-stakes testing" may extend to being those tests which influence significant decisions and critically impact programs, pedagogy, practice, curriculum, individual student achievement, teacher identity and resulting in high-stakes consequences within the educational setting (Blake, 2008).

High-Stakes Testing: Moving Forward

While nearly all states had standardized tests in place prior to NCLB, few states attached such high-stakes penalties and sanctions to the data. Again it is critical to highlight that: federal law takes no position on how states and individual school districts enact or make use of high-stakes test data (Heubert, 2002). However, high-stakes testing has informed the familiar experience of educational culture, thereby creating a situation in play where the familiar and common experience of schooling has created a situation in use where familiarity masks the complexity (Lowenberg Ball & Forzani, 2007) and use of the term "high-stakes" testing in schools operating within the standards reform movement. The stakes associated with performing at the proficient level are high, with NLCB mandates requiring all students to be "proficient" on state, standardized high-stakes tests by 2014 , the degree to which these high-stakes tests differ in design, interpretation and use is alarming (Cronin et al., 2005; Burt & Stapleton, 2010).

It is certain that schooling in the twenty-first century will not be measured as a field of study that Dewey (1938) had envisioned but rather as a defined and contained discipline. High-stakes tests framed as predetermined and prescribed knowledge sidestep what Dewey and others have described as the purpose of schooling - where students engage in a critical and self-reflective education (Dewey, 1938; Giroux, 2010; Sleeter, 2005). Recently, Noddings (2010) concluded that "(W)e have sacrificed richness, depth, and creativity to a dull struggle for higher and higher test scores on material that is quickly forgotten when the test is over" (p. 13). While calling for stakeholders to recognize a common conception and definition of high-stakes testing may seem somewhat less than provocative – it is at its heart a call - not to let history repeat itself once again in viewing "test-score surveillance as a central strategy to supposedly motivate educators to develop children in narrow ways for national purposes" (Onosko, 2011, p. 10). Within the current culture of high-stakes testing we have an opportunity – and a responsibility – to move out and away from the myopic view of teaching and learning (Giroux, 2010) to examine closely the "consequential validity" of such a misunderstood and misconstrued construct as is – "high-stakes testing" (Wilhelm, 2005). Sadly, as the nation races to develop the next generation of high-stakes tests our "schools have become stark, dry, and uninviting. They have become standardized and sterilized" (Hampton, 2005, p. 196). Can a (re)conceptualization and common construct of "high-stakes testing" be formalized from experience?

It is vital that as the nation's educational leaders move to craft a new act to replace NCLB that quality and rigorous research fully describes and informs the process of the next generation of "high-stakes" testing. First, I suggest, the develop standardized assessment tools be based on sound theory and practice that "can stand the weight of accountability, determine high-stakes decisions, direct good teaching, and tell where everyone stands in relation to everyone else – and define what it means to be well educated" (Meier, 2002, p.190). Further, a critical analysis of the complex construct of high-stakes testing use must be closely examined to reveal the policies of assessment and measurement practices within the highly charged, political reform mandates of NCLB where "standards", "accountability" and "high-stakes testing" have become "common cultural buzz words linked to standards-based educational reform movements" (Gabriel & Lester, p. 35). Regrettably, defining the federal view and recognizing the epistemological roots of high-stakes testing has settled into a positivist theoretical frame which looks to teaching and learning as something to be measured – rapidly becoming a misconstrued data footrace in the press to develop the next generation high-stakes tests. Without a public or commonly recognized definition of the complexities and utility (Blake, 2008; Linn, 2000) of high-stakes testing, once again, we may be entering yet another decade of "infinite examination and of compulsory objectification" (Foucault, 1995, p. 189) of teaching and learning where students are constructed, tested, and compared – as widgets on a standardized product line (Duddley-Marling, 2012; Kohn, 2000) – determined by a (mis)construed test.

References

Abrams, L. M., & Madaus, G. F. (2003). The Lessons of High-Stakes Testing. *Educational Leadership*, 61(3), 31-35.

Abrams, L. M., Pedulla, J. J., & Madaus, G. F. (2003). Views from The Classroom: Teachers' Opinions of Statewide Testing Programs. *Theory in Practice*, 42(1), 18-29.

Allington, R. L. (1994a). The Schools We Have. The Schools We Need. *The Reading Teacher*, 48, 14-28.

Allington, R. L. (2002). *Big Brother and The National Reading Curriculum: How Ideology Trumped Evidence*. Portsmouth, NH: Heinemann.

Allington, R. L. (2003). High-Stakes Testing Does Not Improve The Teaching Or Learning of Reading. *The New England Reading Association*, 39(1), 31-36.

Amrein, A. L., & Berliner, D. C. (2002a). High-Stakes Testing, Uncertainty, and Student Learning. *Education Policy Analysis Archives*, 10(18) [On-line], http://epaa.asu.edu/epaa/v10n18.

Amrein, A. L., & Berliner, D. C. (2002b). The Impact of High-Stakes Tests On Student Academic Performance: An Analysis of NAEP Results In States With High-Stakes Results and ACT,

SAT, and AP Test Results In States With High School Graduation Exams. [On-line], http://edpolicylab.org.

Au, W. (2011). Teaching Under The New Taylorism: High-Stakes Testing and The Standardization of The 21st Century Curriculum. *Journal of Curriculum Studies,* 43(1), 25-45.

Au, W. (2009). *Unequal By Design: High-Stakes Testing and The Standardization of Inequality*. New York: Routledge.

Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, 36(5), 258-267.

Berliner, D. C. (2013). Sorting Out The Effects of Inequality and Poverty, Teachers and Schooling. In S.L. Nichols (Ed.), *Educational Policy and The Socialization of Youth for The 21st century. New York Teachers College Press*.

Berliner, D. C. (2006). Our Impoverished View of Educational Research. *Teachers College Record*, *108*(6), 949-995.

Berliner, D. & Biddle, B. (1996). *The Manufactured Crisis: Myths, Fraud and The Attack On America's Public Schools*. New York: Basic Books.

Blake, J. (2008). The Consequential Effects of High-Stakes Testing On Teacher Pedagogy, Practice and Identity: Teacher Voices Disrupt The A Priori. Unpublished Dissertation, University of Tennessee.

Bracey, G. (2000). High Stakes Testing. *Center for Education Research, Analysis, and Innovation* (CERAI-00-32) [On-line], http://epicpolicy.org/files/cerai-00-32.htm.

Burt, W. M. & Stapleton, L. M. (2010). Connotative Meanings of Student Performance Labels Used In Standard Setting. *Educational Measurement: Issues and Practice*, 29(4), 28–38.

Callahan, R. (1962). *Education and The Cult of Efficiency*. Chicago: University of Chicago Press.

Cimbricz, S. (2002). State-Mandated Testing and Teachers' Beliefs and Practice. *Education Policy Analysis Archives*, 10(2).

Cohn, M. M., & Kottkamp, R. B. (1993). *Teachers: The Missing Voice In Education*. Albany: State University of New York Press.

Corbett, H. D., & Wilson, B. (1991). *Testing, Reform, and Rebellion*. Norwood, NJ: Ablex.

Crawford, J. & Impara, J.C. (2001). Critical Issues: Current Trends, and Possible Futures In Quantitative Methods. In V. Richardson (Ed.), *Handbook of Research On Teaching* (4th Ed.Pp. 133-173). Washington, D. C.: AERA.

Cronin, J., Kingsbury, G. G., McCall, M. S., & Bowe, B. (2005). *The Impact of The No Child Left Behind Act On Student Achievement and Growth: 2005edition*. Lake Oswego, OR: Northwest Evaluation Association.

Cronin, J., Dahlin, M., Adkins, D., & Kingsbury, G. G. (2007). *The Proficiency Illusion*: Thomas B. Fordham Institute.

Cross, C. T. (2004). *Political Education: National Policy Comes of Age*. New York: Teachers College.

Cuban, L. (2007). Hugging The Middle: Teaching In An Era of Testing and Accountability, 1980-2005. *Education Policy Analysis Archives*, 15(1)[On-line], http://epaa.asu.edu/epaa/v15n1.

Datnow, A., Hubbard, L., & Mehan, H. (2002). *Extending Educational Reform: From One School To Many*. New York: Routledge Falmer.

David, J. (2011). High-Stakes Testing Narrows The Curriculum. *Educational Leadership*, 68(6), p. 78-80.

Dewey, J. (1938). *Experience and Education*. New York: Touchstone.

Dorn, S. (1998). The Political Legacy of School Accountability Systems. *Educational Policy Analysis Archives*, 6(1) [On-line], http://epaa.asu.edu/epaa/v6n1.html.

Downing, S. M., & Haladyna, T. M. (1996). A Model for Evaluating High-Stakes Testing Programs: Why The Fox Should Not Guard The Chicken Coop. *Educational Measurement: Issues and Practice*, 5(1), 5-12.

Duffy, G. G. (2007). Thriving In A High-Stakes Testing Environment. *Journal of Curriculum and Instruction*, 1(1), 7-13.

Duncan, A. (2011). *Winning The Future With Education: Responsibility, Reform and Results.* Oral testimony of US Secretary of Education Arne Duncan given to the US Congress, March 9, 2011. Retrieved August 20, 2012, from http://www.ed.gov/news/speeches/winning-future-education-responsibility-reform-and-results

Dudley-Marling, C. (2012). Return of The Deficit. *Journal of Educational Controversy, (July), from http://www.nycore.org/wp-content/uploads/Deficit-model-and-testing.pdf*

Dutro, E. & Selland, M. (2012).“I Like To Read, But I Know I'm Not Good At It”: Children's Perspectives On High-Stakes Testing In A High-Poverty School. *Curriculum Inquiry, 42(3),*340-367.

Finn, C. E., & Petrilli, M. J. (2007). Forward. In *Proficiency Illusion*. New York: Thomas B. Fordham Institute.

Foucault, M. (1995). *Discipline and Punish: The Birth of The Prison* (A Sheridan, Trans.2nd ed.). New York: Vintage.

Franzak, J. K. (2004). Constructing Struggling Readers: Policy and The Experiences of Eighth-Grade Readers. In J. Worthy, B. Maloch, J. Hoffman, D. Schallert & C. Fairbanks (Eds.), *53rd Yearbook of the National Reading Conference* (pp. 189-205). Oak Creek, WI: National Reading Conference.

Gabriel, R. & Lester, J. (2011). Race To The Top Era of Education Consulting: A Call To Reform The Reformers. *International Journal of Educational Policies, 5*(1), 33-46.

Giroux, H. A. (2010, May). Dumbing Down Teachers: Attacking Colleges of Education Inthe Name of Reform (Part I). *Truthout*.

Goertz, M., Duffy, M., & Carlson-Lefloch, K. (2000). *Assessment and Accountability Systems:50 State Profiles* (CPRE Research Report Series No. Rr–046). Philadelphia, Pa: Consortium for Policy Research in Education.

Goertz, M., & Duffy, M. (2003). Mapping The Landscape of High-Stakes Testing and Accountability Programs. *Theory into Practice*, 42(1), 4-11.

Goldberg, M. F. (2004). The High-Stakes Test Mess. *Education Digest: Essential Readings Condensed for Quick Review*, 69, 8-15.

Grant, S. G. (2000). Teachers and Tests: Exploring Teachers' Perceptions of Changes In The New York State Testing Program. *Educational Policy Analysis Archives*, 8(14).

Guisbond, L., Neill, M., and Schaeffer, R. 2012. *Nclb's Lost Decade for Educational Progress: What Can We Learn from This Policy Failure?* (Boston: Fairtest),

Gunzenhauser, M. G. (2003). High-Stakes Testing and The Default Philosophy. *Theory Intopractice*, 42(1), 51-58.

Guthrie, J. T. (2002). Preparing Students for High-Stakes Test Taking In Reading. In A. E. Farstrup & S. J. Samuels (Eds.), *What Research Has To Say About Reading Instruction* (Pp. 370-389). Newark, Delaware: International Reading Association.

Guthrie, J. W., & Springer, M. G. (2004). *A Nation At Risk* Revisited: Did "Wrong" Reasoning Result In 'Right' Results? At What Cost? *Peabody Journal of Education*, 79(1), 7-35.

Hampton, E. (2005). Differing Perspectives On The Effects of High-Stakes Testing In West Texas. In A. Valenzuela (Ed), Leaving Children Behind: How "Texas-Style"Accountability Fails Latino Youth (Pp. 179-199). Albany, Ny: State University of New York Press.

Heubert, J. P. (2002). *High-Stakes Testing: Opportunities and Risks for Students of Colornenglish-Language Learners, and Students With Disabilities.* Wakefield, Ma: National Center On Accessing The General Curriculum.

Heubert, J. P., & Hauser, R. (Eds.). (1999). *High-Stakes: Testing for Tracking, Promoting, and Graduation*. Washington, D.C.: National Academy Press.

Hilliard, A. G., Iii. . (2000). Excellence In Education Versus High-Stakes Standardized Testing. *Journal of Teacher Education*, 51(4), 293-304.

Hillocks, G. (2002). *The Testing Trap: How State Writing Assessments Control Learning*. New York: Teachers College Press.

Hoffman, J. V., Assaf, L. C., & Paris, S. G. (2001). High-Stakes Testing In Reading: Today In Texas, Tomorrow? *The Reading Teacher*, 54(5), 482-492.

Horn, C. (2003). High-Stakes Testing and Students: Stopping Or Perpetuating A Cycle of Failure? *Theory Into Practice*, 42(1), 30-41.

Johnson, D. D., & Johnson, B. (2006). *High Stakes: Poverty, Testing, and Failure In American Schools* (2nd Ed.). Lanham, Maryland: Rowman & Littlefield

Johnson, S. M. (2002, July 4, 2008). What Teachers Need In A High-Stakes World. *Harvard Graduate School of Education News*.

Johnston, P. H. (1998). The Consequences of The Use of Standardized Tests. In S. Murphy (Ed.), *Fragile Evidence: A Critique of Reading Assessments*. Mahwah, Nj: Lawrence Erlbaum.

Jones, M. G., Jones, B. D., & Hargrove, T. Y. (2003). *The Unintended Consequences of High-Stakes Testing*. Lanham, Maryland: Rowman & Littlefield.

Kohn, A. (2000). *The Case Against Standardized Testing: Raising The Scores, Ruining The Schools.*Portsmouth, Nh: Heinemann.

Kohn, A. (2002). Standardized Testing: Separating The Wheat Children from The Chaff Children. In S. Ohanian (Ed.), *What Happened To Recess and Why Are Our Children Struggling In Kindergarten.* New York, Ny: McGraw-Hill.

Koretz, D. (2008). *Measuring Up: What Educational Testing Really Tells Us*. Cambridge, Ma: Harvard University Press.

Kozol, J. (2005). *The Shame of The Nation: The Restoration of Apartheid Schooling In America.*New York: Crown Publishers.

Krosnick, J. A., & Schuman, H. (1988). Attitude, Intensity, Importance, and Certainty and Susceptibility To Response Effects. *Journal of Personality and Social Psychology*, 54, 940-952.

Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Linn, R. L. (2000). Assessments and Accountability. *Educational Researcher*, 29(4), 4-16.

Linn, R. L. (2003). Accountability: Responsibility and Reasonable Expectations. *Educational Researcher*, 32(7), 3-13.

Lortie, D. C. (1975). *Schoolteacher: A Sociological Study*. Chicago, Il: University of Chicago Press.

Lowenberg Ball, D., & Forzani, F. M. (2007). What Makes Education Research "Educational"? *Educational Researcher*, 36(9), 529-540.

Madaus, G. F. (1988a). The Distortion of Teaching and Testing: High-Stakes Testing and Instruction. *Peabody Journal of Education*, 65(3), 29-46.

Mathison, S., & Freeman, M. (2003, September 19). Constraining Elementary Teachers' Work: Dilemmas and Paradoxes Created By State Mandated Testing. *Educationpolicy Analysis Archives*, *11*(34).

Mathison, S., & Freeman, M. (2006). Teacher Stress and High-Stakes Testing, How Using One Measure of Academic Success Leads To Multiple Teacher Stressors. In R. G. Lambert & C. J. Mccarthy (Eds.), *Understanding Teacher Stress In An Age of Accountability*. Greenwich, Ct: Information Age Publishing.

Mccarthey, S. J. (2008). The Impact of No Child Left Behind On Teachers' Writing Instruction. *Written Communication*, 25(4), 462-505.

Mcgill-Franzen, A. (2000). Policy and Instruction: What Is The Relationship? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson & R. Barr (Eds.), *Handbook of Reading Research* (Vol. Iii, Pp. 889-908). Mahwah: Nj: Erlbaum Associates.

Mcgill-Franzen, A., & Allington, R. L. (2006). Contamination of Current Accountability Systems. *Phi Delta Kappan*, 87(10), 762-766.

Mehan, H. (1990). The School's Work of Sorting Students. In D. Boden & D. H. Zimmerman (Eds.), *Talk and Social Structure*. Cambridge: Polity Press.

Mehrens, W. A., & Popham, W. J. (1992). How To Evaluate The Legal Defensibility of High-Stakes Tests. *Applied Measurement In Education*, 5, 265-283.

Meier, D. (2002). Standardization Versus Standards. *Phi Delta Kappan,84,(3),190-198.*

Messick, S. L. (1995). Standards of Validity and The Validity of Standards In Performance Assessment. *Educational Measurement: Issues and Practice, 14*(4), 5–8.

Messick, S. (1989). Meaning and Values In Test Validation: The Science and Ethics of Assessment. *Educational Researcher, 18*(2), 5-11.

National Education Policy Center. (2010). Reviews of The Obama Administration's Six Research Summaries for "A Blueprint for Reform."

National Research Council. (1999). *High Stakes: Testing for Tracking, Promotion, and Graduation* (J. P. Heubert & R. M. Hauser, Eds.). Washington, Dc: National Academy Press.

Nichols, S. L., & Berliner, D. C. (2005). The Inevitable Corruption of Indicators and Educators Through High-Stakes Testing. *Education Policy Studies Laboratory* [On-Line], Http://Edpolicylab.Org.

Nichols, S. L., & Berliner, D. C. (2007). *Collateral Damage: How High-Stakes Testing Corrupts America's Schools*. Cambridge, Massachusetts: Harvard Education Press.

Nichols, S. L., & Berliner, D. C. (2008). Why Has High-Stakes Testing So Easily Slipped Into Contemporary American Life? *Phi Delta Kappan, 89*(9), 672-676.

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(1-175).*

Nichols, S. L., Glass, G. V, & Berliner, D. C. (2012). High-Stakes Testing and Student Achievement: Updated Analyses With Naep Data. *Education Policy Analysis Archives, 20* (20).

Noddings, N. (2010, August 31). Teacher Tess In Testing Land. *Education Week.*

Onosko, J. (2011). Race To The Top Leaves Children and Future Citizens Behind: The Devastating Effects of Centralization, Standardization, and High-Stakes Accountability. *Democracy and Education, 19*(2), 1-11.

Paris, S. G., Lawton, T. A., Turner, J. C., & Roth, J. L. (1991). A Developmental Perspective On Standardized Achievement Testing. *Educational Researcher*, 20(5), 12-20+40.

Patton, M. (1990). *Qualitative Evaluation and Research Methods* (2nd Ed.). Thousand Oaks, Ca: Sage.

Pedulla, J. J., Abrams, L. M., Madaus, G. F., Russell, M. K., Ramos, M. A., & Miao, J. (2003). *Perceived Effects of State-Mandated Testing Programs On Teaching and Learning: Findings from A National Survey On Teachers*. Chestnut Hill, Ma: Boston College: Center for The Study of Testing, Evaluation, and Educational Policy.

Perie, M. (2008). A Guide To Understanding and Developing Performance-Level Descriptors. *Educational Measurement: Issues and Practice*, *27*(4), 15–29.

Perkinson, H. J. (1995). *The Imperfect Panacea: American Faith in Education*. Boston: McGraw-Hill.

Ravitch, D. (2010). *The Death and Life of The Great American School System: How Testing and Choice Are Undermining Education.* New York: Basic Books.

Resnick, D., & Resnick, L. (1985). Standards, Curriculum, and Performance: A Historical and Comparative Perspective. *Educational Researcher*, 14(4), 5-20.

Rex, L. A., & Nelson, M. C. (2004). How Teachers' Professional Identities Position High-Stakes Test Preparation In Their Classrooms. *Teachers College Record*, 106(6), 1288-1331.

Riddle Buly, M., & Valencia, S. W. (2002). Below The Bar: Profiles of Students Who Fail State Reading Assessments. *Educational Evaluation and Policy Analysis*, 24(3), 219-239.

Shepard, L. A. (2000). The Role of Assessment In A Learning Culture. *Educational Researcher*, 29(7), 4-14.

Sleeter, C. E. (2005). *Un-Standardizing Curriculum: Multicultural Teaching In The Standards-Based Classroom.* New York, NY: Teachers College, Columbia University.

Smith, M. L., & Fey, P. (2000). Validity and Accountability In High-Stakes Testing. *Journal of Teacher Education*, 51(5), 334-344.

Spillane, J. (2004). *Standards Deviation: How Schools Misunderstand Education Policy*. Cambridge, Massachusetts: Harvard Education Press.

Spring, J. (2004). *American Education*. New York, NY: McGraw-Hill.

Stone, D. (2002). *Policy Paradox: The Art of Political Decision Making*. New York: W. W. Norton & Company.

Thomas, R. M. (2005). *High-Stakes Testing: Coping With Collateral Damage*. Mahwah, New Jersey: Lawrence Erlbaum

Tyack, D. B., & Cuban, L. (1996). *Tinkering Toward Utopia: A Century of Public School Reform*. Cambridge, Ma: Harvard University Press.

U.S. Department of Education. 2012. Race To The Top -- Early Learning Challenge. http://www2.Ed.Gov/Programs/Racetothetop-Earlylearningchallenge/Index. html.

Valli, L., Croninger, R. G., Chambliss, M. J., Graeber, A. O., & Buese, D. (2008). *Test Driven: High-Stakes Accountability In Elementary Schools*. New York: Teachers College Press.

Wilhelm, J.D. (2005). Consequential Validity: Weighing The Sheep, Or Feeding Them, Too? *Voices from The Middle*, 13(2), 41.

About The Author

Jan E. Blake is an Assistant Professor in the Reading and Childhood Education in the College of Education at the University of South Florida St. Petersburg. Her research interests include teacher identity, preparation and development; literacy studies and; education policy. Jan Blake is a former school teacher, special educator and curriculum specialist and has taught all grades K-12.

International Journal of Educational Policies
ISSN: 1307-3842
http://ijep.icpres.org
http://ojs.ijep.info