The International Education Journal: Comparative Perspectives, 2013, 12(2), 7–19
ISSN 1443-1475 © 2013 www.iejcomparative.org

Hiding behind high-stakes testing: Meritocracy, objectivity and inequality in U.S. education

Wayne Au
University of Washington, Bothell

This paper analyses how high-stakes, standardised testing became the policy tool in the U.S. that it is today and discusses its role in advancing an ideology of meritocracy that fundamentally masks structural inequalities related to race and economic class. This paper first traces the early history of high-stakes testing within the U.S. context, focusing on its deep-rooted connections with eugenics and IQ testing in schools. It then turns to the more recent history of high-stakes testing, highlighting the ways that race and class inequality, as well as the ideology of meritocracy, manifest in the United States today as part of a legacy of inequality.

Keywords: high-stakes testing, inequality, meritocracy, standardised testing, assessment subjectivity

High-stakes, standardised testing has become ubiquitous in the United States, where, since the passing of the No Child Left Behind Act (United States Congress, 2002), all U.S. states were mandated to test public school students in grades 3-8, and once in high school, be tested in reading and math, with future provisions for students to also be tested in science. High-stakes, standardised testing has only gained more traction in the U.S. with the Obama Administration's "Race To The Top" initiative, and the impending implementation of national standards vis-à-vis the Common Core (Karp, 2010). In this paper I examine how high-stakes testing became the policy tool in the U.S. that it is today and discuss its role in advancing an ideology of meritocracy that fundamentally masks structural inequalities related to race and economic class. I begin here by tracing the early history of high-stakes testing within the U.S. context, focusing on its deep-rooted connections with eugenics, social engineering, and social efficiency vis-à-vis school tracking. I then sketch the more recent history of high-stakes testing, highlighting the ways that race and class inequality, as well as the ideology of meritocracy, manifest in the United States today as part of a legacy of inequality.

A U.S. APPROPRIATION OF A FRENCH INVENTION

High-stakes, standardised testing in the United States itself began as a recontextualisation of an assessment tool from another country: France. Originally, French psychologist Alfred Binet first developed the IQ test in 1904 to assess if young children were mildly developmentally disabled, producing the "Binet Scale" of intelligence. By dividing the mental age by the second (chronological age), the idea of "intelligence quotient," or IQ was born (Gould, 1996). According to Gould (1996), this testing was to be used specifically on young children only, and it was conceived purely as a practical tool for placement not related to any idea of hereditary or innate intelligence. However, U.S. cognitive psychologists like Henry Goddard, Lewis Terman, and Robert Yerkes recontextualised Binet's testing and measurement of IQ in very specific ways that fit the race and class politics of the United States at the turn of the 20th century. Mainly they distorted the original use of the tests and injected their own underlying presumptions about humans and human ability, presumptions that had very little to do with Binet (Au, 2009b; Gould, 1996). Through the work of these psychologists, and with the explicit support of educational philanthropists like Carnegie (Karier, 1972), IQ in the United States became conceived of as hereditary and fixed, laying the groundwork to use standardised testing to justify the sorting and ranking of different people by race, ethnicity, gender, and class according to supposedly inborn, biologically innate intelligence (Au, 2009b; Gould, 1996).

In 1917, as a psychologist and Army Colonel in charge of the mental testing of 1.75 million recruits during World War I, Yerkes worked with Goddard, Terman, and others to develop the Alpha and Beta Army tests to sort incoming soldiers and to determine their "mental fitness". Yerkes drew several dubious conclusions using this incredibly large pool of data, including that the intelligence of European immigrants could be judged according to their country of origin: The darker peoples of eastern and southern Europe were less intelligent than their fairer-skinned, western and northern European counterparts, and that African Americans were the least intelligent of all peoples (Giordano, 2005). As Karier (1972) explains such testing had deep seated bias built-in:

> Designing the Stanford-Binet intelligence test, Terman developed questions which were based on presumed progressive difficulty in performing tasks which he believed were necessary for achievement in ascending the hierarchical occupational structure. He then proceeded to find that according to the results of his tests the intelligence of different occupational classes fit his ascending hierarchy. It was little wonder that IQ reflected social class bias. It was, in fact, based on the social class order (pp. 163-164).

With the explicit support from these psychologists, eugenicists of the time rallied around the idea that race mixing was spreading the supposedly inferior intelligence genes of African Americans, other non-white peoples, and immigrants (Selden, 1999).

It is important to note that, among others, African American educators were acutely aware of the racism inherent in both the eugenics and IQ testing movements in the U.S. For instance, in 1940, W.E.B. DuBois recalled:

> It was not until I was long out of school and indeed after the (first) World War that there came the hurried use of the new technique of psychological tests, which were quickly adjusted so as to put black folk absolutely beyond the possibility of civilization (as quoted in Guthrie, 1998, p. 55).

Indeed, as Stoskopf (1999) explains, the lower scores of African Americans were regularly used to track Black students into vocational education or for White teachers to explain away any difficulties these students might be having in their classrooms. One of the earliest African American educators to publicly challenge the findings of prominent psychologists involved in the IQ testing and eugenics movements was Horace Mann Bond—the Director of the School of Education at Langston University in Oklahoma. In 1924 Bond critiqued IQ testing and eugenics in *Crisis,* the magazine of the National Association for the Advancement of Colored People.

Despite resistance from African Americans and others in the United States, standardised IQ testing soon found its way into the institution of education, and thus gave rise to systems of academic tracking. As Tyack (1974) explains:

> Intelligence testing and other forms of measurement provided the technology for classifying children. Nature-nurture controversies might pepper the scientific periodicals and magazines of the intelligentsia, but schoolmen found IQ tests an invaluable means of channeling children; by the very act of channeling pupils, they helped to make IQ prophecies self-fulfilling (p. 180).

Then a Stanford University professor of psychology, and under the sponsorship of the National Academy of Sciences, Terman played a key role in adapting the above mentioned army tests into the National Intelligence Tests for school children in 1919, and by 1920 over 400,000 copies of these tests had been sold nationwide. Terman and others also created the Stanford Achievement Test in 1922, and by late 1925, he reported sales of this test to be near 1.5 million copies. Further, a 1925 survey of 215 cities with populations over 10,000 found that 64% of these cities used intelligence tests to classify and sort elementary students, 56% used the tests to classify and sort junior high school students, and 41 did the same for high school students. Another survey of superintendents of school districts in cities with populations over 10,000 people, completed in 1926, produced similar results (Chapman, 1988).

By 1932, 112 of 150 large city school systems in the United States had begun to use intelligence testing to place students into ability groups, and colleges had also begun to use these tests to justify admissions as well (Haney, 1984). As Karier (1972) explains:

> It was men like Thorndike, Terman and Goddard, supported by corporate wealth, who successfully persuaded teacher, administrators and lay school boards to classify and standardize the school's curriculum with a differentiated track system based on ability and values of the corporate liberal society (p. 166).

The "values of the corporate and liberal society" to which Karier refers speaks to the ways that standardised testing was seen as a key to liberal notions of individual equality. Fundamentally, early standardised testing in the United States was viewed as providing a completely objective and value free measurement of human intellect

(Au, 2009b). This view then extended into the logics of how such testing was used. For instance, early creators of the SAT exam, a test often used for entrance into U.S. universities, saw this as a way to challenge entrenched class privileges that gave the rich advantages in attaining higher education. The logic being that a test that objectively measured individuals would give everyone a fair and equal chance at getting to college according to their individual hard work and merit (Lemann, 1999; Sacks, 1999).

The presumed objectivity of standardised testing was similarly applied to school structures in the United States. For instance, educational leaders such as John Franklin Bobbitt thought that schools should be structured to prepare students for their future social roles, and to do so would be to bring schools in line with the ideas of "social efficiency"—that is, for schools to sort children efficiently for their presumed futures either as rich or poor, owners or labourers (Au, 2009b; Kliebard, 2004). Further, Bobbitt (1912) and others thought that structuring U.S. schools like industrial factories, with students as the raw materials and teachers as the assembly line workers, was the best way to achieve their goals. Once again, such thinking was based on the assumptive objectivity of standardised testing, for if the measures of students were accurate then students would be given an education appropriate for their role in society and the economy (Au, 2009b; Kliebard, 2004). Consequently, as discussed above, the assumptive objectivity of standardised testing was thus used to "scientifically" declare the poor, immigrants, women, and non-whites in the U.S. as mentally inferior, and to justify educational systems that mainly reproduced extant socio-economic inequalities.

MODERN-DAY HIGH-STAKES TESTING IN THE U.S.

The modern, high-stakes, standardised testing movement in the United States can effectively be traced back to the publication of *A Nation At Risk* (National Commission on Excellence in Education, 1983). This report triggered a wave of reforms: 54 state level commissions on education were created within a year of its publication. Within three years of publication 26U.S. states raised graduation requirements and 35 instituted comprehensive education reforms that revolved around testing and increased course loads for students (Kornhaber & Orfield, 2001). By 1994, 43 states implemented state-wide assessments for K-5, and by the year 2000 every U.S. state but Iowa administered a state mandated test (Jones, Jones, & Hargrove, 2003). Within the first week of taking office in 2001, President G.W. Bush pushed for federal Title I funding to be tied to student test scores (Kornhaber & Orfield, 2001).

High-stakes testing has always been supported by both major political parties in the United States, and in 2002 the U.S. government passed the No Child Left Behind Act (NCLB) into law (United States Congress, 2002). As a policy, NCLB relies upon high-stakes testing as the central mechanism for school reform, mandating that all students be tested in reading and math in grades 3-8 and once in high school, with future provisions that students be tested at least once at the elementary, middle, and high school levels in science. If schools do not show consistent growth on these tests in subgroups related to race, economic class, special education, and English language

proficiency, among others, they face sanctions such as a loss of federal funding, with the ultimate policy goal of all students reaching 100% proficiency by 2014 (Karp, 2006). NCLB represents the culmination of a 20-year trajectory of education policy that centred on high-stakes, standardised testing as the tool for enforcing educational reform in the United States.

Despite the end of the George W. Bush administration, growing public criticism of specific aspects of NCLB (e.g., the Adequate Yearly Progress provision and it being an unfunded federal mandate, amongst others), and campaign rhetoric about the need for multiple measures of student learning and teacher evaluation (Au, 2009a), the election of President Barack Obama has only intensified the use of high-stakes, standardised tests within education policy in the U.S. Nowhere is this more evident than in President Obama's selection of Arne Duncan to lead the Department of Education and the subsequent promotion of the federal "Race To The Top" program, which included monies for more testing as part of a broader education reform package promoting the flawed use of tests to evaluate teachers, attacks on teachers unions' right to collective bargaining, and the proliferation of charter schools (Kumashiro, 2012). At this point most observers would agree that the use of high-stakes testing has become a matter of widespread common sense (Apple, 2006; Kumashiro, 2008) in educational policy in the United States. The presumed objectivity and general "goodness" of using high-stakes tests to drive education reform suggests that such testing is expected to remain regardless of whom is in political power.

HIGH-STAKES TESTING AND RACIALISED INEQUALITY

Achievement gaps in public education amongst different racial, cultural, and economic groups are a significantly pressing problem in the United States, one that has been persistent over time (Ladson-Billings, 2006). The closing of such test score gaps and working towards educational equality has remained the stated impetus behind every reauthorisation of the Elementary and Secondary Education Act of 1965 in the United States (Jennings, 2000), of which NCLB is a manifestation. Despite such stated intentions, analyses of high-stakes, standardised test data has found that the high-stakes testing policies have not improved reading and math achievement across states and have not significantly narrowed national and state level achievement gaps between white students and non-white students or gaps between rich and poor students (National Research Council, 2011). For instance, dropout rates associated with high-stakes tests are disproportionately high for African American and Latino students. In the U.S. state of Texas, while a 0% dropout rate was reported as proof of the success of their system of high-stakes testing, it was later found that low achieving students, mostly African American and Latino, had instead been "disappeared" from the rosters by school officials in order to boost test scores. The reality in Texas is that up to 50% of African American and Latino students who start the 9th grade do not make it through the final, 12th year. When Massachusetts implemented a high-stakes test-based accountability system in the 1990s, it witnessed a 300% increase in dropouts, and with the implementation of a graduate exit exam, it saw a 4% decline in graduating students.

In both cases the drop-outs and the exit exam failures were disproportionately African American and Latino (Darling-Hammond, 2007).

The historical roots of high-stakes, standardised testing in racism, nativism, and eugenics raises a critical question: why is it that, now over 100 years after the first standardised tests were administered in the United States, we have virtually the same test-based achievement gaps along the lines of race and economic class? Given the historical origins of standardised testing in the social efficiency movement, which sought to educate students according to perceived future social roles (Kliebard, 2004), IQ testing, and the eugenics movement, there is no reason to believe that these testing systems could shake off their racist and classist legacies so easily. For instance, Herrnstein & Murray (1996) claimed that there was a hierarchical ordering of races where African Americans were the least intelligent of all races, followed by Latinos, Whites, and Asian Americans who, according to the authors, were purported to be the most intelligent. Echoing the work of early American psychologists from 100 years ago, Hernstein and Murray based their conclusions on an analysis of standardised test scores. Despite the substantive, critical responses rejecting the arguments put forth in *The Bell Curve* (see, e.g., Fraser, 1995), the racist, eugenicist position is still popularly upheld by some scholars. For instance, Rushton and Jensen (2005) in their analysis of "Thirty years of research on race and cognitive ability," assert that there are genetically based racial differences in IQ (Jensen is professor emeritus of educational psychology at University of California, Berkeley). Others, such as Barrow and Rouse (2006), a senior economist at the Federal Reserve Bank of Chicago and a professor of economics and public affairs at Princeton respectively, examined the relationship between education, race, and pay. In their study they explicitly rely on the work of Hernstein and Murray (1996) as a baseline for their analysis. The ghosts of eugenicists and the standardised intelligence test-makers from the early 20th century still haunt us via the very racialised and class-disparate outcomes of the modern day, high-stakes, standardised testing movement. Further, that the analyses of Hernstein and Murray (1996), Rushton and Jensen (2005), and Barrow and Rouse (2006) are taken seriously in contemporary public debates clearly illustrates the ideological and historical grounding of U.S. high-stakes testing in race and class-based inequality as well as the eugenics movement here.

MERITOCRACY AND U.S. HIGH-STAKES TESTING

Historically, standardised testing in the United States has been positioned in a dual, seemingly contradictory ideological role. As noted above, based upon the presumed objectivity of the tests, psychologists, philanthropists, and educators saw the tests as a way to accurately sort students based on measured ability (which conflated with ethnicity, race, and class), and thus served ideologically to justify existing socio-economic inequalities. Ironically, drawing on the same presumption of objectivity, early advocates of testing also saw standardised testing as a means of challenging class privilege. As Sacks (1999) explains, "Standardised testing, rendered with complete objectivity and couched in terms of an empirical 'science,' would be the death knell to

the insidious influences of class privilege perpetuated by the blueness of one's blood (p. 264)." Under the assumption that standardised tests provide fair and objective measurement of individuals, such testing seemingly held the promise that every test taker is offered a fair and equal shot at educational, social, and economic achievement. Problems like racism and class privilege are thus supposedly ameliorated through testing.

This characterisation of standardised testing then (and high-stakes, standardised testing now) as a means of challenging inequality is rooted in the ideal that the United States operates as a meritocracy. That is to say, regardless of social position, economic class, gender, or culture (or any other form of difference), all have an equal chance at becoming "successful" based purely on individual merit and hard work – which by extension also means that any failure is simply due to the individual's own deficit (Lemann, 1999; Sacks, 1999). Thus, as Karier (1972) explains:

> Most testers refused to admit the possibility that they were, perhaps, servants of privilege, power and status, and preferred instead to believe and "hope" that what they were measuring was, in fact, true "merit." This was also an act of faith, a faith based on the belief that somehow the "prestige hierarchy of occupations" and the people in it who provided the objective standard upon which the tests were based, were there not because of privilege, wealth, power status and violence, but because of superior talent and virtue. This was a fundamental axiom in the liberal's faith in meritocracy which emerged in twentieth century American education (p. 169).

Consequently, the ideology of meritocracy masks structural inequalities under the guise of "naturally" occurring aptitude amongst individuals (Bisseret, 1979). Vis-à-vis the ideology of meritocracy, the low achievement on standardised tests of working class people, non-white populations, and some immigrant groups can then be simply and neatly attributed to the failure of individual students, individual groups, or individual cultures, and not attributed to existing structural inequalities.

However, the idea of individuals freely competing based on their own merit to achieve in the realm of education has not been born out by the reality of standardised testing. As Berliner (2012) explains, test scores in the U.S. are more determined by structural conditions affecting students than individual effort:

> Virtually every scholar of teaching and schooling knows that when the variance in student scores on achievement tests is examined along with the many potential factors that may have contributed to those test scores, school effects account for about 20% of the variation in achievement test scores….
>
> On the other hand, out-of-school variables account for about 60% of the variance that can be accounted for in student achievement. In aggregate, such factors as family income; the neighborhood's sense of collective efficacy, violence rate, and average income; medical and dental care available and used; level of food insecurity; number of moves a family makes over the course of a child's school years;…provision of high-quality early education in the neighborhood; language

> spoken at home; and so forth, all substantially affect school achievement (n.p.).

Socio-economic factors simply have an overwhelming effect on educational achievement, and this reality is effectively masked by the ideology of meritocracy embedded in high-stakes testing in the United States.

The meritocratic assumptions of high-stakes testing in the U.S. are also belied by many of the logics that underpin the tests themselves. For instance, akin to systems of capitalist economics, systems of accountability built upon high-stakes, standardised testing cannot function if everyone is a "winner" – for both ideological and technological reasons. Ideologically, if everyone passed the tests there simply would be no way to justify elite status or any form of disparity of education performance at all: every student would qualify for the most elite colleges and jobs, thereby rendering the very hierarchy of elitism obsolete. While this might sound good to many (myself included), there is a long history of elite groups resisting U.S. educational reforms that challenge elite status (Au, 2005; Oakes, Welner, Yonezawa, & Allen, 1998), such that it is more than unlikely real educational equality could be reached by levelling out status hierarchies established and maintained vis-à-vis high-stakes testing.

Technically speaking, the statistical logic of standardised tests requires some students to fail (Popham, 2001). Further, if everyone passed a standardised test (or in the language of the U.S.'s NCLB policy, if all students achieved 100% proficiency), the results of that test would immediately be called into question on technical grounds (e.g., there is something wrong with the test itself), on ethical-political grounds (e.g., someone must have cheated), or on both. Further, 100% passing on a standardised test is a statistical impossibility that no country or system anywhere has ever accomplished (Linn, 2003).

These ideological and technical points are particularly important when it comes to understanding how high-stakes testing fits into the discourse of race and class issues in the education reform movement in the United States, a movement which explicitly seeks to close racial and economic achievement gaps in high-stakes testing scores. One of the great ironies about this discourse is that closing the achievement gap does not mean having everyone be successful on high-stakes tests. Rather, closing the achievement gap actually means having proportional rates of failure and success amongst different groups. If education in the U.S. closed the high-stakes test score achievement gap amongst different groups it would simply mean that equal numbers of rich kids and poor kids pass and fail, equal numbers of white kids and African American kids pass and fail, etc. If high-stakes tests are the sole measure of equality in educational achievement, then at best we in the U.S. can only hope to evenly distribute failure across groups.

THE NON-OBJECTIVITY OF U.S. HIGH-STAKES TESTING

Seeing high-stakes, standardised testing as meritocratic in the United States also assumes that such testing is objective: testing cannot be a measure of individual hard work and merit without assuming that such measurement is accurate, objective, and essentially bias free. However, when we look closely at both the numbers and the tests we can see how this presumed objectivity is far from reality. For instance, a U.S. Department of Education's National Centre for Education Statistics report (Schochet & Chiang, 2010) found a statistical error rate of 35% when using one year's worth of test data to measure a teacher's effectiveness, and an error rate of 25% when using data from three years. Other research in the U.S. has found that one time, randomly occurring factors like whether or not a child ate breakfast on test day; if a window was open and a distracting dog was barking outside during the test; whether or not a child got into an argument with parents or peers on the way to school; which other students happened to be in attendance while taking the test; whomever happened to be administering the test; etc., accounts for 50-80% of any gains or losses on a given student's standardised test score (Kane & Staiger, 2002). While there are other significant technical problems with the objective accuracy of standardised testing (Baker et al., 2010), these two research examples highlight just how problematic the use of these tests are when used to make high-stakes decisions regarding the performance of students, teachers, schools, etc., in the United States: the tests simply are not as accurate as assumed.

The lack of objectivity also shows up in the scoring of the high-stakes tests themselves. As Farley (2009b), a former employee in the U.S. testing industry of 15 years reflects:

> …[T]he test-scoring industry cheats. …It cheats on qualification tests to make sure there is enough personnel to meet deadlines/get tests scored; it cheats on reliability scores to give off the appearance of standardization even when that doesn't exist; it cheats on validity scores and calibration scores and anything else that might be needed. …Statistical tomfoolery and corporate chicanery were the hallmark of my test-scoring career, and while I'm not proud of that, it is a fact. Remember, I was never in the testing business for any reason other than to earn a pay check, just like many of the testing companies are in it solely to make a buck (n.p.).

Farley's experience working in the testing industry was not anomalous, and the lack of objectivity is perhaps most clear when it comes to the grading of standardised writing tests specifically. Like Farley, DiMaggio (2010) worked as a writing test grader for the U.S. testing company, Pearson, for several years and explains what it is like to be a part time worker on a temporary contract there:

> In test-scoring centers, dozens of scorers sit in rows, staring at computer screens where students' papers appear (after the papers have undergone some mysterious scanning process). I imagine that most students think their papers are being graded as if they are the most important thing in the world. Yet every day, each scorer is expected to read hundreds of papers. So for all the months of preparation

> and the dozens of hours of class time spent writing practice essays, a student's writing probably will be processed and scored in about a minute.
>
> Scoring is particularly rushed when scorers are paid by piece-rate, as is the case when you are scoring from home, where a growing part of the industry's work is done. At 30 to 70 cents per paper, depending on the test, the incentive, especially for a home worker, is to score as quickly as possible in order to earn any money… (n.p.).

Perhaps even worse, DiMaggio explains how he and other scorers were told to change their scores in order to create results consistent with the previous year's tests:

> Usually, within a day or two, when the scores we are giving are inevitably too low (as we attempt to follow the standards laid out in training), we are told to start giving higher scores, or, in the enigmatic language of scoring directors, to "learn to see more papers as a 4." For some mysterious reason, unbeknownst to test scorers, the scores we are giving are supposed to closely match those given in previous years. So if 40 percent of papers received 3s the previous year (on a scale of 1 to 6), then a similar percentage should receive 3s this year (n.p.).

Similar stories have been chronicled in detail by Farley (2009a) and have been reported at Salon.com, the *New York Times,* and the *Minneapolis City Pages*. These stories and the myriad of technical issues highlight the non-objectivity of high-stakes, standardised test scores.

CONCLUSION

High-stakes, standardised testing, once adopted within the United States first as standardised testing and later with high-stakes attached, took on a dual role of both legitimating and masking structural race and class inequalities. As Karier (1972) notes, this makes perfect sense within the context of U.S. history:

> The nativism, racism, elitism and social class bias which were so much a part of the testing and Eugenics Movement in America were, in a broader sense, part of the *zeitgeist* which was America. This was the land of the Ku Klux Klan, the red scare, the Sacco-Vanzetti and Scopes trials as well as the land of real opportunity for millions of immigrants. It was this kind of contradictory base in which the corporate liberal state took firm root, building a kind of meritocracy that even Plato could not envisioned. Just as Plato ascribed certain virtues to certain occupational classes, so too, Lewis Terman assigned numbers which stood for virtue to certain occupational classes. It was clear to Terman that America was the land of opportunity, where the best excelled, and the inferior found themselves on the lower rungs of the occupational order (p. 163).

Given the racism, class inequality, and other forms of structural oppression present in the United States of the time (Zinn, 1995), and also given the positivistic objectivity ascribed to the "scientific" measurement of humans through such testing, the use of standardised tests in this manner would almost seem inevitable: standardised tests

have simply proven to be effective in justifying elitism. Again, Karier (1972) proves insightful:

> The many varied tests, all the way from I.Q. to personality and scholastic achievement, periodically brought up-to-date, would serve a vital part in rationalizing the social class system. The tests also created the illusion of objectivity which, on the one side served the needs of the "professional" educators to be "scientific," and on the other side served the need of the system for a myth which could convince the lower classes that their station in life was part of the natural order of things (p. 167).

Further, when we honestly confront the present-day reality of persistent, test-defined race and class-based inequality, an inequality that nearly mirrors the general outcomes of the last 100-plus years of high-stakes, standardised testing in the United States (Madaus & Clarke, 2001), we are essentially left with a choice between one of two possibilities. Either the tests are providing objective and accurate measures of human intelligence and learning (thereby potentially validating the claims of eugenicists and those that believe in the biology of IQ), or the tests are neither objective nor accurate and may in fact be contributing to the very inequality they are purporting to measure (Au, 2009b). Indeed, as I've highlighted here, high-stakes standardised tests are far from being the objective measures that proponents in the U.S. would have us believe. As such, both historically and contemporarily, high-stakes, standardised testing has functions to mask the reality of structural race and class inequalities in the United States.

REFERENCES

Apple, M. W. (2006). *Educating the "right" way: Markets, standards, god, and inequality* (2nd ed.). New York: Routledge.

Au, W. (2005). Power, identity, and the third rail. In P. C. Miller (Ed.), *Narratives from the classroom: An introduction to teaching* (pp. 65-85). Thousand Oaks, California: Sage Publications.

Au, W. (2009a). Obama, where art thou?: Hoping for change in U.S. education policy. *Harvard Educational Review*, 79(2), 309-320.

Au, W. (2009b). *Unequal by design: High-stakes testing and the standardization of inequality*. New York: Routledge.

Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., Shepard, L. A. (2010). Problems with the use of student test scores to evaluate teachers. *Economic Policy Institute*, Briefing Paper #278. Retrieved from http://www.epi.org/files/page/-/pdf/bp278.pdf.

Barrow, L., & Rouse, C. E. (2006). The economic value of education by race and ethnicity. *Economic Perspectives*, 30(2), 14-27.

Berliner, D. C. (2012). Effects of inequality and poverty vs. teachers and schooling on America's youth. *Teachers College Record*, 116(1). Retrieved from http://www.tcrecord.org (2014 paper publication).

Bisseret, N. (1979). *Education, class language and ideology*. Boston: Routledge & Kegan Paul.

Bobbitt, J. F. (1912). The elimination of waste in education. *The Elementary School Teacher*, 12(6), 259-271.

Chapman, P. D. (1988). *Schools as sorters: Lewis M. Terman, applied psychology, and the intelligence testing movement, 1890-1930*. New York: New York University Press.

Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'no child left behind'. *Race, Ethnicity, and Education*, 10(3), 245-260.

DiMaggio, D. (2010). The loneliness of the long-distance test scorer. *Monthly Review*, 62(7). Retrieved from http://monthlyreview.org/2010/12/01/the-loneliness-of-the-long-distance-test-scorer

Farley, T. (2009a). *Making the Grades: My Misadventures in the Standardized Testing Industry*. San Francisco: Berrett-Koehler Publishers.

Farley, T. (2009b). My Misadventures in the Standardized Testing Industry. Retrieved from http://voices.washingtonpost.com/answer-sheet/standardized-tests/-gerald-martineaupost-today-my.html

Fraser, S. (1995). *The bell curve wars race, intelligence, and the future of America*. New York: BasicBooks.

Giordano, G. (2005). *How testing came to dominate American schools: The history of educational assessment*. New York: Peter Lang.

Gould, S. J. (1996). *The mismeasure of man* (Rev. and expanded. ed.). New York: Norton.

Guthrie, R. V. (1998). *Even the rat was white: A historical view of psychology* (2nd ed.). Boston: Allyn and Bacon.

Haney, W. (1984). Testing reasoning and reasoning about testing. *Review of Educational Research*, 54(4), 597-654.

Herrnstein, R. J., & Murray, C. A. (1996). *The bell curve: Intelligence and class structure in American life* (1st Free Press pbk. ed.). New York: Simon & Schuster.

Jennings, J. F. (2000). Title I: Its legislative history and its promise. *Phi Delta Kappan*, 81(7), 516-522.

Jones, G. M., Jones, B. D., & Hargrove, T. Y. (2003). *The unintended consequences of high-stakes testing*. New York: Rowman & Littlefield Publishers, Inc.

Kane, T. J., & Staiger, D. O. (2002). Volatility in school test scores: Implications for test-based accountability systems. In D. Ravitch (Ed.), *Brookings Papers on Education Policy 2002* (1st ed., pp. 235-284). Washington, D.C.: The Brookings Institution.

Karier, C. J. (1972). Testing for order and controld in the corporate liberal state. *Educational Theory*, 22(Spring), 159-180.

Karp, S. (2006). Leaving public education behind: The Bush agenda in American education. *Our Schools/Our Selves*, 15(3), 181-196.

Karp, S. (2010). School reform we can't believe in. *Rethinking Schools*, 24(3). Retrieved from http://www.rethinkingschools.org/restrict.asp?path=archive/24_03/24_03_NCLBstan.shtml

Kliebard, H. M. (2004). *The struggle for the american curriculum*, 1893-1958 (3rd ed.). New York, NY: RoutledgeFalmer.

Kornhaber, M. L., & Orfield, G. (2001). High-stakes testing policies: Examining their assumptions and consequences. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers?: Inequality and high-stakes testing in public education* (pp. 1-18). New York: Century Foundation Press.

Kumashiro, K. (2008). *The seduction of common sense: How the right has framed the debate on America's schools*. New York: Teachers College Press.

Kumashiro, K. (2012). *Bad teacher!: How blaming teachers distorts the bigger picture*. New York: Teachers College Press.

Ladson-Billings, G. (2006). From the achievement gap to the education debt: understanding achievement in U.S. schools. *Educational Researcher*, 35(7), 3-12.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy*. New York: Farrar, Straus, and Giroux.

Linn, R. L. (2003, July). *Accountability, responsibility and reasonable expectations*. Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California, Los Angeles Retrieved 27th February 2006 from http://www.cse.ucla.edu/products/reports_set.htm

Madaus, G. F., & Clarke, M. (2001). The adverse impact of high-stakes testing on minority students: Evidence from one hundred years of test data. In G. Orfield & M. L. Kornhaber (Eds.), *Raising standards or raising barriers?: Inequality and high-stakes testing in public education* (pp. 85-106). New York: Century Foundation Press.

National Commission on Excellence in Education. (1983) *A nation at risk: The imperative for educational reform*. United States Department of Education, Washington D.C. Retrieved from http://datacenter.spps.org/uploads/sotw_a_nation_at_risk_1983.pdf

National Research Council. (2011). *Incentives and test-based accountability in education. Committee on incentives and test-based accountability in public education*, M. Hout & S.W. Elliott (Eds.). Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education, Washington, D.C.: The National Academies Press.

Oakes, J., Welner, K., Yonezawa, S., & Allen, R. L. (1998). Norms and politics of equity-minded change: Researching the "zone of mediation". In M. Fullan (Ed.), *International handbook of educational change* (pp. 953-975). Norwell, MA: Kluer Academic Publishers.

Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, Virginia: Association for Supervision and Curriculum Development (ASCD).

Rushton, P. J., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, 11(2), 234-294.

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

Schochet, P. Z., & Chiang, H. S. (2010). *Error rates in measuring teacher and school performance based on test score gains*. U.S. Department of Education, Institute of Educational Sciences, National Center for Educational Evaluation and Regional Assistance, Washington D.C. Retrieved from http://ies.ed.gov/ncee/pubs/20104004/pdf/20104004.pdf.

Selden, S. (1999). *Inheriting shame: The story of eugenics and racism in America*. New York: Teachers College Press.

Stoskopf, A. (1999). An untold story of resistance: African-American educators and I.Q. testing in the 1920's and '30's. *Rethinking Schools*, 14(1). Retrieved from http://www.rethinkingschools.org/archive/14_01/iq141.shtml

Tyack, D. (1974). *The one best system: A history of American urban education*. Cambridge, MA: Harvard University Press.

United States Congress. (2002). *No Child Left Behind Act of 2001*, 107-110 607. Retrieved from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Zinn, H. (1995). *A people's history of the United States: 1492-Present*. New York: HarperPerennial.

Wayne Au is an Associate Professor at the University of Washington, Bothell, and he is an editor for the social justice teaching magazine and non-profit publisher, *Rethinking Schools*. He writes extensively on critical education theory, particularly education policy and curriculum theory. Amongst other publications, he is author of *Unequal By Design: High-Stakes Testing and the Standardization of Inequality* and co-editor of, *Pencils Down: Rethinking High-Stakes Testing and Accountability in Public Schools.* wayneau@u.washington.edu