Can We Test for Liberation?
Moving from Retributive to Restorative and Transformative Assessment in Schools

Wayne Au
University of Washington, Bothell


Abstract
Schools in the United States are inundated with high-stakes, standardized tests, which are used as the central tool for educational policy and accountability systems there - often under the guise of promoting racial justice and civil rights. In this article the author uses empirical research on the impact of high-stakes to argue that, rather than promote educational equality, high-stakes testing in fact causes harm to working class and Black and Brown students as a form of retributive justice, which seeks to punish "wrongdoers" rather than addressing the actual material issues and conditions that contribute to educational achievement. Alternatively, in this article the author suggests that we can conceive of forms of restorative and transformative assessment that can be healing to our schools and communities as well as activist in nature.
“Education is the great Civil Rights issue of our time.”
Former President George W. Bush (CNN, 2002, n.p.)

The educational achievement gap, “is the civil rights issues of our time.”
Former U.S. Secretary of Education Rod Paige (Feinberg, 2004, n.p.)

“Education is the Civil Rights Movement of our generation.”
Former U.S. Secretary of Education Arne Duncan (2010, n.p.)

“Education is the Civil Rights issue of our time.”
Former President Barack Obama (Cooper, 2011, n.p.)

“Education is the Civil Rights issue of our day.”
Former U.S. Secretary of State Condoleeza Rice (CBS News, 2012, n.p.)

“[E]ducation is the civil rights issue of our time.”
President Donald Trump (Halper, 2017, n.p.)

Starting with the No Child Left Behind Act of 2002, the mantra has been this: There are race-based gaps in standardized test scores. Closing those should be the goal for achieving racial equality in education. Connected to clear punishments for failure, we can use those tests to achieve this equality (Au, 2009b; Karp, 2003, 2007). Since then politicians from both parties have said over and over again, educational achievement gaps are the Civil Rights movement of today (see, e.g., Paige & Witty, 2010). This is why a number of mainstream Civil Rights organizations spoke out against the movement to resist high-stakes testing in May of 2015. Their argument: without the data, we can’t hold schools and teachers accountable for racial inequality (Brown, 2015).

The logic behind using high-stakes tests to fight racial inequality in schools is a simple one. The standardized tests produce data that we can look at and identify “achievement gaps” amongst different groups of students. Then, if we don’t see the scores of low performing groups go up and don’t see these gaps closing, teachers, schools, students, and, ultimately their communities, will be held “accountable” to punishment for not improving. Such punishments have included cutting education funds, closing schools, firing an entire school staff and reconstituting it with new hires, converting a public school to a charter school, withholding diplomas from students, holding kids back in grade levels, firing teachers, and chastising in public by the media. The idea being that these kinds of threats will lead to the increased achievement of low income and kids of color (Fabricant & Fine, 2013; Picower & Mayorga, 2015).

Within this logic we have to recognize that there is an important truth underlying the impulse to use high-stakes test scores as a racial accountability measure in education. In the United States we have such a long history of racism, white supremacy, and the reproduction of poverty in our public school system (Au, Brown, & Calderon, 2016) that when communities of color express that they want to make sure that their children are being educated and not harmed,
Can we test for liberation?

it has to be taken seriously and understood as based in the reality of community experience across generations. However, despite the racial, cultural, emotional, and historical truth of this impulse for accountability, there are some faulty assumptions underlying the entire argument for using high-stakes tests to achieve the ends of race and class equality in education.

The Faulty (Bio)logics of Standardized Testing

The arguments for using high-stakes testing for racial equality all assume that our standardized tests provide accurate measurements of teaching and learning so that an increase in test scores equal educational improvement. This presumption has does not hold true since test scores correlate most strongly with family income, neighborhood, educational levels of parents, and access to resources – all factors that are measures of poverty that exist outside of schools (Amrein-Beardsley, 2014; Berliner, 2010, 2013). This is not to say that schools and teachers are not important in student learning and achievement. It is to say that, while schools are teachers are central to how our children learn and experience education, the tests offer such narrow measures that they miss most of the social relations that constitute teaching and learning. It is also to say that, as we have seen in so many “miracle” schools, a rise in test scores has been about gaming the system, being selective about student enrollments, or losing low achievers through attrition, as opposed to being connected to real gains in student learning (Nichols & Berliner, 2007).

Another assumption underlying current policy is that high-stakes tests are race-neutral tools that are in fact capable of promoting racial equality. This assumption is also faulty. At their origins one hundred years ago, standardized tests were used as weapons against communities of color, immigrants, and the poor. Early concepts of aptitude and I.Q. were foundational for these tests such that Whites, the rich, and U.S. born were “scientifically” found to be biologically more intelligent than other groups, because of the presumed objectivity of standardized tests – thereby justifying the race, economic class, and cultural inequalities that existed then (Au, 2009b). For instance, in 1916, based on standardized test scores, Stanford Professor Lewis Terman—one of the founding fathers of standardized testing in the United States, argued that certain races inherited “deficient’ IQs and that, “No amount of school instruction will ever make them intelligent voters or capable citizens.” He further asserted that “feeblemindedness” was, “very, very common among Spanish-Indian and Mexican families of the Southwest and also among negroes [sic].” Terman then suggested that, “Children of this group should be segregated in special classes and be given instruction that is practical…” because, “They cannot master abstractions, but they can often be made efficient workers” (as quoted in, Blanton, 2003, pp. 43–44). Thus, the test scores were used to justify educational tracking for kids of color, immigrants, and the poor, as this kind of testing was brought en masse into the growing public school system (Au, 2009b).

This kind of test-informed inequality stems in part from the design paradigm of our standardized tests themselves, which were (and are) based on a foundational assumption that human intelligence is based on biological aptitude that is “naturally” distributed on a bell curve across populations (Bisseret, 1979; Gould, 1996). This bell curve logic decrees that, if a test was given to a large population, some would score high, a lot would be in the middle, and some would do very poorly. Within this paradigm, when the distribution of test success and failure happens to match existing racial and economic inequality, it ends up being viewed as coincidental and part of the “natural” order of things (Weber, 2015). Additionally, even when we used standards-based tests where, theoretically, everyone could “meet standard,” we are still
submitted to the logic of unequal distribution of a bell curve: In order to translate the standards-based test scores to make student comparisons, statisticians put raw test data through a complicated statistical process of scaling, equating, and normalizing that still ends up producing a bell curve of student scores (Tan & Michel, 2011).

The bell curve underlying the construction of high-stakes, standardized means that they will always produce high-scorers and low-scorers, will always produce winners and losers. Put differently, the assumed unequal distribution of intelligence built into the tests makes it impossible to reach an equality of test scores, which also means that they make it impossible to reach equality among test-takers. This reality paints all discussion of closing the achievement gap in test scores in a new light as well, because, given how the tests are designed to produce inequality, then “closing the achievement gap” does not mean everyone succeeds on the tests. Rather, “closing the achievement gap” really only means proportionate success and failure between groups such that we have equal numbers of rich and poor students passing and failing, equal numbers of Black, White, Asian, Native, and Latinx passing and failing, etc.

It is important to recognize that this notion of the unequal distribution of biological intelligence along a bell curve is not some outlandish idea relegated to the distant past. For instance, in the mid-1990s, the authors of The Bell Curve (Herrnstein & Murray, 1996) claimed that there was an 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. Like Yerkes before them, Herrnstein and Murray based their conclusions on an analysis of standardized test scores. More recently, Rushton and Jensen’s (2005) analysis of “Thirty years of research on race and cognitive ability,” asserted that there are genetically based racial differences in I.Q. (Jensen was 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 Herrnstein and Murray (1995) as a baseline for analysis. Further, we should not lose track the fact that overall our race-based standardized test score gaps today parallel the same general outcomes of the blatantly racist standardized tests of 100 years ago, nor should we forget that our current test scores and the gaps they produce continue to be used today to justify systems of tracking and unequal access to resources (Au, 2009b).

A third faulty assumption underlying the idea of using high-stakes tests to achieve racial equality in education is the theory that threat of discipline punishment will bring about individual and systemic changes that will raise test scores and increase achievement. Again, this hasn’t proven to be correct. While it is true that, for instance, test scores in Math and English/Language Arts for Black and Latino students have generally risen since the implementation of nationally required testing through the No Child Left Behind Act (NCLB), the test scores of white students have risen even more (National Research Council, 2011; Ravitch, 2014). So, after over a decade of the very intense test-and-punish policies of NCLB—the bulk of which still exist in ESSA (Karp, 2016)—the threats made against, and the disciplining of, teachers, students, and schools have exasperated racial test score gaps instead of closing them.
High-Stakes Testing and Discipline

It is critical to understand that the idea of discipline is at the heart of policies and practices aimed at holding students, teachers, and schools “accountable” for raising test scores. There is physical discipline: Students are bound by a set amount of testing time. They must remain silent. Their movements are restricted and generally must stay seated. They either cannot access resources beyond the assessment or are limited to only using specified resources (e.g., calculators). Testing happens under the gaze of an authority (a test proctor) watching for transgressions against these disciplinary restrictions. There is also a physical discipline of enduring the stress and physical limits placed on them for hours at a time, as well as the emotional discipline of handling the stress of the high-stakes of the tests. High-stakes, standardized tests also discipline knowledge: They determine what knowledge and content is considered legitimate for teaching and learning in the classroom. They also discipline teachers’ pedagogy in how they teach content because they compel teacher to teach to the test and place restrictions on depth and breadth of subject matter (Au, 2007, 2011; Madaus, Russell, & Higgins, 2009). Further, the tests discipline educational resources, as money is focused on test-aligned textbooks, teaching materials, and professional development. In Discipline & Punish, Michel Foucault (1995) talks about how the process of examination enforces discipline because it:

...assures the hold of power that is exercised over them. It is the fact of being constantly seen, of being able always to be seen, that maintains the disciplined individual in his subjection. And the examination is the technique by which power, instead of emitting the signs of its potency, instead of imposing its mark on its subjects, holds them in a mechanism of objectification. In this space of domination, disciplinary power manifests its potency, essentially, by arranging objects. (p. 187)

High-stakes tests objectify teaching and learning, objectify teachers and students. They seek to arrange through comparison, sorting, and ranking. This is their power to discipline us within systems of accountability. This is their power in the establishment of a system of surveillance and an environment of teaching and learning under constant threat (Au, 2009b; Vinson & Ross, 2003). The question remains, can this kind of test-based surveillance and discipline, be used to promote educational equality and racial justice?

Disciplining Black and Brown Bodies

While high-stakes, standardized testing disciplines all students, because the tests concentrate failure in their schools and communities, non-white and low income students are disciplined with disproportionate intensity. This means that low income and kids of color are: tested more; experience the greatest loss of time spent on non-tested or less-tested subjects like art, music, science, and social studies; don’t have multicultural, anti-racist curriculum made available to them because those areas are on the tests; and lose opportunities for culturally relevant instruction because the test tend to inhibit processed based, student-centered instruction in favor of rote memorization (Au, 2009a; Nichols & Berliner, 2007). Further, because of the increased intensity and more restrictive curriculum and educational environment, for children of color high-stakes testing serves to acculturate them to a norm of being disciplined by state authorities (Au, 2009b; Vinson & Ross, 2003). The disciplining of Black and Brown children by
high-stakes, standardized testing also manifests in very concrete and material ways: A study by the Economic Policy Institute found that in one state having a high-stakes exit exam correlated with a 12.5% increase in the rate of incarceration (Baker & Lang, 2013). Again, given that these tests fail kids of color disproportionately, this study suggests that high-stakes, standardized tests are a conduit for the school-to-prison pipeline.

**High-Stakes Testing as Retributive Justice**

Ultimately, high-stakes, standardized testing, with its focus on surveillance, discipline, and punishment, represents a form of *retributive justice*. Built around the concept of “retribution” for a crime, retributive justice seeks vengeance for a wrongful act, and it is the cornerstone of our entire system of criminal justice in the United States (Zehr, 2011). Zehr (2011) explains that retributive justice operates along three questions:

- What rule has been broken?
- Who is to blame?
- What punishment to they deserve? (n.p.).

Educational “accountability” based on high-stakes testing are an expression of retributive justice. They criminalize test failure and, instead of trying to fix the problem, they seek to punish students, teachers, schools, and communities for that failure.

Can we test for racial and educational justice? Within the current system of high-stakes, standardized testing, we cannot. The paradigm of test and punish does not promote justice in any form. Indeed, looking at its impact, high-stakes test-based policies promote injustice and inflict acute damage on kids of color especially. However, there are other models of justice, such as restorative and transformative justice (Zehr, 2011), to explore as well. This raises the questions: If high-stakes testing is retributive, can we conceive of assessments based on these other forms of justice? Can we test for liberation?

**Dreaming of Restorative and Transformative Assessment**

Restorative justice has become an increasingly popular alternative to models of disciplinary, retributive justice all too common in schools (see, e.g., Annamma, 2015). According to Zehr (2011), restorative justice asks a different set of questions than the retributive model:

- Who has been hurt and what are their needs?
- Who is obligated to address these needs?
- Who has a “stake” in this situation and what is the process of involving them in making things right and preventing future occurrences? (n.p.)

Following the spirit of this model, and centering racial justice for the sake of this discussion, a restorative model of assessment would begin by recognizing that students of color have been hurt by institutionalized racism and white supremacy in our schools (Au et al., 2016; Dumas, 2014; Patel, 2016) and that our current assessments have perpetuated this hurt (Au, 2015). Further, our assessments would then need to explicitly name who is obligated to address the institutionalized racism and white supremacy faced by our students in their schools. We would need to think...
through who has voice and power in determining the assessment and as well as their role and responsibility in making things right. Restorative assessment would also take seriously the idea of healing our kids and communities. Imagine the possibility of an assessment that would be a part of a process of healing the hurt caused by white supremacy and institutionalized racism. Imagine an assessment that was culturally responsive in form and content, one that assessed students for identity development, knowledge of self, cultural knowledge, and confronting internalized oppression/colonization.

Aspects of restorative assessment exist in some places. For instance, much of Christensen’s (2009, 2017) work, which asks students to write powerfully through pain, or to consider race, class, and the power of language in their lives, begins to point the way towards forms of restorative assessment in the classroom. The practices and pedagogy of the now-banned Mexican American Studies program in Tucson, Arizona also orient us towards restorative assessment, as they engaged their students in deep learning that sought to decolonize curriculum and their cultural selves (Acosta & Mir, 2012; Romero, Arce, & Cammarota, 2009; Tintiangco-Cubales et al., 2015). The revitalized ethnic studies movement statewide in California, as well as in cities like Portland, Oregon, and Seattle, Washington (Ware, 2017) also create potential curricular space for restorative assessment, because the fight for ethnic studies signals to the world that there is knowledge that is important for students to learn about themselves, each other, and their communities. We can also see it in the anti-racist writing assessment being done by Inoue (2015), which challenges the norms for grading writing as colonizing and racist.

Restorative assessment also invites a conversation about the very forms our assessments take. For instance, as opposed to numerical scores generated by the typical high-stakes, standardized test, processed-based portfolios and performance assessments allow for more nuanced, human, and complex expression of student learning, growth, and development (Foote, 2007; Meier & Knoester, 2017). Similarly, restorative assessment puts the very idea of A-F and decimal grading systems into question since those singular grades and decimal scores generally perform the same functions of surveillance and punishment as standardized tests (Kohn, 2013). The types of critical self-reflection and self-evaluation that can be built into strong portfolios and performance assessments, as well as more dialogic narrative evaluations by teachers (Meier & Knoester, 2017), are more restorative since they ask us to assess through conversation, relationships, and centering the power of assessment more within students themselves.

However, restorative justice has been critiqued on multiple grounds, including its focus on individual events and its lack of recognition of the conditions that contributed to and shape wrongdoing (Nakagawa, 2003). In response to these critiques to restorative justice, activists have also been advocating for models of transformative justice. Transformative justice asks a set of distinctly different questions than both retributive and restorative models of justice:

- What social circumstances produced the harmful behavior?
- What structures exist between this structure and others like it?
- What measures could prevent further occurrences? (Zehr, 2011, n.p.)

If we applied these to thinking through what transformative assessment might look like, then it suggests a series of skills and a knowledge base that we want to make sure our students are
learning. For instance, we might assess students on their understanding of what historical and socio-economic circumstances produced the institutionalized racism and white supremacy we see in our schools (Picower & Mayorga, 2015), or the role that schools play in maintaining the settler-colonial state (Tuck & Gaztambide-Fernandez, 2013). Given the continued focus on high-stakes testing in education policy and practice, transformative assessment could also check for student understanding of how testing itself reproduces inequalities and maintains white supremacy (Au, 2015), as well as what can be done to mitigate the transfer of structural inequalities in both our schools and our assessments.

As is implied by the name, transformative assessment could also assess students on their understanding of, and capacities for, institutional and community transformation, including their knowledge of strategies for challenging institutional racism in their schools, districts, and communities. From an activist and transformational perspective, there are concrete skills, understandings, and forms of resistance that we want to foster and develop in our students as potentially powerful individuals and collectives. How do we assess those in ways that are not retributive and ways that are not remotely connected to our current norms and assumptions about standardized testing and accountability? For instance, in May of 2016, students at Forest Grove High School outside of Portland, Oregon protested a Trump-inspired “build a wall” banner that was hung at their school. Hundreds of Forest Grove High School Students walked out and rallied against the act of racism. Soon, as word spread through social media, they were joined by students from six other high schools in their area (KGW Staff, 2016). In the following days students from Portland high schools and Portland Community College joined for a rally in the city of Portland itself (Gallivan, 2016). It was a powerful moment for student organizing in the Portland area that helped sharpen peoples’ consciousness about anti-immigrant racism in the region.

The example of Forest Grove High School students, as just one possibility among many in recent years (e.g., McKay, Regunberg, & Shea, 2014), shows that there are basic skill sets for organizing that we could see in student learning through transformative assessment. The Forest Grove example also shows that there are forms of consciousness and political orientations that are anti-racist and willing to oppose structural authority that help students understand and act on their collective power in transformative ways. Indeed, activism like that of the Forest Grove students also points to the important relationship between restorative and transformative assessment: Restorative assessment, with its focus on healing, cultural self-knowledge, and decolonization can help foster the kinds of critical consciousness that can contribute directly to student involvement in mass mobilizations and movements for social justice we might connect with transformative assessment. Further, student organizing in Forest Grove and elsewhere suggests the possibility that we can create and wield restorative and transformative assessments to foster this kind of activism as a central aspect of public education. Such assessments would have to challenge racism and white supremacy (aligned with other aspects of our intersectional identities too) with a focus on cultural and community healing and radical institutional transformation. Such assessments would also have to be constructed to openly challenge hegemonic power and require a complete break from our current system and logics of high-stakes, standardized testing.


**Can we test for liberation?**

Reclaiming Assessment

The possibility of restorative and transformative assessment is also a call for educators and students to reclaim the very idea of “assessment” from the high-stakes standardized tests and corporate education reformers using them for profits and privatization. As educators we are, after all, interested in student learning, and using assessment to keep track of that learning is a part of the process of teaching and learning. The issue is what kind of assessment – in terms of form and process, as well as to whom and what ends our assessments serve. Right now there are real fights in the streets about police killings of Black men and women, about anti-immigrant xenophobia, nationalism, and White supremacy, about native sovereignty and the protection of the environment against the ravages of a fossil fuel economy, about gentrification, livable wages, access to healthcare, epidemic homelessness, and the lack affordable housing. Students are struggling with these issues in their lives and are out protesting in the streets. Our students need restorative understandings to help them make sense of their lives and times, and they need to know about transformative strategies and skill sets that they can use in these fights. The possibility of restorative and transformative assessment is a call for educators to create forms of schooling and assessment that are liberatory, partisan, and in the interest of justice.

**References**


Christensen, L. (2017). *Reading, writing, and rising up: Teaching about social justice and the power of the written word* (2nd ed.). Milwaukee, WI: Rethinking Schools, Ltd.


Can we test for liberation?


**Author**

Wayne Au is a Professor in the School of Educational Studies at University of Washington, Bothell, and he is an editor for the social justice teaching magazine, *Rethinking Schools*. 