Wayne Au and Karen Gourd

Asinine Assessment: Why High-Stakes Testing Is Bad for Everyone, Including English Teachers

High stakes, standardized exams have been billed as a panacea for our educational ills. . . . This is a sham and an appallingly bad educational strategy which guarantees poor results, reduced motivation and legions of graduates without the skills necessary to live a decent and fulfilling life.

—Peter Henry, “The Case Against Standardized Testing”

High-stakes standardized testing has become ubiquitous in US education. We argue that not only is such testing bad for education, but it also contradicts curriculum and instruction aligned with professional standards promoted by NCTE and innovative educators. Additionally, contradictory to the concept of “no child left behind,” high-stakes standardized tests have negative impacts on students of color. However, we also know that English teachers can make a positive difference in the education of youth through creative curricular resistance and by keeping the big picture of the discipline as the heart and soul of curriculum, instruction, and assessment.

The Bipartisan Politics of High-Stakes Testing

The modern-day, high-stakes standardized testing movement can effectively be traced back to the 1983 publication of the National Commission on Excellence in Education, A Nation at Risk: The Imperative for Educational Reform. This Reagan-era report sounded an alarm within public education in the United States, and, even though much of the report’s education crisis was found to be manufactured, it had an enduring impact on public education: Fifty-four state-level commissions on education were created within one year of the report’s publication. Within three years, 26 states raised graduation requirements, and 35 states instituted reforms that revolved around testing and increased course loads for students. By 1994, 43 states had implemented statewide assessments for K–5, and by 2000 every US state but Iowa had administered a state-mandated test. Within the first week of taking office in 2001, with the overwhelming support of Democrats and Republicans alike, President G. W. Bush pushed for federal Title I funding to be tied to student test scores, resulting in the passing of the No Child Left Behind Act (NCLB) of 2002 (Au).

As policy for more than a decade, the required testing under NCLB is well known to educators: high-stakes testing is mandated in reading and math in grades 3–8 and once in high school. In science, students are to be tested at least once at the elementary, middle school, and high school levels. If subgroups of students related to race, economic class, special education, or English language proficiency do not show consistent improvement in test scores, schools face sanctions such as a loss of federal funding. Wrapped in language about ending educational inequality and closing achievement gaps, the policy goal of NCLB was that by 2014 all students in all subgroups would reach 100 percent...
Despite growing public criticism of NCLB and official rhetoric about the need for multiple measures, the election of President Barack Obama resulted in policies that intensified inappropriate use of high-stakes standardized tests to rate schools and assess teachers.

The Ignored History of Standardized Testing

Given the ubiquity of high-stakes standardized testing, it is surprising that many people do not know that such tests have their origins in IQ testing and the eugenics movement in the United States. In 1904, French psychologist Alfred Binet developed the IQ test to determine if some young children were mildly developmentally disabled. By dividing the first (mental age) by the second (chronological age), the idea of “intelligence quotient,” or IQ, was born. According to the late evolutionary biologist Stephen Jay Gould, Binet’s testing was to be used with young children only, and it was conceived purely as a practical tool for identifying children who might benefit from specific early interventions. It was not connected to any idea of hereditary or innate intelligence. However, cognitive psychologists in the United States such as Goddard, Terman, and Yerkes distorted Binet’s testing and measurement of IQ by advancing the idea that intelligence was hereditary and fixed, thus laying the groundwork for testing to be used to sort according to supposed innate intelligence.

According to Gould, Yerkes played a particularly important role in this process. 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 from these 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 African Americans were the least intelligent of all peoples.

Standardized IQ testing soon found its way into the institution of education. David B. Tyack provides an explanation: “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 invaluable means of channeling children; by the very act of channeling pupils, they helped to make IQ prophecies self-fulfilling” (180). 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 or deny admissions (Au).

Despite the rhetoric of NCLB to reduce achievement gaps between the rich and poor and between racial and ethnic groups, we have yet to reconcile the historical racism, nativism, and eugenics that lay at the roots of high-stakes standardized testing. We have yet to reconcile the historical racism, nativism, and eugenics that lay at the roots of high-stakes standardized testing.

We have yet to reconcile the historical racism, nativism, and eugenics that lay at the roots of high-stakes standardized testing.
Asinine Assessment: Why High-Stakes Testing Is Bad for Everyone, Including English Teachers

Entwined Assumptions of Standardized Testing

High-stakes standardized tests are based on assumptions that raise serious questions about their efficacy in ending educational inequality. Perhaps most glaring of these problematic assumptions is that, to be viewed as valid and reliable measures of human populations, the statistical logic of standardized tests require that some students fail. Put differently, if everyone passed a standardized test (or in the language of NCLB, if all students achieved 100 percent 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 both.

Much like our current system of economics, the system of high-stakes standardized testing cannot function if everyone is a “winner”; this point is particularly important when it comes to the discourse of race and class issues surrounding current education reform and the hyper-reliance on high-stakes testing. Such testing has been promoted under the guise of achieving racial and economic equality through a focus on closing the test-score achievement gap. One of the great ironies about this discourse is that “closing the achievement gap” does not mean having everyone succeed on the tests. “Closing the achievement gap” really just means that equal numbers of rich kids and poor kids pass and fail, that equal numbers of white kids and African American kids pass and fail, that equal numbers of native English speakers and English language learners pass and fail, etc.

Another problematic assumption underlying standardized tests is that the tests provide objective and accurate measurement of student learning and, by extension, effective teaching. Standardized test scores are far from objective or accurate. For instance, Peter Z. Schochet and Hanley S. Chiang report that there is a statistical error rate of 35 percent when using one year’s worth of test data to measure a teacher’s effectiveness, and an error rate of 25 percent when using data from three years. As another instance, Thomas J. Kane and Douglas Staiger, in their research presented to the think tank the Brookings Institute, found that one time, randomly occurring factors (e.g., 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, which other students happened to be in attendance while taking the test, whomever happened to be administering the test) account for 50 to 80 percent of any gains or losses on a given student’s standardized test score.

Inaccuracy in Scoring Standardized Writing Tests

There are other issues that raise serious questions about the objectivity and accuracy of high-stakes standardized testing as well, such as the inability of the tests to account for cognitive transfer of skills and knowledge across subjects or teachers and the high correlation of test scores with non-school factors (e.g., inadequate access to health care, food insecurity, and poverty-related stress). However, perhaps the biggest issue is the lack of objectivity and accuracy in the scoring of the tests themselves. Todd Farley, a former employee of 15 years in the testing industry, described his work experience in a 2009 Washington Post blog:

[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.

The rhetoric around standardized tests is that they are less biased and more accurate indicators of what students know than assessments completed by classroom teachers who know and interact with the students over time. Yet the reality is that the unbiased statistical value of standardized tests is suspect, and increased accuracy based on unbiased scoring using a rubric and a scorer who does not know the students is a myth.

The lack of objectivity and inaccuracy of high-stakes test scores is perhaps sharpest in the grading of standardized writing tests themselves.
Dan DiMaggio worked as a writing test scorer for several years and talks about what it is like to be a paid temp worker for Pearson Education:

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.

Perhaps even worse, DiMaggio explained how he and other scorers were told to change their scores 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.

With so much inconsistency and subjectivity going into the scoring, it is unconscionable that we are using such tests to make high-stakes decisions for students, teachers, principals, and schools.

Testing and English Language Arts Instruction

On the whole we know that high-stakes testing is controlling both what and how subjects are taught: untested subjects are being reduced in the curriculum and teachers nationwide are also moving toward more teacher-centered, lecture-based pedagogies that encourage rote learning in response to the pressures of the tests. Further, and perhaps ironically, given the stated policy focus on addressing inequality, test-based changes in classroom instruction negatively affect non-white students disproportionately and the pressures of high-stakes standardized testing are greatest in states with high populations of students from low-income families and students of color (Au).

English language arts has certainly not been immune to these national trends, and the instruction encouraged by high-stakes tests contradict decades of research on teaching of writing and literature. For example, in Massachusetts, where students participate in the Massachusetts Comprehensive Assessment System (MCAS), teachers report that how they teach writing, as well as the types of writing students are asked to perform, are being controlled by high-stakes tests. In a study by Catherine Luna and Cara Livingstone Turner, one teacher from an urban Massachusetts high school explained, “I find in my class that I’m teaching to the test right now. I’m drilling on five paragraph essays,” while another colleague says, “You know, we’re not really teaching them to write. We’re teaching them to follow a format” (83).

In studying North Carolina, Maika Watanabe found that under the pressures of the high-stakes writing assessment there, writing has “become less ‘like a real writer writes’ in that the focus of writing has shifted to form over content and product over process. In addition, teachers emphasize that they teach students to submerge their voices as they write to inauthentic audiences. They further provide fewer opportunities to integrate reading and writing assignments and to write a variety of genres with the demands of testing looming over them” (343). Likewise, writing instruction for the teachers in Watanabe’s study has succumbed to time, genre, length, and process constraints required by the tests, resulting in less student voice in writing, less integration of writing with other content, less time
for students to explore diverse genres, and more formulaic writing generally.

Research has also found that multicultural content is being pushed out of the language arts curriculum because the tests do not assess such content. For example, Jane Agee studied the experience of an African American teacher working in a high-stakes testing environment. This English teacher graduated from a progressive teacher education program with the goal of bringing multicultural education into her practice, including an explicit commitment to using more multicultural literature. However, this teacher reluctantly gave up on including multicultural voices in her English curriculum because these voices were not included on the tests.

Challenging Standardized Tests with Creative Curricular Resistance

High-stakes tests are asinine. They are rooted in racism and classism, and as a 2011 National Research Council report tells us, a focus on high-stakes testing for nearly a decade has not closed achievement gaps, and in cases like the use of high school exit exams required for graduation, they have made it worse. The current focus on testing continues to build upon this legacy by limiting curriculum and opportunities for academic achievement for all students, while simultaneously limiting effective and engaging practices. However, the English classroom can be a place for teachers and students to engage in what Herbert R. Kohl calls “creative maladjustment”: “Creative maladjustment consists of breaking social patterns that are morally reprehensible, taking conscious control of one’s place in the environment, and readjusting the world one lives in based on personal integrity and honesty—that is, it consists of learning to survive with minimal moral and personal compromise in a thoroughly compromised world and of not being afraid of planned and willed conflict, if necessary” (130). The English classroom is the perfect site for creative maladjustment, even if working within a framework intended to create curriculum conformity.

First, English teachers can have students research, read, and write about the issues around standardized assessment. Bill Bigelow does this in his lesson, “Testing, Tracking, and Toeing the Line,” which looks at how the histories of both testing and tracking are tied to eugenics and socioeconomic stratification. Similarly, in her article “High-Stakes Harm,” Linda Christensen discusses the racist history of the SAT with her students and asks them to critically address the cultural biases in the test by asking them to create their own, homegrown version. What could be more relevant to current US students than how high-stakes testing affects them? Having students write about things they care about can motivate them to develop necessary communication skills and to recognize learning as a useful process that opens doors to creativity and independence. Indeed, having students informed about current educational issues can increase the pool of informed voters post-graduation.

Second, English teachers can commit to resisting external pressure to become teachers of rote curriculum and low-level cognitive skills. Young people are capable of more than low-level skills and are more likely to acquire high-level skills through creative, challenging work than by un-engaging drill-and-skill approaches. Students who are skilled at high-level cognitive processes are likely to also be able to complete low-level cognitive tasks while the reverse is less likely. Students who are skilled at low-level tasks may not be able to apply these skills in creative and innovative ways, especially if the more complex skills are only rarely used. The more complex things are, the more exposure and practice we need to become proficient. The easy stuff can be learned in the context of complex skill development. For example, writing essays about relevant topics, such as high-stakes testing, for a real audience provides opportunities for exercising complex cognitive skills and practicing grammatical rules.

Third, English teachers need to find ways to make multicultural literature a permanent component of the curriculum. Multicultural literature offers a richness of ideas and experiences to teachers and students. Further, attending to multicultural literature helps keep teachers in touch with a key principle: Our curriculum is more engaging if it is grounded in the lives of our students. A pragmatic strategy to keep multicultural literature an integrated part of the curriculum is to start by making the canon relevant to students. When students learn to love Shakespeare and Steinbeck, parents, peers, and administrators are unlikely to find fault with the teacher’s curriculum. Then compare and contrast perspectives from canonical and multicultural
literature, recognizing differences and similarities. Students can be asked to challenge or defend universality as a condition of the human experience. In addition, considering multiple perspectives supports the development of divergent thinking, a cognitively demanding process useful in science, technology, math, sociology, and the humanities.

Creative maladjustment may appear radical to some teachers; however, students would benefit if teachers recognized creative maladjustment as a professional response to unjust educational practices. We have the choice to explore enduring themes relevant to individuals and humankind. We can continue to determine the English language arts curriculum and to use instructional strategies that inspire enthusiasm, creativity, and higher order thinking for teachers and students.

Works Cited


Wayne Au, a former high school history and language arts teacher, is an associate professor in the education program at the University of Washington, Bothell, and is an editor for the progressive education magazine Rethinking Schools. Karen Gourd, a former high school teacher, is an assistant professor in the education program at the University of Washington, Bothell. Contact the authors at wayneau@u.washington.edu and KGourd@uwb.edu.

READWRITETHINK CONNECTION
Lisa Storm Fink, RWT

The article encourages students to research, read, and write about the issues around standardized testing. The ReadWriteThink.org lesson plan “Chasing the Dream: Researching the Meaning of the American Dream” invites students to explore the meaning of the American Dream by conducting interviews, sharing and assessing data, and writing papers based on their research to draw their own conclusions. http://www.readwritethink.org/classroom-resources/lesson-plans/chasing-dream-researching-meaning-30925.html

English Journal 19