Standardized students: The problems with writing for tests instead of people

Bronwyn T. Williams

Williams teaches at the University of Louisville in Kentucky. He is editor of the Literacy and Identity Department.

I'm not usually a huge fan of bumper stickers, even the ones I agree with, because of the way they shout out simplistic positions on complex issues. I still remember, though, when I saw the one that read "Standardized Testing Produces Standardized Students." I smiled and nodded my head a bit in agreement. Simplistic as that phrase may be, over the years it comes back to me when I am involved in conversations about testing that seem bound in reductive and simplistic arguments about standards, rigor, and accountability. I don't use the phrase because I don't think using a bumper-sticker argument I agree with makes it superior to my opponents' bumper-sticker arguments. But I do try my best to nudge arguments about testing toward a more complex consideration of the myriad implications of the concept of standardization.

Some levels of standardization I rely on. I take great comfort that there are identifiable standards for inspecting elevators, for example. I also appreciate that most medical doctors go through some kind of standard training in human anatomy. And I believe that we teachers have a responsibility to assess whether students are learning the concepts, ideas, and ways of thinking that we believe are important and also to hold ourselves and our pedagogies accountable if we are failing to reach most of our students.

Like many of my colleagues, however, I am not convinced that literacy assessment is best achieved through standardized tests given to huge groups of students in high-stakes situations. More and more, it seems as if the point of literacy education—of all education—is becoming standardized assessment and rankings rather than learning. Standardized assessment differs from assessment that attempts to determine whether students are learning what we are trying to teach. No one I know is against the latter. What recent trends in standardized assessment emphasize, however, is not learning but the comparison and ranking of failure. Incessant testing regimes, such as the infamous No Child Left Behind (2002) law in the United States (known ruefully among many teachers as No Teacher Left Standing), focus on broad comparisons of students, with little regard to their differences, and severe punishments for schools and teachers who fail to meet the "standards." As long as students meet the standards, what they may actually be learning seems to be beside the point.
The fervor for this kind of standardized testing reinforces the kind of ranking games that are a particular enthusiasm of Americans and are certainly not unknown in other countries. We’re willing and eager to rank anything—from the 100 greatest movies to the 250 best cities in which to raise children. Never mind that the criteria for such rankings are hazy at best; if we can’t put a number on it and rank it, then what good is it? Numbers seem scientific and technological. So we test and test and test, oblivious or resistant to the possibility that standardized literacy testing often produces numbers with about as much utility or connection to reality as ranking songs on the old American Bandstand television program. “It sounds good, has a good beat, and I can dance to it. I give it an 87.”

**Testing as punishment**

My concerns about the increasingly pervasive reliance on standardized testing in literacy education are about more than questionable methods of assessment and measurement, however. I am also deeply troubled about the implications for issues of literacy and identity. What effect does the unrelenting emphasis on standardized literacy testing have on students’ perception of the purposes and possibilities of literacy? By extension, what effect does such testing have on their perception of the possibilities for themselves as readers and writers?

Many concerns about identity and standardized testing have been framed in terms of race and social class and have been well documented and well argued by others (McNeil, 2000; Murphy, 1997; Ohanian, 1999; Orfield & Kornhaber, 2001). These teachers and researchers have argued that standardized testing works not from a set of objective standards somehow as constant as the North Star but from a set of cultural conceptions about literacy that are neither objective nor static. Students whose race or social class is not part of the dominant culture often face more complex challenges in meeting the standards of that dominant culture. Much of the impulse behind standardized tests and their illusion of objectivity seems to be a drive to punish, ridicule, and marginalize those who already feel punished, ridiculed, and marginalized by the institutions of education. At the same time the standardized tests from the dominant culture reassure its members about the quality of their educational institutions as well as their children.

Using literacy tests to reinforce dominant privileges and exclude others is nothing new; all we have to do is remember the literacy tests used to reject immigrants to the United States or to keep African Americans off U.S. voting rolls in the last century. The standardized testing movement today is just better able to cloak such motives within the rationale of “not leaving children behind.”

I do have another concern about these methods of literacy assessment in terms of student identity. Most teachers have stories of bright students who “don’t test well.” I’ve seen such students at every level, from middle school to university, and all of them could do innovative, creative, fascinating work on a project or a paper. But for reasons stretching from learning disabilities to personality traits and cognitive ways of processing and communicating information, they could not score well on timed, standardized tests. Nevertheless, there are students who blossom in such test-taking situations—students who understand the rhetorical demands and structures of standardized exams, and whose minds organize and recall certain kinds of information quickly and efficiently.

I see one of each kind of student when I look at my twin adolescent sons, who were born just 15 minutes apart and raised in the same circumstances. One son excels at taking standardized tests of all kinds by understanding the rhetorical structure of the questions and the institutional demands of the exam. The other, though in some ways a more powerful writer and just as strong a student in school, has always found standardized tests rigid and bewildering. If all I knew about them were the results of their annual...
standardized tests, I would no doubt rejoice in how the school system was succeeding to educate one while worrying over its failure to reach the other. Their wildly divergent test scores tell me nothing about their abilities and nothing about the quality of teaching they receive.

Yet U.S. culture clings to standardized literacy tests as a means of providing meaningful information about students, teachers, and schools because such tests offer the illusion of scientific rigor (as well as those all-important quantifiable numbers) to an endeavor that ultimately can't be measured in a lab and for which numbers are meaningless. This infuriating numbers game allows politicians and media pundits to make facile judgments, and cynical proclamations, about education that they turn into a relentless cycle of testing, criticism, and punishment. From the administrators to the teachers and students, testing drives the curriculum, and the curriculum shapes student identities in terms of literacy practices.

**Writing as a human endeavor**

Standardized testing, to be standardized, must create questions and answers that leave no room for interpretation. Such rigid questions and answers remove the importance of context from literacy practices and allow for no independent meaning making from students. Yet it is in that moment when an individual makes meaning in writing and reading in a specific cultural context that identity and literacy come together. When literacy education becomes more about standardized assessment, it becomes less about writing and reading by individuals who make meaning and have something to say. In the drive to assess and quantify, what is forgotten is why we want students to read and write in the first place. Reading and writing is about communicating with other human beings—about being part of a society and its ongoing conversations. I think most literacy teachers dream that students, once their days of school are over, will be inspired and educated to read and write about what matters to them. That is the kind of literate identity we want for our students.

However, the increasing pressure of standardized testing disconnects literacy education from human concerns. Students face writing prompts and reading tests that have no connection to their lives, communities, or interests. The tests are created and then read by disconnected, uninvested, anonymous readers—and now perhaps even computers. Literacy practices become less about communicating with people and more about communicating with a faceless system or a machine. What students, like administrators and teachers, learn from this system is that only the numbers matter, not the meaning or the communication.

Students and teachers have also learned that, like any system, standardized tests can be "gamed." When work is written only to be assessed rather than to communicate ideas, the activity becomes more about ensuring that certain qualities are present (e.g., the use of examples, the complexity of sentences, transition devices, vocabulary) regardless of the overall effect of the piece of writing. In the United States, one example of such an activity is the brief writing section recently added to the Scholastic Aptitude Test (SAT)—the test taken by thousands of university-bound students that is supposed to indicate their abilities to succeed in higher education. Using a 25-minute writing sample to determine a student's ability to write the kind of extended critical prose required in university is like using a person's ability to back a car down a driveway to determine whether the driver can make it through rush-hour traffic. The writing samples on the SAT are apparently scored on such generalizable characteristics of writing as smooth transitions and varied sentences rather than on content or overall effect (Klein, 2005).

So let's say specific examples are highly prized, particularly those that are not about personal experiences. Now, it doesn't matter if the specific examples are made up, as long as they are specific, and so pretty soon we'll find students sprinkling their essays with impressively "specific"
examples. What the students will know, and what we as teachers will have to admit, is that writing as a means of communicating ideas does not matter in this situation; it's racking up the right number of smooth transitions and specific examples that does. It is difficult to imagine such a situation creating the conditions to inspire students to think of themselves as writers and readers and to engage in writing with any sense of ownership or passion.

**Gaming the testing machines**

Even more troubling is the recent trend toward evaluating student writing through computer software. No longer is the student writing for any person, even an anonymous person; instead it is writing done to be judged by a series of algorithms that look for quantifiable characteristics such as transitional phrases and complex syntax. But any of these programs can be easily “fooled.” Give one an essay that is gibberish but includes the proper characteristics of syntax and vocabulary, and you receive high marks.

In fact, computer assessment software can be gamed by other computer software. Hesse (2005) pointed this out in a recent presentation. First he used the online Essay Generator (www.EssayGenerator.com) to create an essay. If you’ve never used this website, it is great fun. You enter any word or phrase and are immediately rewarded with an essay on that topic with complete sentences; sections on social, economic, and political factors; and even important-looking citations and a graph. But reading the essay more closely reveals its delightful ridiculousness. For example, I just entered the phrase “standardized testing” and received an essay with the following opening sentences:

The issues involving standardized testing has been a popular topic amongst scholars for many years. In depth analysis of standardized testing can be an enriching experience. Though standardized testing is a favorite topic of discussion amongst monarchs, presidents and dictators, there are just not enough blues songs written about standardized testing.

The essay continues with complex and syntactically correct sentences that make little or no sense at all. Hesse (2005) took his randomly generated essay, full of similar nonsense, and then submitted it to “intelligent” essay assessing software and received a report that praised the high quality of the writing, including the mature command of language and the effective use of examples and transitional devices. I replicated Hesse’s experiment with similar results (and it’s worth trying yourself if you want to be simultaneously amused and horrified). Perhaps this is the wave of the future: We can have computers write essays to be read and evaluated by other computers and leave students out of the process altogether.

Even so, the important concern in the use of computer assessment, as well as standardized testing, is not whether such systems can be fooled. Those advocating the development of such software maintain that someday computer software may be able to read student essays for content as well as for syntactic characteristics. They may be right. But the sophistication of the software is not the point. The more disturbing question is about what writing becomes when it is produced to be read and evaluated by a computer. What is the point of writing for a computer? Will there be writers who will care as deeply about what a computer thinks about their writing in the way that they care about pleasing a human audience? Will anyone take writing for a computer seriously? Students may care whether they pass a test, but what will they learn about literacy and identity when there is no human connection to what they write?

I’m a die-hard humanist on this issue. Writing and reading are about touching the mind of another person, whether in a proposal, a poem, or a polemic. When advocates of computer software or standardized testing argue that their systems will be more objective, it is an argument that again strikes me as beside the point. Writing for
other human beings—even when they are teachers—is by nature subjective. Everything about writing depends on context, culture, and the occasional unpredictability of human response. Such is the challenge and joy of writing, along with its intermittent frustration. Even as I write this column, thinking I have some sense of my audience and the context of this journal, I know that I cannot predict all the possible responses. I can’t think of a writer who doesn’t know in her or his bones that writing will always be responded to and judged subjectively. Yet standardized literacy testing and evaluation would have us pretend to students and their parents that such subjective responses can be overcome with scientific methods and better technology, and that such methods can generate a set of numbers that are meaningful about the quality of that writing. And we forget that even standardized test readers and computer software programmers are people with their own biases and preferences. We do these same students a significant disservice by acquiescing to this pretense of objectivity and by not talking to them about how real humans in real contexts read and how best we can try to identify and respond to such real people and situations.

In conversations with students about how real audiences in real situations respond to writing is where meaningful assessment can happen. If I want to know if students are learning to write, I first want to see those students write about subjects that matter to them for an interested audience and reflect on the reasons for their writing choices. Once students get a real response from an audience, I want them to reflect again on that response and what it might teach them about what they wrote and why. I always talk with students about assessment and how their work could be assessed. Then I discuss my choices during assessment and how I evaluate both their writing and their reflections on it. I’m not against assessing writing. I tell students that every piece of writing read by another person is assessed in some way, even if only in the reader’s choice to read all the way to the end. I’m just against thinking that we can assess writing through some pseudoscientific, technology-driven, one-size-fits-all, a-contextual test.

Making a difference with human response

Sommers and Saltz (2004), in their longitudinal study of more than 400 university student writers, found that the students who made the most progress in terms of writing shared two characteristics. First, these students, even if they began their university careers as relatively weak writers, were the ones who brought ideas and issues that mattered to them to their choice of courses and their responses to assignments. Equally important for these students was the ability to see writing assignments as something more than just fulfilling a requirement for a grade:

When students begin to see writing as a transaction, an exchange in which they can "get and give," they begin to see a larger purpose for their writing. They have their first glimmerings of audience; they begin to understand that they are writing for flesh-and-blood human beings, readers who want them to bring their interests into a course, not simply teachers who are poised with red pens ready to evaluate what they don’t know. (p. 139)

In addition, Sommers (2005) found that the responses from teachers that resulted in the most substantial improvement in student writers took the work and the ideas of the writers seriously and, along with constructive criticism, provided comments and questions that pointed the students toward what they might improve in future pieces of writing. Writing for a real audience about ideas that engage individual interests and intellects and having that audience provide thoughtful individualized responses not only helps students become better writers, it also helps them create their identities as writers. Such an approach to writing is the antithesis of standardized testing.

Of course, the sad fact is that there are people in and out of education who are not
concerned about whether writing with passion for real readers will inspire students and help them develop identities as confident writers. This camp wants to ensure that students attain a functional literacy that makes them productive in the workplace—in clerical and service industry jobs. As I've argued before in this column, the way literacy instruction is conceived and enacted is often connected to issues of social class, and it's the same with standardized tests as well. One benefit of standardized testing as assessment is that it is much cheaper than the kind of individualized assessment I advocated earlier. Affluent students will be taught to "game" the standardized testing system. But these same students will also have abundant opportunities in small classes in elite schools to write for real readers and to read for meaning, pleasure, and enrichment. By contrast, it is working class and poor students, for whom pleasure and the construction of literate identities is deemed unimportant, who will encounter a curriculum driven by the fear of standardized testing. The administrators and teachers of those students, facing large classes, few resources, and the threat of punishments over low scores, will be forced to create a curriculum in which students will be taught to write only for the anonymous readers and machines that evaluate such tests. Certainly that will be cheaper than providing the resources and trust in poorer schools that would allow teachers to face the same class sizes and use the same approaches for teaching writing that affluent schools enjoy. But students from poorer schools will encounter fewer and fewer opportunities to think of themselves as writers communicating their interests and passions for real readers.

What is to be done?

Many teachers, even when facing the pressure of standardized testing, continue to design assignments that fulfill the needs of assessment while providing students with places to do writing and reading that matters. And systems of assessment that are not bound up in concerns about time and technology, such as writing portfolios, can provide responsible assessment that is valuable for students, teachers, and institutions. However, writing portfolios can become as rigid and impersonal as any approach to assessment if the focus of the writing becomes the assessment itself rather than the communication of ideas.

The pressure toward standardized testing is such that more must be done than just designing good pedagogy around the margins of a test-driven curriculum. It is time for individual teachers at every level to take back the debate about assessment from the people who say the only valid evaluation of writing is timed, standardized, and faceless. It is time to take back the debate that maintains that only assessment that results in quantifiable numbers is valid. Engaging in this debate does not happen on a grand stage, it happens as a persistent, continuing conversation with the people in our communities. It is the responsibility of teachers to keep explaining to students, parents, administrators, neighbors, and newspaper editors that writing and reading that matters for real audiences is what creates literate citizens. It is the responsibility of teachers to explain how such writing can be assessed, that such instruction can be responsible and accountable, but that assessment and instruction happen differently in literacy education than in other fields. Finally, it is the responsibility of teachers not only to keep talking about what the best practices and outcomes for student literacy education should be but also to publicize our successes.

Reading and writing matter, and we teachers care deeply about them. If we want students to think of themselves as readers and writers, then those activities have to matter to students beyond learning how to game the test to avoid being punished. It is time we reclaimed the idea that having standards does not necessarily mean accepting or aspiring to standardization.

REFERENCES


---

**A LATE STARTER**

British writer Andrea Levy’s fourth novel, *Small Island* (2005, Picador), is the story of two couples, one English and one Jamaican, whose paths cross in London after World War II. Levy’s earlier novels drew on her background as the child of Jamaican immigrants growing up on a working class, north London housing estate. But the latest one is about people like her parents who were part of the postwar “Windrush generation”—so named for the former troopship SS Empire Windrush that carried hundreds of West Indian migrants to England in 1948.

*Small Island* has won some major literary prizes since its publication in Great Britain (2004, Headline). They include the Orange Prize for women’s fiction, the Commonwealth Writers’ Prize, and the Whitbread Book of the Year award. And yet Levy came to writing—and to reading—relatively late. Nowadays her house is full of books, but she acknowledges that she did not read a novel until she was 23. Television has been one of the strongest influences on her style. Levy explains that, when she writes, she sees things in her head as if they were “on the telly,” and she thinks that “television and film are more of an influence on our storytelling than we care to admit.”