

This is Not Only a Test: Exploring Structured Ethical Blindness in the Testing Industry

by Bob Broad, Illinois State University

The other articles in this issue (Elliot, Slomp, Poe & Cogan, and Cushman) gather resources from the disciplines of philosophy, validity theory, law, and cultural studies to forge a new praxis of ethical writing assessment. My contribution takes a different approach to evaluative ethics, an approach we could term political. On the current political scene of writing assessment in the United States,[1] one heavy ethical burden lies neglected: the adverse educational consequences of mass-market standardized testing, especially the testing of writing ability. This article examines failures by two prominent testing corporations to take up their ethical responsibilities to protect the “primary good” (Rawls) of educational quality. One corporation’s failure appears at the end of an evidently serious and good-faith effort to engage its responsibilities; the other corporation’s failure results from its apparent lack of effort and commitment to that task. I explain both failures by pointing to the phenomenon of “structured ethical blindness,” in which truths that stand in the way of someone’s pursuit of profit are difficult or impossible for that person to see and to take seriously. My goal is to understand structured ethical blindness in mass-market writing assessment and to suggest how our profession and our society might best respond to this phenomenon. Based on my analyses, I conclude that testing corporations cannot fairly or realistically be expected to run their businesses with careful attention to the educational consequences of their products. Their unavoidable self-interest prevents them from seeing those consequences clearly. I therefore propose federal regulation and oversight as the only apparently functional mechanism by which to counterbalance testing corporations’ pursuits of private profit with the U.S. public’s right and responsibility to protect the quality of students’ educations.

1.0 The Fall from Testing Innocence of a Student, a Teacher, and a Society

When I was a student in grade school (in the 1960s), high school, and college (in the 1970s), I nearly always enjoyed taking standardized tests. Apparently I was one of those test-takers who could guess what the test-makers were up to and select my forced-choice answers accordingly. Plus, I could come up with a decent first-draft text on a timed-impromptu writing test. As a student, not for a moment did I consider how the tests I took might have shaped my understanding of reading, writing, and learning. The standardized tests I took seemed to match what that voice on our black and white television would occasionally say about tests of the Emergency Broadcast System: “This is a test. This is *only* a test.” I frankly doubt the tests I took affected how my teachers taught me. I don’t think my teachers thought much about the SAT, ACT, or other standardized tests I was taking. *They didn’t have to.* Unfortunately, those innocent days are gone, and in the 21st century, a high-stakes, mass-market, standardized test is *never* “only a test.”

Caught up in the illusion that test scores tell us how well students have learned and how well teachers have taught, U.S. society insists on tying students’ scores to higher and higher stakes, including: students’ high-school graduation, college admissions, and college course placements, as well as teachers’ salaries and tenure. As a result of this ratcheting up of the educational consequences of tests, studying and preparing for exams (such as those formerly connected with No Child Left Behind and, more recently, the Common Core State Standards Initiative and the Every Child Succeeds Act of 2015) now have the full attention of teachers and administrators. In my home state of Illinois, vendors are making a great deal of money selling pre-packaged curricula that claim to do nothing other than raise students’ scores on the PARCC test. These new test-prep curricula have inevitably displaced more “traditional” curricula such as reading and writing assignments (and assessments) designed and judged by teachers in our schools. In this political and educational context, “Many teachers put their best judgment and best lessons aside and drill children on how to score high on multiple-choice tests” (Glass, n.d.). In other words, a test produced and sold by PARCC or SBAC (for example) is not only a test: it is a profit-making business venture, a curriculum, a pedagogy, and ultimately an educational policy affecting our society in far-reaching ways.

That early encounter with the conflicts between teaching and testing ultimately led me to a career of research and theory in writing assessment. I have committed my professional life to promoting assessment practices that support the best teaching and learning of the rhetorical arts, and along the way I have frequently condemned the standardized testing industry for the damage I believe it does to U.S. education, and especially to education in writing.

2.0 High-quality Education as a “Primary Good” We Must Protect

In *A Theory of Justice*, John Rawls introduced the key term “primary goods,” defined as “...things which it is supposed a rational man [*sic*] wants whatever else he wants” (Rawls, 1999, p. 79).

Elsewhere Rawls provided explanations as well as lists and categories of examples of primary goods:

With more of these [primary] goods men can generally be assured of greater success in carrying out their intentions and in advancing their ends... The primary social goods, to give them in broad categories, are rights, liberties, and opportunities, and income and wealth. (1999, p. 79)

I don't find anywhere in *A Theory of Justice* that Rawls specifies *education* among the primary goods. (As an ethical philosopher, Rawls goes out of his way to avoid concrete examples.) In another book, however, Rawls did comment directly on the crucial importance of education to the common good.

The state's concern with [children's] education lies in their role as future citizens, and so in such essential things as their acquiring the capacity to understand the public culture and to participate in its institutions, in their being economically independent and self-supporting members of society over a complete life, and in their developing the political virtues... (Rawls 2001, p. 157)^[2]

3.0 Mass-market Standardized Tests Harm the Quality of Public Education

As a key part of my argument, I assert that high-stakes, mass-market standardized testing harms U.S. education generally and harms rhetorical education especially. Lacking space to review the entire body of literature grounding this claim, here I will lay out a few samples of reasoning and evidence to support it.

Let's begin with the positive case: Valid writing assessment, by definition and by necessity, yields positive educational consequences. A valid writing assessment will provide teachers and students (among other stakeholders) with meaningful information about each student's writing performance *and* provide the student with information and guidance that will help her grow and improve as a writer. Brian Huot (1996) put it this way: "a valid procedure for assessing writing must have positive impact and consequences for the teaching and learning of writing" (Huot, 1996, p. 556). Looking at the negative case, Slomp et al. (2014) reinforced Huot's affirmative claim and pointed to its origins in contemporary psychometric theory: "Cronbach argues that if the negative consequences accruing from test use are severe enough, they can on their own provide sufficient cause to discontinue the use of the test..." (Slomp, Corrigan, & Sugimoto, 2014). A test with severely negative educational (and other) consequences should, Cronbach (1988) declared, be discontinued. Note that Cronbach's formulation clearly focuses on negative consequences of a particular test, not the testing industry as a whole. However, if the harms of the industry were found to be "severe enough," his argument could reasonably be applied to the entire industry. So how severe are the negative consequences of standardized tests in general?

Researchers have found consistently that one of the most damaging effects of large-scale, big-stakes standardized testing in the schools has been: (1) to oversimplify what's taught in school; and (2) to severely constrict what is taught to only those items most likely to appear on an upcoming standardized test. (Sacks, 1999, p. 128)

According to Sacks's survey of testing research, standardized tests corrupt and diminish both curricula and pedagogy, the chief professional concerns of teachers. Madaus and Russell elaborated on the harms to teachers' professionalism: "...high-stakes testing can degrade teaching skills by reducing teaching to narrow test preparation... Focusing solely on test scores also devalues teachers' judgments about the achievement of their students..." (Madaus & Russell, 2010, p. 22).

Diane Ravitch's complementary condemnation highlights those on the receiving end of this dynamic, pointing out its harms to students' intellectual development:

A steady diet of multiple-choice questions over a dozen years may impair students' ability to think critically and to reflect on alternate solutions to problems; instead, they are taught to guess "the right answer."

If they are learning to pick the right answer rather than ask the right question, they are not getting a good education. (Ravitch, 2013, p. 269)

And Perelman offers his famously colorful critique of the effects of standardized testing in the specific area of teaching and learning writing: "Education should be the enemy of bullshit... Yet mass-market testing practices and organizations, *especially when assessing writing* [emphasis added], have whole-heartedly embraced bullshit" (Perelman, 2012, p. 427). The educational researchers quoted above strongly agree that standardized tests harm the quality of public education, and that—honoring Cronbach's principle—these tests should be discontinued and replaced with assessments that support the best teaching and learning and are therefore valid and fair. This leads to the question of whether and how those who design and sell such tests engage ethical questions about possible damages caused by their products.

4.0 Two Testing Corporations' Attempts at Self-Study of Educational Consequences

Fortunately, the two best-known purveyors of standardized tests—the Educational Testing Service (ETS) and Pearson, Inc.—both have published ethical guidelines for their businesses. The critical question for this discussion is how well they attend to the *educational consequences* of their tests, and what they do about it if and when they discover that their tests are causing educational harm. By Rawls's argument and definition, this is their ethical obligation to the societies in which they do business.

4.1. ETS Standards for Quality and Fairness

ETS Standards for Quality and Fairness (Educational Testing Service, 2015) gets off to an excellent start, affirming the organization's commitment to values every educator will embrace: "our mission [is] to advance quality and equity in education for learners worldwide" (p. 1). ETS goes on to take an admirably strong stand connecting the on-the-ground performance of its tests to its lofty mission: "Every ETS product or service must be in conformity with the ETS mission" (p. 5). By its own ethical commitments, ETS insists that every test it sells "advance quality and equity in education."^[3]

ETS "Standard 4.3: Obtaining and Documenting the Evidence" (2015, p. 17) calls for test-makers to collect sixteen kinds of evidence to validate a test: e.g., "training and monitoring of raters," "statistical relationships among parts of the test," and "utility of test results in making decisions about the allocation of resources." *None* of those sixteen kinds of evidence makes any mention of educational consequences. This absence gives me a feeling of deep foreboding, because (as section 3 above demonstrates) we cannot fulfill our shared commitments to "quality and equity of education" unless we know the impact of the tests in question.

From the standpoint of the writing construct, the most robust assessment design of which I'm aware is the writing portfolio (Hamp-Lyons & Condon, 2000; Cambridge, Cambridge, & Yancey, 2009). Key elements boosting the validity of the writing portfolio over many other writing assessment designs are:

- Multiple rhetorical performances
- Performances in a variety of rhetorical situations or genres (on different topics, for diverse audiences, forums, and purposes)
- Writer's choice and development of topics about which she strongly cares; questions, topics, and issues that matter to the writer
- Robust writing strategies and processes: invention, drafting, research, response, revision, and proofreading
- Significant time (weeks and months) for pursuing various writing strategies
- Multiple and different audiences and evaluators

The good news is that teachers of writing use such assignments and assessments all the time. When they do, their classroom-based writing assessments can be valid and fair, because this approach to writing assessment meets the key requirement Cronbach, Huot, and Slomp (and others) have laid out: that a valid assessment will have "positive impact and consequences for the teaching and learning of writing" (Huot, 1996, p. 556). The bad news is that no mass-market "writing test" can meet this standard of adequate construct representation. I applaud ETS Standard 4.5, and I take it as a given that the employees of ETS fully intend to carry it out. Upton Sinclair helps us understand why they never have: The organization is blocked (phenomenologically) from seeing where their own standard is not met and also blocked (economically and existentially) from meeting their own standard, even while common practices in the teaching of writing (such as writing portfolios) clearly point the way to achieving those standards.

A final note on the *ETS Standards*. Chapter 13 presents "Test Takers' Rights and Responsibilities" (2015, pp. 51-53). As in so much of this publication, there is a great deal of material with which educators are likely to agree and that they are likely to admire, including ETS's commitments to:

- Inform test takers of their rights and responsibilities
- Provide access and information
- Obtain informed consent before releasing personally identifiable test results
- Inform test takers about how to register complaints and question scores

However, illustrating how far the blinders have closed in over the course of the fifty pages since ETS boldly proclaimed its commitment to "advanc[ing] quality and equity in education for learners worldwide," there is no mention in Chapter 13 of any "Test Takers' Rights" to quality or equity in education. By the end of the book, the organization has fallen silent on its commitment to protect these aspects of test-takers' lives, and it has averted its gaze from those places it could and should look to see whether and how it is fulfilling its stated commitments. I have tried to offer a clear, if uncomfortable, explanation for this ominous and damaging silence: "It is difficult to get a man to understand something when his salary depends on his not understanding it."

4.2 Pearson's Code of Conduct and Statement of "Social Impact"

As ETS begins its *Standards for Quality and Fairness*, so too the Pearson Corporation starts its *Code of Conduct* (Pearson, 2013) with inspiring (and aspiring) commitments to the betterment of education, ethics, and society:

We aspire to [do]... not only what is right for the business, but also what is right ethically in the wider context of culture and society. (n. pag.)

Everything we do must be driven by the needs of learners... At all times, we are focused on the best interests of learners... (p. 1)

Here are explicit commitments from Pearson to run its business with close attention to and action on behalf of ethics, society, and students.

Unfortunately, Pearson's *Code of Conduct* (2013) turns out to be a document in a very different genre from ETS's *Standards*. The *Code of Conduct* simply does not grapple with the details of how a corporation that profits from selling tests needs to investigate the consequences of its business and act to redress injustices or resolve other problems. Instead, in the course of its approximately 1600 words, the *Code of Conduct* briefly states Pearson's general commitments to, for example:

- hire a diverse workforce
- pay its employees fairly
- comply with laws
- compete fairly and honestly, and
- operate in a way that is sensitive to—and improves—the environment

When I realized the *Code of Conduct* proved inadequate for my investigation, I turned to Pearson's web page entitled "Social Impact" (Pearson, n.d.). There I hoped to find more specific, substantive, and durable commitments from the organization for investigating the educational (and other) consequences of its tests, commitments like those spelled out very clearly in the *ETS Standards for Fairness and Quality*. I was disappointed to find Pearson offering this generic marketing blurb instead: "...every product we make and sell can be measured and judged by the outcomes it helps to achieve..." (Pearson, n.d.). As with ETS, I (as an educator) admire and applaud Pearson's commitment to its products being judged by their results. (This article is my attempt to follow through on precisely that commitment from both corporations.) However, nowhere did I find specific strategies or even promises from Pearson regarding how it takes responsibility for evaluating those educational "outcomes." This lacuna resonates with the disturbing silence in the ETS document around gathering evidence of educational consequences and test-takers' rights to the promised "quality and equity in education."

Meanwhile, Pearson's *Code of Conduct* does not even begin to answer (or even seriously to ask) the most urgent question implied in several spots in the document: "What are the educational consequences of our tests and what are we committed to doing about those consequences?" Despite all the limitations of the Educational Testing Service's treatment of this question, their *Standards* document demonstrates a depth of thought, maturity, and seriousness about business ethics and educational responsibility that Pearson's *Code of Conduct* neither begins nor attempts to emulate. The ETS document appears completely sincere in laying out a series of ethical commitments that the corporation is prevented from fulfilling because, to paraphrase Sinclair, their employees' salaries quite literally depend on their not understanding or seeing the destructive educational consequences (see §3) of its products. There may exist, hidden away somewhere from my efforts to find it, a document from Pearson that attempts to do what ETS commendably attempted to do. However, based on my study of the two relevant Pearson documents I could find (*Code of Conduct* and "Social Impact"), Pearson Inc. is either unable or unwilling (or both) to attempt the ethical self-study and self-regulation that ETS attempts, albeit with devastatingly limited success.

5.0 "Test Busters": A New Regulatory Unit Within the U.S. Department of Education

One important function of government is to protect the public, the common good, from harms inflicted by those in pursuit of profits. Though neo-liberal discourse (Gallagher, 2011) tries to persuade people to view every governmental authority or action with suspicion and hostility, a quick glance at U.S. history reveals a long list of instances in which federal intervention was required to protect U.S. citizens against:

- discrimination and persecution by culture, ethnicity, gender, religion, age, class, and sexual orientation (among other characteristics)
- threats to their health from polluted water and air
- threats to their health from unsafe products (automobiles, guns, drugs, tobacco, alcohol)
- threats to their well-being from entire industries (climate change wrought by burning fossil fuels; widespread carnage wrought by poorly controlled gun sales)

Huot (2008) proposed establishment of a federal commission charged with protecting our education system from the widespread problems with standardized testing laid out above (§3). In another context (Broad et al., 2009), I proposed that the field of writing assessment eschew governmental regulation in favor of a more grass-roots approach to protecting the education system against the numerous problems that commercial, mass-market writing assessments bring. In the current scenario, however, I sign on wholeheartedly to Huot's proposal. Those of us in the field of writing assessment know enough about the consequences of standardized testing to recognize that the *laissez-faire* status quo, in which legislators and other educational decision-makers are lavishly lobbied by testing industries (Simon, 2015), is too dangerous to be educationally sustainable.

I am not an expert in federal regulatory agencies, but I will humbly point the way, as far as I can see it, that my proposal might proceed. The U.S. Department of Education (DOE) mission statement affirms that agency's commitment to "fostering educational excellence," so protecting the education system from harm seems to fall under DOE's jurisdiction. DOE already has in place at least one board and one committee dealing with assessment, but neither seems right for the responsibility on which my proposal focused.

The National Assessment Governing Board (NAGB) sets policy for the National Assessment of Educational Progress (NAEP). Because NAEP is its own system of assessment, I believe the entity I am proposing to monitor and prevent harms of standardized testing must be distinct from NAGB. The National Advisory Committee on Institutional Quality and Integrity sounds like a promising candidate for the job, but its focus is on accreditation and certification of institutions of higher education. Educators, students, parents, and concerned others will need to lobby for a new board, overseen by DOE and protected from the influence of the standardized testing industry, charged with responsibility for monitoring and preventing the harms of mass-market standardized tests. While I don't think "Test Busters" should be the new board's official name, as a nickname it has a certain ring to it.

The proposed agency will:

1. Be led by prominent scholars of teaching and learning who also have expertise in assessment.
2. Critically examine any test or assessment program with close attention to social and educational outcomes (both intended and unintended).
3. Critically examine the construct of interest with regard to its stability and variability across social, cultural, and racial divides.
4. Construct maps that clearly indicate how assessment items and scoring systems reduce construct-irrelevant variance while maximizing construct representation.
5. Identify populations of interest and design a sampling plan to ensure meaningful disaggregation of data for those populations.
6. Develop a plan for analyzing consequences (intended and unintended) of proposed assessments with special attention to how the testing program affects educational *quality, access, and opportunity to learn*.
7. Address questions of justice, how negative consequences from testing programs will be remedied (including by revision and/or suspension of the program).
8. Conduct ongoing review of any large-scale, high-stakes assessment program to ensure resources are sufficient to the task and to articulate processes by which revisions to assessment programs will be carried out.

The status quo of the testing enterprise in U.S. education is dangerous and unsustainable. Extensive damage to our educational system from mass-market tests goes unchecked, as the testing corporations gather more contracts and profits. To address this ethical crisis, the federal government needs to step in and manage this industry for the same reasons and in similar ways that it manages other risky, profitable industries. John Rawls proposes two possible courses of action when institutions are discovered to be "unjust": those institutions "must be reformed or abolished." Please note that I am proposing the milder, more liberal of these two courses of action (reform, not abolition) in response to demonstrated structured ethical blindness in the standardized testing industry.

References

- Broad, B., Adler-Kassner, L., Alford, B., Detweiler, J., Estrem, H., Harrington, S., ... Weeden, S. (2009). *Organic writing assessment: Dynamic Criteria Mapping in action*. Logan, UT: Utah State UP.
- Cambridge, D., Cambridge, B., & Yancey, K. B. (Eds.). (2009). *Electronic portfolios 2.0: Emergent research on implementation and impact*. Sterling, VA: Stylus Publishing.
- Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test Validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.
- Educational Testing Service. (2015). *ETS Standards for Quality and Fairness*. Educational Testing Service.
- Gallagher, C. W. (2011). Being there: (Re)Making the assessment scene. *College Composition and Communication*, 62(3), 450–476.
- Glass, G. V. (n.d.): Education in two worlds: Why I am no longer a measurement specialist. (n.d.). Retrieved August 19, 2015, from <http://ed2worlds.blogspot.com/2015/08/why-i-am-no-longer-measurement.html?m=1>
- Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio: Principles for practice, theory, and research*. New York, NY: Hampton Press.
- Huot, B. (1996). Toward a new theory of writing assessment. *College Composition and Communication*, 47(4), 549–66.
- Huot, B. A. (2008, April). *From writing assessment theory and practice to regulation*. Paper presented at the Conference on College Composition and Communication, New Orleans, LA.
- Madaus, G., & Russell, M. (2010). Paradoxes of high-stakes testing. *Journal of Education*, 190(1/2), 21–30.

Pearson. (2013). *Code of conduct*. Pearson. Retrieved from https://www.pearson.com/content/dam/corporate/global/pearson-dot-com/files/coc/2013/COC_Playbook_English_September_2013.pdf

Pearson. (n.d.). Social impact. Retrieved from <https://www.pearson.com/social-impact.html>

Perelman, L. (2012). Mass-market writing assessments as bullshit. In N. Elliot and L. Perelman (Eds.), *Writing assessment in the 21st century: Essays in honor of Edward M. White* (pp. 425–437). New York, NY: Hampton Press.

Ravitch, D. (2013). *Reign of error: The hoax of the privatization Movement and the Danger to America's Public Schools*. New York, NY: Knopf.

Rawls, J. (1999). *A Theory of Justice* (Rev. ed.). Cambridge, MA: Belknap Press.

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.

Simon, S. (2015, February). No profit left behind. *Politico*. Retrieved from <http://www.politico.com/story/2015/02/pearson-education-115026.html>

Slomp, D., Corrigan, J. A., & Sugimoto, T. (2014). A framework for using consequential validity evidence in evaluating large-scale writing assessments. *Research in the Teaching of English*, 48(3), 276–302.

Notes

[1] I am aware that my first-hand knowledge of education systems is limited to the United States. For this reason, I attempt to acknowledge the U.S. as the society about which and to which I am mainly trying to speak. At the same time, my more modest knowledge of education systems in Canada, the United Kingdom, Israel, and France leads me to hope and believe that my analyses and arguments may prove fruitful in contexts beyond the U.S.

[2] Thanks to Norbert Elliot for suggesting this later book by Rawls with its more direct commentary on education as an “essential” common good.

[3] I note with interest that the articles by my colleagues Poe & Cogan and Elliot tend to emphasize the *equity* part of ETS's two-part formulation (“quality and equity in education”), whereas Slomp and I seem more focused on the *quality* element. Of course both elements—equity and quality—deserve careful attention in a discussion of the ethics of writing assessment.

Copyright © 2021 - *The Journal of Writing Assessment* - All Rights Reserved.