

Zack Quaratella

Teaching to the Test: the Good, the Bad, and the Necessary

This literature review will cover in broad strokes the progression of thought on standardized testing since the late 1970s. First, to give historical context, the history of the standardized test in the United States will be briefly considered. The purpose of such a reflection is a reminder to oneself that history often reveals trends that the fog of present opinion can obscure.ⁱ We will then give a general account of the current state of standardized testing, though the literature is admittedly clouded with constraints presented by politics, student and teacher privacy, and private corporations. We consider the merits of extrapolative projects conducted by biased test-makers and narrow case studies presented by practitioners. Having reviewed the history of the standardized test and considered (to the best of our ability) the present state of testing in the United States, this essay will then discuss the various pedagogical methods proposed in the literature for helping students succeed on standardized tests.

In the United States, standardized tests developed as a fact of life for students early in the twentieth century. Starting in the 1920s, norm-referenced tests were used to test incoming army personnel, and by the 1930s the SAT sold “about 1.5 million booklets per year.”ⁱⁱ Following World War II, and into the Cold War, standardized testing came to be synonymous with creating a competitive global citizenry, as embodied by the National Defense Education Act of 1958. Again in the 1980s, with the frightening outlook posted by *A Nation at Risk*, testing became a desperate attempt to shore up obvious faults in the nation’s education system. All this led Chris Pipho of the Education Commission of the States to comment, “Nearly every large education reform effort of the last few years has either mandated a new form of testing or expanded uses of existing testing.”ⁱⁱⁱ Finally, at the turn of the 21st century, testing has once again taken on an even

more significant role in students' and teachers' academic lives with the advent of "high stakes" testing. More specifically, "Since 1998, social studies achievement tests have been present in about half of the states, with various states adding, dropping, or changing structure over time" and by 2010 eleven states featured social studies exit exams.^{iv} Despite recent appeals by preeminent educational historian Diane Ravitch to "opt out" of state testing, for many teachers who rely on their jobs as a source of income and security from unemployment and homelessness, this is simply not a feasible option.^v Therefore, the current debate lies with persons who hold more authority than teachers: scholars, school board members, test makers, and politicians.^{vi}

Of those who are able to engage in the debate, which groups support standardized testing, and what are some of their arguments? In 1977, Herbert Rudman wrote an article published in *The Phi Delta Kappan* attacking the "antitesting" movement. He characterizes his opponents condescendingly as "free lance writers; professional association staff members; and professors of biology, physics, mathematics, psychology, engineering, anthropology, and medieval history."^{vii} Having discredited his opponents' professions and claiming them to be outliers, Rudman makes the case for supporting standardized testing in schools. He begins by describing the process in which tests are crafted, claiming that they require 8-10 years of careful research. Anticipating the complaints of his adversaries, Rudman claims that this process ensures "curriculum is *reflected* by the achievement test and not determined by it."^{viii} This is accomplished by gathering a group of content experts who aggregate national social studies curricula and write items that anticipate "where the curriculum will be eight years from the time [they] start to write and revise the test, and where it will likely be for the greater part of the life of the test."^{ix} Rudman does not describe how the content experts are able to anticipate where curriculum will be a decade in the future. Addressing another common criticism that tests were used too heavily to decide student tracks,

Rudman claims “Tests yield data when responded to by children. Teachers and administrators translate these data into information.”^x He feels that “anthropomorphizing” tests is sensationalist; tests do not determine student outcomes, teachers do.

While much has changed since 1977 in the ways standardized tests are defended, much has also remained the same. Like Rudman, a recent collection of essays commences with a foreword by John Fremer that attacks the credibility of opponents of standardized testing. Fremer claims that there is a media bias against testing because “The sources most often cited are individuals or groups that are fundamentally opposed to all major applications of high-stakes testing in school and society.”^{xi} Interestingly, too, the book’s first essay describes a positive public opinion of standardized testing that *also* labels opponents of testing simply as overly vocal outliers.^{xii} According to this essay, public opinion has always supported testing. Moving into specific arguments for the merits of testing, Dean Goodman and Ronald Hambleton of the University of Massachusetts at Amherst argue there are common misconceptions about educational assessments. They contend that comments which claim tests do not align with state standards are not fair because the *No Child Left Behind Act of 2001* mandates that tests align with state standards.^{xiii} Goodman and Hambleton also claim that despite being capstone graduation tests, there is actually not too much emphasis placed on standardized tests because “students are given multiple opportunities to pass these tests.”^{xiv} Aside from simply discussing misconceptions, advocates for testing also argue that they “raise standards” and “hold schools accountable.” After the Virginia SOLs were passed by only 3% of schools in 1999, Virginia Board of Education President Kirk Schroder responded that, “these failure rates demonstrate beyond a shadow of a doubt that the Virginia Board of Education has raised academic expectations for our students in a major, yet reasonable, way.”^{xv} On the whole, supporters’

arguments are generally grounded in positivism and a sense of objectivity.

By contrast, individuals and groups who oppose standardized testing typically use more qualitative methods to make their arguments. Because privacy issues or protective testing companies often limit access to critical data, critics of standardized testing often resort to unfortunately shoddy academic works. However emotional or vitriolic their arguments can sometimes be, they are often grounded in the opinions of actual practitioners (which is usually not the case for proponents of testing). For example, in James Popham's response to Rudman's claim that schools do not teach to the test, Popham writes, "I have spoken with a good many teachers and school administrators who, although unwilling to be quoted, report devising outright 'let's bump up our standardized test scores' programs."^{xvi} Early on, the two sides had difficulty conversing in reasonable academic debate.

As the debate matured, authors began to construct qualitative meta-analyses on the major complaints against standardized tests. In 1989, Walter Haney and George Madaus gave a list of four major categories:

Standardized tests: give false information about the status of learning in the nation's schools; are unfair to (or biased against) some kinds of students; tend to corrupt the processes of teaching and learning, often reducing teaching to mere preparation for testing; and focus time, energy, and attention on the simpler skills that are easily tested and away from higher-order thinking skills that are easily tested and away from higher-order thinking skills and creative endeavors – the Achilles' heel of the nation's education system today, in the view of many observers.^{xvii}

As these are still the main rallying cries of anti-test apologists, I will treat the first two individually and the second two together with examples from the literature.

The first charge, that tests give educators false information, has grown tremendously in importance since the rise of high-stakes testing in the early 2000s. However, as is reflected by Leslie Salmon-Cox in 1981, teachers simply do not prefer using test scores to solely evaluate students because they do not provide a full account of that student's ability. When conducting a

survey of teachers, Salmon-Cox found that “Teachers tend to assess students by using observational and interactional techniques.”^{xviii} Put metaphorically, “Teacher preference, in effect, is for continuous movies, in color with sound, while a test score, or even a profile of scores, is more akin to a black-and-white photograph.”^{xix} Despite findings like these, testing has grown monumentally in importance since Salmon-Cox wrote; modern policy-makers, it seems, prefer black-and-white photographs. D. Monty Neill and Noe J. Medina make a cognitive argument in regard to testing fallacies: “Unitary test scores,” they write, “and linear scaling of scores ignore the true complexity of human intelligence and thus provide a deceptive picture of individual achievement, ability, or skills.”^{xx} The information will *necessarily* be false because it is attempting to describe a characteristic in humans that simply does not exist. As our understanding of cognitive psychology expands, our ability to treat human brains as a single mass that can be tested and scaled will become increasingly more questionable. From a mathematics perspective, tests also often incorrectly label students as underperforming: “Ability or achievement tests that use a linear scale can label a student’s performance as incorrect or substandard when it is really only a normal variation; such tests can also mask real differences.”^{xxi} From this, students can be placed in certain tracks because of inadequate sample sizes or varying rates of cognitive development. As standardized tests are increasingly becoming exit examinations, teachers inevitably glean even more precarious data; concepts that one class of students struggled with may come easily to the following year’s students. Given that social studies curriculum in high school is broken into content areas (World History I, World History II, World Geography, Virginia and U.S. History, and Virginia and U.S. Government), it is difficult for administrators and teachers to show student growth from year to year.

The charge of bias in standardized testing is an attempt to help explain why the

achievement gap has not been closed. It strikes at the heart of the testing movement: if tests are not truly objective (meaning equal for all), then why switch from subjective teacher grading anyway? Neill and Medina argue that in the case of standardized tests, “*objective* merely means that the tests can be scored by machines, not by subjective human scorers.”^{xxii} The tests themselves pretend to have objectivity, but “test developers generally validate the content of a test by asking subject-area experts to make qualitative judgments about the relationship between individual test items and the trait or traits that the test seeks to measure – a method occasionally referred to as BOGSAT (Bunch of Guys Sitting Around a Table).”^{xxiii} They simply seem to be trading one type of subjective opinion with another, swapping teachers and local education leaders for unknown “subject-area experts.” To that end, Rita Platt begs the question in her article “Standardized Tests: Whose Standards Are We Talking About?” After teaching in a remote part of Alaska, Platt contends, “If students throughout America were being held accountable to standards that did not match the standards of their cultures or communities, most students would fall short of the test targets.”^{xxiv} While Platt presents an extreme example of “outsiders” both geographically and culturally, other authors contend that minority and low-income students in the contiguous United States also suffer from similar effects as the Alaskan students. When asked what to do about a small cut, for example, researchers found that students of the majority class responded “apply a Band-Aid” while students in low-income situations responded “suck on it.” Those students may never have owned or used a Band-Aid in their lives.^{xxv} While supporters of testing will argue that these claims are overblown and controlled for by test designers, opponents pose an interesting dilemma: if standardized tests are norm referenced, someone must fail them. The question becomes, who fails? In the employment sector, bias against minorities and women was recognized in hiring exams; under Title VII of the

Civil Rights Act employers are largely barred from using the very tests schools rely upon to evaluate student achievement.^{xxvi}

Teachers themselves best answer the question of whether achievement tests corrupt teachers' classrooms and the depth of thinking students are doing in preparing and taking tests. In the intensive case study conducted by Kathe Taylor and Sherry Walton, the authors found that "Despite the rich learning environment and despite teachers' assertions that students routinely demonstrate their knowledge and skills, this school's median achievement test scores declined steadily for three years."^{xxvii} Likewise, a study by Nelson Maylone proposes a new word to describe the skill of taking a test: "TestThinking". Maylone argues that "the skills standardized tests truly measure may be useful in only one context: taking tests."^{xxviii} While supporters of testing claim that this is not necessarily a negative concept, others worry that testing damages more "authentic" curricular efforts. Scott W. DeWitt, et al. recently conducted a four state analysis of whether state tests construct "higher-order" or "lower-order" items. They found that even when state standards demand higher-order thinking, the items on state tests typically ignore those standards because there is a difficulty in constructing multiple-choice questions that "require students to demonstrate higher-order cognitive skills."^{xxix} Therefore, it is easy to see why teachers complain that they do not have the time to work on higher-order thinking with students: tests do not expect students to exhibit them.

With the history of standardized testing and a review of the scholarly debate in mind, where do we as educators fit into this debate? While I agree with the view of Linda Darling-Hammond that "as teaching looks more like testing, U.S. students are doing less writing, less science, less history, reading fewer books, and even using computers less in states that will not allow their use on standardized tests," high-stakes testing will not evaporate suddenly as a result

of these drawbacks.^{xxx} Therefore, as argued by Royal Van Horn, “I am reminded of my least favorite quote of all time, ‘Perception is everything.’ I don’t like the fact that peoples’ perceptions often differ from reality, but you really ought to want your class not only to do well but also to be perceived as doing well.”^{xxxix} Teachers generally seem to understand that high-stakes tests do not accurately assess their students’ abilities, but because other groups (administrators, parents, presidents) believe this to be the case, teachers must also understand that they owe it to their students to do everything pedagogically and ethically that they can to help students succeed on the exams.

Although “TestThink” was devised largely as a critique of standardized testing, Maylone’s concept can be useful for teachers. Some students, Maylone argues, are TestThinkers; they “always simply focus on what the test-makers want – even though, without a stretch, at least two of the choices in this item make sense.”^{xxxix} While Maylone seems to consider this sort of skill to be innate or culturally acquired, teachers can also show students how to view items in this light. As Taylor and Walton contend, when students are subjected to a workshop drilling specific test-taking skills, their scores and attitudes about testing will improve. They argue for these workshops in absence of the social studies content area.^{xxxix} So, while Maylone is somewhat defeatist, Taylor and Walton embrace the notion that standardized tests affect students’ lives, so teachers have a duty to help student do as well as they can.

Royal Van Horn offers several suggestions for improving students’ scores on their standardized tests. Many of his suggestions are simply about good teaching, for example maximizing one’s resources and ensuring students are being productive with their time. However, Van Horn also notes that “teaching to the test” does not always mean teaching memorization tricks: “A skill-by-skill approach to raising achievement test scores is not wise...

Educators who do not understand this principle are the kind of people who think it makes sense to go to an auto parts store and buy all the little pieces necessary to build a car.^{»xxxiv} Van Horn argues anecdotally that holistic approaches to education help students improve test scores: “I can always pick out the drill-and-practice teachers and those who use a more balanced approach that includes a lot of student writing and emphasizes children’s literature. Guess what, the students who write a lot and read a lot do much better on the tests than those who do lots of dittos.”^{»xxxv} This sort of ambitious teaching practice is touted by S.G. Grant, who argues that teachers “not let the curriculum and test dominate what [they] believe makes most sense for [their] students with regards to relevancy or engagement.”^{»xxxvi} Grant assures teachers that if they continue their authentic lessons, students will do well on these tests. Van Horn also argues quantitatively to address the needs of exceptional students; high and low scores can greatly affect the average score of the class and also of the school. As teachers, we also hold the power to decide when we give students our benchmark exams, so Van Horn offers the obvious advice to avoid giving students several standardized tests in a row.^{xxxvii} If this requires coordinating with other teachers, do so; if this requires bartering with school administrators, do so with the knowledge that these tactics are now part of the profession of teaching.

The mission of getting students to pass standardized tests has become an integral part of the teaching profession. The ultimate utility of these kinds of tests is debatable: some argue that testing should be ramped up further to get more accurate data on the problems of education, while others lambast the testing movement for destroying the sanctity of the classroom and choking the creativity out of students. The purpose of expanding on the debate in the pages above was to show that these arguments are not above teachers’ heads. Supporters of standardized testing show teachers the value of data-driven learning; in large classes, tests can be

used to prevent students from falling through the cracks. Opponents warn against relying too heavily on testing for any important educational function; humans are too complex to be accurately quantified in a single multiple-choice test. In my classroom this spring, I will balance the theoretical underpinnings of this paper with the practical applications of the final few authors. Realizing the benefits and flaws of tests, I can accurately use them to help students' learning in my classroom. I can use the district-wide benchmark multiple-choice tests to quickly assess how well students are following my lessons. On the other hand, I will be cautious about making long-term judgments about students' academic skills based on testing. Finally, knowing that student's lives are now decided by standardized testing, I understand that I must use strategies to help them succeed. This does not mean "teaching to the test" using lectures, but it means constructing an authentic curriculum based on the standards implemented by the state in which I teach. Ultimately, it is my task as a social studies teacher to balance the needs of students with the demands of the tests; these strategies and theories offer the best way of doing so.

ⁱ Van Hover, et al. "From a Roar to a Murmur: Virginia's History & Social Science Standards, 1995-2009" *Theory and Research in Social Education* 38:1 (Winter 2010), 83.

ⁱⁱ Walter Haney and George Madaus, "Searching for Alternatives to Standardized Tests: Whys, Whats, and Whithers," *The Phi Delta Kappan* 70:9 (1989), 684.

ⁱⁱⁱ Chris Pipho, "Tracking the Reforms, Part 5: Testing- Can it Measure the Success of the Reform Movement?," *Education Week*, 22 May 1985, 19.

^{iv} Scott W. DeWitt, et al. "The Lower-Order Expectations of High-Stakes Tests: A Four-State Analysis of Social Studies Standards and Test Alignment," *Theory & Research in Social Education* 41:3 (July 2013), 390.

^v Diane Ravitch, "How Standardized Tests Harm Students and How to Stop This Harm," *Diane Ravitch's Blog* (blog), September 26, 2013, <http://dianeravitch.net/2013/09/26/how-standardized-tests-harm-students-and-how-to-stop-this-harm/>.

^{vi} James Popham, "The Standardized Test Flap Flop: An effort to refute major arguments offered by Herbert Rudman." *The Phi Delta Kappan* 59:7 (1978), 470; Van Hover, et al. "From a Roar," 82-115.

^{vii} Herbert Rudman, "The Standardized Test Flap: An effort to sort out fact from fiction, truth from deliberate hyperbole." *The Phi Delta Kappan* 37:4 (1977), 179.

^{viii} *Ibid.*, 180.

^{ix} *Ibid.*

^x *Ibid.*, 181.

^{xi} John Fremer, "Foreword: The Rest of the Story," in Richard Phelps, editor, *Defending Standardized Testing* (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2005), vii.

-
- ^{xii} Richard Phelps, "Persistently Positive: Forty Years of Public Opinion on Standardized Testing," in Richard Phelps, editor, *Defending Standardized Testing* (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2005), 21.
- ^{xiii} Dean Goodman and Ronald K. Hambleton, "Some Misconceptions About Large-Scale Educational Assessments," in Richard Phelps, editor, *Defending Standardized Testing* (Mahwah, New Jersey: Lawrence Erlbaum Associates, 2005), 93. I am aware this is a circular argument.
- ^{xiv} *Ibid.*, 95.
- ^{xv} Van Hover, et al., "From a Roar," 96.
- ^{xvi} Popham, "The Standardized Test Flap Flop," 471.
- ^{xvii} Haney and Madaus, "Searching for Alternatives to Standardized Tests", 684.
- ^{xviii} Leslie Salmon-Cox, "Teachers and Standardized Achievement Tests: What's Really Happening?" *The Phi Delta Kappan* 62:9 (1981), 634.
- ^{xix} *Ibid.*, 632.
- ^{xx} D. Monty Neill and Noe J. Medina, "Standardized Testing: Harmful to Educational Health" *The Phi Delta Kappan* 70:9 (1989), 689.
- ^{xxi} *Ibid.*
- ^{xxii} *Ibid.*, 691.
- ^{xxiii} *Ibid.*, 690.
- ^{xxiv} Rita Platt, "Standardized Tests: Whose Standards Are We Talking About?" *The Phi Delta Kappan* 85:5 (2004), 387.
- ^{xxv} Neill and Medina, "Standardized Testing," 692.
- ^{xxvi} Haney and Madaus, "Searching for Alternatives," 685.
- ^{xxvii} Kathe Taylor and Sherry Walton, "Co-Opting Standardized Tests in the Service of Learning," *The Phi Delta Kappan* 79:1 (1997), 66.
- ^{xxviii} Nelson Maylone, "Testthink" *The Phi Delta Kappan* 85:5 (2004), 383.
- ^{xxix} Scott W. DeWitt, et al. "The Lower-Order Expectations of High-Stakes Tests: A Four-State Analysis of Social Studies Standards and Test Alignment" *Theory & Research in Social Education* 42:2 (July 2013), 388.
- ^{xxx} Linda Darling-Hammond, "Building a system for powerful teaching and learning," in R.L. Wehling (editor), *Building a 21st century U.S. education system*, (Washington, D.C.: National Commission on Teaching and America's Future, 2007), 71.
- ^{xxxi} Royal Van Horn, "Improving Standardized Test Scores," *The Phi Delta Kappan* 78:7 (1997), 585.
- ^{xxxii} Maylone, "TestThink," 385.
- ^{xxxiii} Taylor and Walton, "Co-Opting Standardized Tests," 69.
- ^{xxxiv} Van Horn, "Improving Standardized Test Scores," 584.
- ^{xxxv} *Ibid.*, 584.
- ^{xxxvi} S.G. Grant, "The Road to Ambitious Teaching: Creating Big Idea Units in History Classes," *Journal of Inquiry & Action in Education* 2:1 (2009), 8.
- ^{xxxvii} Van Horn, "Improving Standardized Test Scores," 585.