

Introduction: This Is *Not* a Test

[Testing] is the mechanism we have. We may not like it.

—Tony Bennett, Indiana State Superintendent of Public Instruction, 2008

This is an actual emergency. Our schools are under attack, and with them, the future of our young people. What's more, this assault isn't being perpetrated by some foreign power bent on our destruction. No Red Hand, no Shining Path, no al-Qaida. This assault is coming from within.

What's worse, the assault on our schools and our children's future is one we, the people, have allowed to happen and seem powerless to resist. We are referring to the barrage of standardized testing besetting our schools and districts. No Child Left Behind is only its most recent, and most punishing, incarnation. And the Obama administration's proposals for "amending" NCLB reflect a similarly misguided reliance on test scores as the primary measures of success for students, teachers, and the education system as a whole. For decades more and more tests have been seeping into our schools, sapping the energy and enthusiasm of educators and draining the life from children's learning. And while some of the motivation for this burgeoning movement is clearly commercial, it is at least partly driven by what we have come to think of as "the tyranny of good intentions."

Those are some pretty dire charges, but we argue that, because of the way we insist on using these tests, they damage society in ways that far outweigh the minimal benefits they confer—whether the tests are used for "measuring achievement" in the K–12 schools or for helping to determine admission to college. We hope to persuade you that you need to look beyond the good intentions expressed by those who continue to support this plague of tests, and

we plan to arm you with some basic understanding of standardized tests and the assumptions—almost never discussed in public forums—that underlie them. As the title of our book suggests, these assumptions are largely mythical. Like myths, they embody a system of beliefs that characterize a particular culture. But unlike myths, we use the assumptions that underlie standardized tests to make policies and drive practices that influence the lives of our children and so greatly affect our future.

If you're experiencing serious *déjà vu* about now, we sympathize. If the thought of reading still more complicated and confusing information about standardized testing—including all those arcane details about margins of sampling error and reliability coefficients—makes you want to run screaming from the room, take it from us, we're right behind you. So why, then, have we subjected ourselves to the prolonged project of creating this book, and why do we hope that you'll agree that the time you give it is time well spent?

The answer lies in that very sense of *déjà vu*. We have all heard the complaints about standardized testing before. Walter Lippmann was among the first to make them—in the 1920s. And Banesh Hoffmann, back in 1962, raised serious questions in *The Tyranny of Testing* about what standardized tests could really tell us about the achievement or the aptitude of our youngsters. (This book has been reissued in paperback, by the way, and is available on Amazon.com and other websites.) More recent critiques have come from Peter Sacks, Alfie Kohn, W. James Popham, Daniel Koretz, Gerald Bracey, and many other commentators and researchers, all sounding off on the proper role of testing in America. They've warned us of the distortions that inevitably arise when we ignore the capabilities of our technologies. And they've told us that many of the test-based inferences that politicians and pundits routinely draw about American schools and children are unsound. In addition, FairTest and its Assessment Reform Network, along with other groups devoted to the rational use of tests, work hard every day to make the limitations of standardized tests clear.

So haven't we heard such criticism often enough? Maybe not. Look at that list of names in the preceding paragraph. How many do you recognize? We're betting that most readers will be unfamiliar with nearly all of them, and we think that, as educators, parents, and concerned citizens, you will want to learn more about them and their work. If the messages these commentators have been sending for decades had been received and digested by the public, we wouldn't be asking you to spend a few hours with us and think this problem through. We would all have moved on to the myriad other problems that confront schools and educators every day, problems that won't be fixed by another truckload of test scores.

But though the messages have been sent, they haven't gotten through. The repeated remarks of America's opinion leaders make this fact painfully clear. For example, Peter Wood is executive director of the National Association of Scholars, an organization of conservative academics. Presumably, he's smart enough to recognize nonsense when he hears it—nonsense like the negative characterization of U.S. students' performance in science, which Bill Gates derived from the trends of international test scores and offered to the nation's elected leaders. Wood's own argument about the role of American culture in driving students away from the serious study of science may or may not hold water. That's an argument for another book. We quote him here because he accepts unquestioningly Gates's assumption that the trends in scores on international tests are a fair and accurate reflection of math and science education in America. (We take a brief look at what international comparative assessments can be good for in chapter 1; it's not a long section.) Wood summarizes Gates's words without apology or explanation, and presumably with a straight face:

Our record on high-school math and science education is particularly troubling. International tests indicate that American fourth-graders rank among the top students in the world in science and above average in math. By eighth grade, they have moved closer to the middle of the pack. By 12th grade, our students score near the bottom of all industrialized nations. As a result, too many of them enter college without even the basic skills needed to pursue a degree in science or engineering.¹

Of course, Bill Gates has *numbers* to back him up. And although we Americans say we hate math—and Gates's comments about poor learning in math and science fall squarely within that cultural comfort zone—we nonetheless love numbers. We treasure them and bring them out for display on special occasions. Batting averages, game-winning RBIs, free-throw percentages, the Dow Jones average. You name it, and if it's got numbers that we can arrange in an ordered list, then we want that list. And if you have numbers over an extended period of time, then we can create trends. We love those trend numbers even more. We want to compare the "home team" with everyone else. We want to know: Who's in the lead? Who's moving up? Who's moving down? And like the seers of days gone by, we use our numbers to try to predict the future.

The Book of Lists was a best seller in the 1970s, and if anything, our national obsession with ranking has only gotten stronger with the arrival on the scene of computers and the World Wide Web. Today, anyone can crunch up some crispy numbers, and everyone can be a list maker. Think of the reports

in your local newspaper whenever your state releases test scores for local schools. Do you see the equivalent of a league table? An ordered list of the winners and losers? We thought so. And here in our own university town, each year some faculty members rail against the proliferation of college and department rankings—the one by *U.S. News and World Report* may be the most famous, but it is far from alone. Yet when a list maker rates one of Indiana University's schools or departments in the top ten, that's news—front-page news in the local paper and news for dissemination by the university's public relations staff. Numbers, whatever their provenance, are deemed to measure some underlying reality.

There's little harm in playing around with numbers and lists as long as you understand where the numbers come from and what—if anything—they mean. However, we greatly fear that this understanding is largely missing from today's public discussions of test scores, and we're not at all sure that very much talk about what the scores mean even takes place. And that talk should be taking place all the time, because the outcomes of these tests are being used to determine what your children learn and what opportunities will be open to them.

But most Americans seem to hold views about testing that can best be described as “intuitive,” though not necessarily correct. Henry Braun and Robert Mislevy have written about “Intuitive Test Theory,” and they liken it to the kind of intuitive beliefs about physics that people adopt as children—and then cling to for the rest of their lives.² These are beliefs like “heavy objects fall faster than light ones.” Seems to make sense and works well in daily life, except perhaps in physics class. But it's totally wrong. Such ideas work pretty well for navigating the world, but you wouldn't want to base a moon landing on them. While we won't be dealing explicitly with the intuitive ideas cited by Braun and Mislevy, many of the assumptions about standardized testing that we treat are the result of such widely held—and fallacious—beliefs.

We hope that this book will enable you to think less intuitively about tests. We think every concerned citizen ought to be raising serious questions about the standardized tests used in their schools, about the decisions that are based on the outcomes of those tests, and about the potential for harm as a result of those published school “report cards.” We hope to enable you to ask questions of the people who are making the decisions—as citizens first, but also as parents and educators yourselves. Talk with your children's teachers and your school's administrators. You might be surprised to find many of them less than comfortable with the current situation. Question your local superintendent, your school board, and your elected officials at the community and state levels. One of our goals is to provide you with enough information and especially with sources of additional information to enable you to ask hard

questions—and to follow up on the superficial answers that we fear you will encounter all too often.

Why do we think many of the answers you get from school leaders and politicians are likely to be superficial? During the politically charged month before the 2008 elections, Indiana's candidates for state superintendent of schools took part in a community forum in Bloomington. (Now, we don't intend to pick on Indiana's education leaders any more than on the leaders of other states; we just happen to live here. But we think you'll find the comments of our leaders in line with what you hear where you live.) Both candidates were reasonable and experienced superintendents. However, when the subject of standardized testing came up, their good judgment deserted them. First, they pointed to problems that special education students and those still learning English have with the tests, then they noted that tests provide only "one measure of school success," and finally they expressed a desire to communicate outcomes more clearly to parents—sensible, if not deep, positions. But then the eventual winner of the election argued, unchallenged by his opponent, that testing is here to stay. "That is the mechanism we have. We may not like it."³

AN AERIAL VIEW

Every military leader knows it's always valuable to have an aerial view of the terrain to be contested. Before we address the individual assumptions that underlie standardized testing, we look briefly at this "big picture" in chapter 1. That aerial photograph will give you the lay of the land that we'll be traversing together. In chapter 1, we look briefly at the accountability system we've created—mostly in the past twenty-five years—and examine the confusion of purpose that it reflects. We take up recent proposals to improve that accountability system in chapter 10, and we sort out the confusion of purpose that afflicts our schools in chapter 11, concluding that many of the perceived problems of today's schools are better understood and would be better resolved by working to return the schools to their historical purpose: the preparation of the next generation of citizens for our democratic republic.

Finally, in a short concluding section, chapter 1 briefly explores the terra incognita of international assessments of educational achievement. We ask both what they are good for and what they're *not* good for. And we find that, like the unknown territory at the edges of a medieval map, in the land of international assessments, "There be dragons." In short, the international assessments are mostly a distraction for the public and for policy makers and are useful primarily for what professional educators and students of pedagogy might be able

to learn from some of the better ones conducted in recent decades, especially those efforts that have sought to go behind the classroom door.

UNQUESTIONED ASSUMPTIONS

But, to paraphrase our state superintendent, if we don't like the mechanism we have, we can change it. We put it in place. Yet in order to see that we can change, we have to stop for a moment and think through what the tests are capable of telling us and what we really want to know. They're not always the same things, and they are things we rarely pause to think about.

Indeed, the idea for this book grew out of discussions we've had over the years about exactly what are the assumptions that underlie standardized testing. Here are some questions that will help you focus on the assumptions we'll be dealing with in some detail in the following chapters. Think back and ask yourself about the many ways most of us never pause to consider what's up with standardized tests:

- Have you ever thought about how well students' knowledge and skills can be assessed by the limited sample of content included in a forty-five-question test? What does a score on that test tell you about the vast range of content that simply can't be included? (See chapter 2.)
- Have you ever talked about the high achievement at a particular school when all you really knew about the school was the average test scores of its students? (See chapter 3.)
- Have you ever argued—or heard someone argue—that what we need is objective information about student achievement? For most people that word *objective* used in a school context automatically means standardized test scores and very little else. (See chapter 4.)
- Have you or your school system ever handed out punishments or rewards to schools, to teachers, or to individual children based on their test scores? How motivational are such practices? (See chapter 5.)
- Have you ever thought that improvement in scores on “high stakes” tests is a sound indicator of improvement in learning? (See chapter 6.)
- Have you ever wondered about whether the tests have an effect on the curriculum and on classroom life? Have you ever questioned what's left out to make time for the tests themselves and for the often extensive preparation for them? (See chapter 7.)
- Have you ever given more weight to an “indirect” measure (a standardized test score) of student achievement than to a “direct” assessment of achievement? Direct assessments range from judgments teachers make

to your own reading of your children's work to the response of those who attend a school performance or a school open house. (See chapter 8.)

- Have you ever thought that moving to a district or attendance area with high test scores would mean high achievement and success in life for your children? How well do standardized tests forecast future success—in school, of course, but also throughout life? (See chapter 9.)

There are literally dozens of other assumptions that all of us—from policy makers to school officials to ordinary citizens—routinely make about standardized tests. We will focus on those that grow out of the questions listed above, but once you start to think this way about the tests, you'll come up with other questions that will reveal other infrequently examined assumptions. If you come up with some other questionable assumptions and you'd like some feedback, feel free to share them at <http://thoughtsonstandardized-testing.blogspot.com>. We'd be happy to hear what you think, and we think the discussions generated might be both revealing and useful.

Here are a few additional assumptions that we don't deal with in any detail but are surely worth thinking about. The tests are timed. Is faster necessarily better? Better for everything? When your state reports proficiency, what does that mean and who says so? Is answering a question—whether filling in a bubble or writing in a short response—in any way equivalent to finding, posing, and solving a real problem in context? Why do standardized achievement tests assess the particular array of skills that they do? Who decides on that array, and how?

We'll stop here, but you don't have to. Think about what we tacitly assume to be true about the system of assessment that is consuming ever more time in children's school days. And ask yourself if what we're learning from our assessment system is worth the price.

MODUS OPERANDI

We promise this will be our only Latin heading. We pledge not to cave in and go for something close, like “valedictory.” This is our shorthand way of saying, “Here's how we're going to address the questions and assumptions we just listed.”

We will work in two ways. First, we will appeal to your sense of logic and common sense. However, in the world of assessment, common sense can often deceive us, so we'll also refer you to the experts in the field who will tell you in their own words what's appropriate to assume about tests and what's not. We'll examine the logic of the assumptions underlying standardized testing, draw on

the judgments of experts in the field, and relay that information to you, along with our opinion of what it all means for our schools and our children.

From the general tenor of these pages, you won't be surprised to learn that we dispute many of the basic assumptions embodied in the questions we listed above. We don't dispute them because we have some ax to grind where testing is concerned. We were once fans of the tests, and we've included our own personal testing histories—or Testing Autobiographies—as brief interludes throughout the book so that you can form some opinion of what led us to our current position. We hope that seeing how we moved from naive but essentially *pro*-testing positions to where we stand today will persuade you to consider the arguments we make and then to explore some of the sources we cite. If you do so, we believe your testing history will follow a similar path. All of those to whom we showed early drafts of this book automatically began to recall their own experience with standardized tests, and we invite you to reflect on your experience and that of your children and any other individuals whose lives you know well. What role did the testing play in their lives and in the decisions they made?

As our testing autobiographies make clear, we acquired our skepticism about the assumptions from a combination of our own experience and because the large body of psychometric science just does not support them.⁴ We didn't conduct this research ourselves, and we are not assessment experts. Instead, we base our judgments on the published works of a number of eminent psychometricians and policy analysts who have examined these issues carefully through the lens of their professional expertise. Sometimes they have conducted original research, and sometimes they have gathered and synthesized the research of others. We find their conclusions persuasive—for statistical and logical reasons and because they mesh well with our own experience. We will quote from these experts liberally, in part to give you the flavor of the conclusions in the authors' own words and in part to persuade you to seek out their published works. If you care about where our schools are headed and about the future of all our children, we encourage you to follow up and seek more information than we have the space or the expertise to give you here. In a sense, we hope that this book will function somewhat like a Web portal: you won't find all the answers you seek here, but you'll find the connections and signposts that will guide your exploration of this complex and socially charged matter.

We've also asked a few people who have a range of experience and expertise in testing and measurement to give us a few brief comments on potential uses for standardized tests. We believe there are some, but we also believe they're not the uses most people are familiar with. We'll include their views in a series of sidebars interspersed throughout the text.

We also do not intend to refight old academic battles. The academic fight over the use of a single test score to determine eligibility for a program or

benefit or to impose sanctions or to retain a student in a grade for an additional year has long been concluded. The test-'em-till-they-improve team lost. The three main professional associations that deal with psychometrics—the American Psychological Association, the American Educational Research Association, and the National Council on Measurement in Education—issued a revised set of standards for the use of testing nearly a decade ago.⁵ (A panel to revise the standards was appointed in September 2008, but its work is not likely to be done in time for us to make use of it.)

You need to know up front that the standards adopted by these three associations are frequently violated by all manner of programs and policies, some federal, some state, some local. The 1999 version of the standards organized the field of assessment into the broad areas of test construction, fairness, and application. Most of the violations we refer to have taken place in the area of application, where judgments are frequently made about individual children on the basis of tests designed to assess group performance, where children are routinely denied high school diplomas as a result of a test score, where access to special programs for high achievers depends on a test score. If you believe that the leading professional associations and the leading psychometricians and analysts know what they're talking about—and we do believe them—then you'll agree that the way we use tests in public schools needs to change and come into line with the best thinking in psychometrics.

Those who continue to advocate for the misguided uses of standardized tests tend to be policy makers, corporate leaders, and politicians, rather than educators. But the important point is not who they are but that their arguments are not based on sound evidence. They are, in fact, based on equal measures of rhetorical sleight of hand, wishful thinking, and a widespread public faith in the almost magical inerrancy of numbers.

In addressing the underlying assumptions of standardized tests, we'll do more than introduce you to the conclusions of the nation's leading psychometricians. We'll take a few shots of our own. And while we marshal the evidence of the experts in quotations and notes, we will draw on our own understanding of human beings and human societies to address the widespread cultural appeal of the "pro-testers." We acknowledge the attractiveness of simple answers to complex questions, but, like H. L. Mencken, we believe they are highly likely to be wrong.



Just about everyone—politicians, school leaders, teachers, parents, and indeed almost all citizens—has grown up with these tests. They became a part of the background of life in the last half of the twentieth century, and their

importance has only intensified in the twenty-first. Standardized testing is to Americans as water is to a fish. We don't even see that things could be different. We hope this book will make it clear just how wet we all are.



A MEDLEY OF VIEWS

A great many people have given the matter of standardized tests and their use a great deal of thought. We asked some of them for brief responses to this question: *Please describe what, in your view, are the appropriate uses of standardized tests in U.S. schools.* In this introduction and at the end of chapters 1, 2, 6, 7, 8, 9, and 10, we present A Medley of Views that we received. Some commentators said that, when certain guidelines are observed, standardized tests can be quite useful; some saw no possibilities whatever for the tests. But no one we asked endorsed the current use of the tests for accountability purposes under No Child Left Behind or the new administration's proposed uses in its "blueprint." We didn't ask Secretary of Education Arne Duncan or his predecessor Margaret Spellings.



"READING" THE READING TESTS BY SUSAN OHANIAN

Early in my teaching career, I thought standardized tests had some minimal use. I thought they let me know if my students, in general, were able to do what experts expected. But the more knowledgeable I became about child development and the closer I looked at the tests, the more I began to suspect the so-called expertise of the test producers. So I tried skewing the results. For example, guessing that test makers would consider apostrophe placement a big deal for third-graders and guessing that they would try to lure kids with sins of commission, before the test I told my students, "You know, we don't use apostrophes in this class, and I'll break the knuckles of anybody who says the right answer is to add an apostrophe."

Predictably, the test had two such items, and my class, grouped together as the worst readers in third grade, obeyed my threat and scored above grade level in Language Arts Usage. I think my method is saner

than drilling hapless third-graders all year on the difference between possession and plurality.

They also scored high in spelling even though they were abominable spellers. On standardized tests spelling is really proofreading, and since my students did a *lot* of reading, they were adept at recognizing what “looked” right. Later, it was hard to convince parents that their children really weren’t fine spellers when McGraw-Hill claimed they were “above grade level.”

Now, as I collect absurd standardized test questions from across the country, my misgivings grow. In testing comprehension, some test makers use vile passages constructed by work-for-hire temps. Others use “authentic” literature, degrading that literature in the process. When I saw the mutilation done to D. B. Johnson’s *Henry Hikes to Fitchburg*, I burst into tears. Surely, D. B. Johnson did not create this work so kids will identify an adjective when they see it. The MCAS (Massachusetts Comprehensive Assessment System) asked tenth-graders to read a passage from *The Grapes of Wrath* by John Steinbeck and then answer this question:

The sentence “From her position as healer, her hands had grown sure and cool and quiet; and faultless in judgment as a goddess” begins with

- (a) a split infinitive.
- (b) an independent clause.
- (c) a prepositional phrase.
- (d) a gerund phrase.

As though Nobel Laureate writers write to provide children with grammar lessons. Or as if this information has anything to do with reading comprehension.

Test manufacturers would never be able to keep the lid on just how outrageous their reading comprehension tests are if looking at these tests weren’t a felony in many states. Teachers and parents and the public aren’t allowed to have a clue.

For anyone who administers standardized tests to children or whose children take standardized tests, *Children and Reading Tests* (Hill and Larsen, JAI Press, 2000) is a must-read. Must. No matter what your level of expertise in deciphering reading tests, this book will knock your socks off. Using methods of discourse analysis, the authors examine representative material from actual reading tests, and they discuss

children's responses. In short, they talk to children about why they chose the answers they did. In a sophisticated and nuanced revelation we see both how tests fail to tap into children's worldviews and how convincing children's "wrong" answers are.

Susan Ohanian is a teacher and writer who lives in Vermont. She hosts the website www.susanohanian.org, which was founded in opposition to No Child Left Behind.