

The War on Standardized Testing

Kill the Messenger

With a foreword by
Herbert J. Walberg

and a preface by
J. E. Stone

Richard P. Phelps

An Executive Summary Published by the Association of American Publishers



Educational testing is an essential activity in every school, in every school district and in every state. Standardized tests are used to evaluate students and schools; to help improve teaching and learning; and to generate important information from which educational policy decisions can be made. With the passage of the No Child Left Behind Act of 2001, which requires tests to be the primary measure of school accountability, testing has taken on added significance.

Yet, when the results of these tests are not as good as one would like, there has been an increasing tendency by some to blame the test. In other words, if you don't like the results it maybe easier to kill the messenger than fix the underlying problem the test revealed.

The Association of American Publishers' Test committee created this brochure, which is an executive summary of a book written by Richard Phelps entitled *Kill the Messenger*, to help policy makers and the public better understand the growing debate about the use of standardized tests in our nation's schools.

– The AAP Test Committee

This document excerpts from *Kill the Messenger: The War Against Standardized Testing*, by Richard P. Phelps, published by Transaction Publishers. Copyright © 2003 by Transaction Publishers, Rutgers University, 35 Berrue Circle, Piscataway, NJ 08854 ISBN 0-7658-0178-7 / cloth / 331 pp. / 2003 \$39.95.

Information about quantity discounts of *Kill the Messenger* is available from: Special Orders Department, Transaction Publishers Distribution, 390 Campus Drive, Somerset, NJ 08854 www.transactionpub.com


Kill the Messenger: The War on Standardized Testing

As a judge in a popular television program asserted to public school officials gathered in his courtroom for a trial, “We give you our children.” We also give the public school system lots of our money. Given the scale and the stakes, one would think that U.S. public education professionals would be highly accountable to the public and deeply imbued with an ethic of public service. Unfortunately, the opposite seems too often the case.

Though a novice to the debate over standardized testing can easily become confused by the flurry of arguments and counterarguments over details, some of them silly and arcane, the key, essential point of debate is *who* gets to measure school performance—the education “professionals” or those of us who are footing the bills and giving up our children. The essential point of debate is whether testing, and other methods of quality control, should be done “internally” or “externally.”

Two Sides

The more one looks at those education policy issues in our society that stimulate controversy, the clearer it becomes that there are two very distinct interest groups involved, groups whose interests are often diametrically opposed. There are education providers—education professors, teachers’ unions, and a proliferation of education administrator groups, all with large memberships and nationwide organization. With some notable, brave exceptions, these enormously wealthy



Most of the arguments used against standardized testing could be used just as validly against weather reporting, or any one of thousands of other ordinary, familiar systems of measurement.

and politically well-connected groups, oppose “external” testing. They favor only that standardized testing which they, themselves, fully control from initial development or purchase through the scoring, analysis, reporting, and interpretation of results.

On the other side of education debates are education *consumers*—the general public, parents, students, and employers—who, in overwhelming proportion, strongly support standardized testing and, in particular, standardized testing with consequences (i.e., “stakes”). With the exception of employers, however, this group is neither well-organized nor well-financed. Though education consumers outnumber education providers in vast numbers, they do not have the money, organization, time, or political power to compete with the provider organizations in a fair fight.

Though there are two sides to the debate on external standardized testing, then, often only one side gets an opportunity to talk. One can visit several dozen websites of professional organizations of education professors, research centers, administrator associations, and, with one exception (the American Federation

of Teachers), those of teacher groups, and find gobs of “research studies” on standardized testing. A quick glance will show, however, that, with only rare exceptions, the selection of research is completely one-sided. Professional educators, those to whom we “give our children” get to hear only one side of the story, as their leaders allow only one side of the story to be told.

A couple of years ago, I contacted the webmasters at a couple dozen professional educator organization websites, suggesting that their listings of research reports were one-sided, and I provided links to research reports that would provide some balance and perspective to the issue of standardized testing. With only one exception out of a couple of dozen organizations, I was given the cold shoulder.

And, the Superior Alternative to Testing is ... What?

Tests provide information. Usually, they do it more reliably, more accurately, more objectively, and less expensively than the alternatives. Why be opposed to information? The reason usually is: because one does not want the results known. Students who work hard want the evidence of their work to be known. Some students who do not, do not. Those who want to improve the schools want information that will guide them toward improvement. Those who do not want to improve the schools, do not.

Opposing the use of standardized testing is like opposing weather research. Weather reports sometimes predict bad weather and then, sometimes, we get bad weather. So, one could suppose,

we can fix the problem of bad weather by banning reports that predict bad weather. Moreover, weather reports are not completely reliable. In fact, they are sometimes wrong! So, perhaps we should ban weather reports because they sometimes give us information we do not like to hear and because they are not perfect.

As *Kill the Messenger: The War on Standardized Testing* demonstrates, most of the arguments used against standardized testing could be used just as validly against weather reporting, or any one of thousands of other ordinary, familiar systems of measurement. While observing the war on testing, the healthy skeptic should keep in mind two questions—if we were to do this, who would benefit? ...and, are the alternatives better? Who would benefit if we were to ban weather reporting, and are the alternatives better?

Critics do not usually compare standardized testing to available alternatives but, rather, to an unattainable or nonexistent utopian perfection. Neither standardized tests nor the manner in which they are administered will ever be perfect. But, the alternatives to standardized testing are far from perfect, too.

The most prominent available alternative to standardized testing is a system of social promotion, with many levels of nominally the same subject matter taught in high school, ranging from classes for self-motivated kids to those for kids who quit trying years before, and whom the system has ignored ever since. It is a system that graduates functional illiterates. Another available alternative is the large-scale institution of remedial

programs in colleges, to compensate for the degradation of instruction in the elementary and secondary schools.

Another available alternative is an absence of reliable information of student performance anywhere outside a student's own school district. Given the arbitrary and idiosyncratic character of high school graduation credentials in the United States, few employers have a clue about the skills their job applicants based on student transcripts alone (that is, where it is even possible to obtain transcripts).

Still another available alternative to high-stakes standardized testing is a school culture in which students are routinely passed whether or not they earn it, and the few teachers brave enough to assign failing grades may well have their marks erased and replaced by school administrators, thus allowing failing students to graduate and controversy to be avoided. In high schools where a student can graduate despite doing no work, other students, and their parents, will assume that they, too, can pressure school administrators for selfish advantage. Behind-the-scenes prerogatives become implicit academic standards.

Yet another available alternative is an exclusive reliance on teacher grading and testing. Are teacher evaluations free of the problems listed in the anti-testing canon? Of course not. For example, grades are susceptible to inflation with ordinary teachers, as students get to know a teacher better and learn his idiosyncrasies and how to manipulate his opinion. That is, students can hike their grades for reasons unrelated to academic achievement by gaming a teacher's system.



Tests provide information. Usually, they do it more reliably, more accurately, more objectively, and less expensively than the alternatives. Why be opposed to information?

Standardized Tests Versus the Actual Alternatives

Judging standardized tests against a benchmark of utopian perfection that does and cannot exist means standardized tests always look bad. But, the available alternatives to standardized testing, it turns out, contain the same drawbacks that testing opponents ascribe, rightly or wrongly, to standardized testing. How would testing critics’ most common accusations look compared to the actual, available alternatives? The table below provides a glimpse.

There are a number of problems with teacher evaluations, according to numerous researchers. Teachers tend to consider “nearly everything” when assigning marks, including student class participation, perceived effort, progress over the period of the course, and comportment. Actual achievement in mastering the subject matter is just one among many factors. Indeed, many teachers express a clear preference for non-cognitive outcomes such as “group interaction, effort, and participation” as more important than test scores. It’s not so much what you know, it’s how you act. Being enthusiastic and group-oriented

gets you into the audience for TV game shows and, apparently, also gets you better grades in school.

The actual alternative to standardized testing is an absence of standardized curriculum and instructional practice (because they would be unenforceable). If the curriculum is never evaluated, we cannot know if any of it works. Without standardized tests, moreover, no one outside the classroom can reliably gauge student progress. No district or state superintendent. No governor. No taxpayer. No parent. No student. Each has to accept only whatever each teacher says and, without standardized tests, no teacher has any point of comparison, either.

Standardized Testing in Practice

Each of us faces many tests in life. Only a small portion of those tests are given in school. Of those school tests, only a tiny portion are “standardized tests.” Of those school tests that are standardized, only a minority are “external” standardized tests with “high stakes”—predetermined consequences for the students who take them based on the results.

But, oh boy, do those *external* high-stakes standardized tests cause a ruckus! What makes these tests so different from other school-based tests and so controversial? Education providers may not control the process nor the effect or the interpretation of the results.

Someone else does. That “someone else” may include elected state officials, college admissions committees, citizen committees that write academic standards, local parents, and general public opinion. For U.S. education professionals

Common Standards and Tests	Teachers/Schools do as they Wish
Teaching to the Test	
<p>Testing opponents repeatedly assert that teachers will teach the material that will be on a test. But, if the tests are high-stakes, and they are kept under lock and key until the day of the test administration, how can teachers know what material will be on the test, except in the most general terms?</p>	<p>Teachers always “teach to” something. If they are not teaching a required curriculum matched to a standardized test, what are they teaching? ...and why should we be so sure that it is better than the required curriculum? In the absence of common standards and tests, teachers still teach to something, but that something is arbitrary; it is whatever the teacher personally prefers. And, why should the taxpayers pay for that?</p>
Narrowing the Curriculum	
<p>Standards prescribe a common curriculum. It is a curriculum, however, based on public decisions, and not on an individual teacher’s personal preferences. As there are a fixed number of hours in a day and days in a year, it is not possible for a common curriculum to have less content than a teacher-arbitrary curriculum. If, when standards are introduced, a teacher drops one topic, she also adds another.</p>	<p>What teachers and schools do in the classroom absent any adherence to common standards is not necessarily any “broader” than what happens with common standards. Indeed, it is likely to be “narrower” as it is determined by nothing more than one individual’s personal preferences. Nor is it necessarily any better, more profound, or more beneficial to the students. It’s merely more arbitrary.</p>
Cheating by Students	
<p>Cheating is far easier to prevent and detect with standardized tests. Different forms can be used in the same classroom making copying unrewarding. Computer programs can be run after the fact that look for telltale patterns.</p>	<p>Cheating in regular classroom work has become epidemic. The overwhelming majority of students admit to cheating in polls. Teachers and schools are ill-equipped to monitor or detect most cheating. Meanwhile, the Internet makes cheating far easier than it used to be.</p>
Cheating by Teachers	
<p>The fact that cheating by teachers on high-stakes standardized tests is well publicized is testament to how easily that cheating can be detected.</p>	<p>Testing opponents argue that teachers have an incentive to cheat on high-stakes tests and no incentive to cheat otherwise. Nonsense. Social promotion and grade inflation provide the contrary evidence. In surveys, the majority of teachers claim overwhelming pressure to give high grades to and promote undeserving students.</p>
Preferred Instructional Methods	
<p>Testing critics claim that what happens in the classroom absent standardized testing is wonderful (rich, innovative curriculum; the joy and magic of learning, and so on).</p>	<p>But, they provide no evidence of any of their claims. We are left with no choice but to take their word on it.</p>
Opposition to Norm-Referenced Tests	
<p>Testing critics excoriate the use of norm-referenced standardized tests as unfair (i.e., it is unfair to simply rank kids, rather than measure them against standards).</p>	<p>But, they would have us rely on grade-point-averages, which are norm-referenced measures, normed at the school level</p>
Preference for Teacher-Made Classroom Testing	
<p>Standardized tests are developed by cadres of tests and measurement Ph.D.s. The most capable measurement experts in the world work in North America developing standardized tests.</p>	<p>Few teachers have had more than a cursory exposure to training in measurement. But, standardized testing critics would have us rely exclusively on their measurement decisions.</p>

who, unlike their counterparts in most other countries, have gotten used to pretty much running all of the education enterprise as they see fit (or, as their critics charge, as best suits their own purposes)—and evaluating their own performance—this outside “interference” comes as quite a shock.

Until recently, the United States had a testing system that matched the preferences of school officials. In most places, standardized tests were used only for “diagnostic” purposes. Local school officials purchased the “nationally normed” tests directly from test publishers, administered them on their own and in their own way, and then sent the completed test forms back to the test publishers for scoring. Those scores provided school officials with some rough comparison of their own students’ academic progress to national averages.

Commercial test publishers liked this system well enough. They made many separate sales of, essentially, the same product to many separate customers. There was no extra work involved in “aligning” the tests to state standards or curriculum; the tests were sold right “off-the-shelf.” Since local school officials could use the tests (and the resulting scores) any way they wanted to use them, they were generally happy with the system, too.

Then, in the 1980s, a West Virginia physician named John J. Cannell investigated a statistical anomaly that he had discovered: statewide average scores for students on some widely used standardized tests were above the national average

in every state in which they were given. It was dubbed the “Lake Wobegon Effect” after the fictional public radio community where “all the children are above average.”

Reform of Test Use

The Lake Wobegon controversy led to calls for more state government control over test content and administration and less local discretion. In most states, those calls were answered. Probably the single most important recent innovation to improve the quality and fairness of testing in the United States is the addition of managerial and technical expertise in state education agencies. At that level, it is possible to retain an adequate group of technically-proficient testing experts, adept at screening, evaluating, administering, and interpreting tests, who are not “controlled by commercial publishers” or naive about test results. They, along with governors and legislatures, are currently calling the shots in standardized testing. Some of the most important decisions that affect the design and content of standardized tests, and the character of the testing industry and the nature of its work, are today being made by state testing directors.

State testing directors, for example, can utilize a number of relatively simple solutions to the problems of test-score inflation (and other alleged problems with standardized tests, such as curricular narrowing and teaching to the test), including: do not reveal the contents of tests beforehand; do not use the same test twice; include items on the test that sample very broadly from the whole domain

of the curriculum tested; require that non-tested subjects get taught (or test them, too); and maintain strict controls against cheating during test administrations. Most states now either use tests custom built to their state standards and curricula or that are adapted for that purpose from commercial publishers' huge test item banks. Indeed, all of the apparent causes of the Lake Wobegon Effect are fairly easily avoided.

The Lake Wobegon Effect overwhelmingly involved the use of “no-stakes” tests purchased by local education officials, whose results and interpretation were entirely under the control of those same local educators. Some testing opponents, however, assert that, despite the overwhelming *prima facie* evidence to the contrary, the Lake Wobegon Effect is, really, more applicable to the use of high-stakes tests. This is like arguing that the police initiating new patrols in a high crime neighborhood are the *cause* of the crime that occurred in the neighborhood before they showed up.

Testing opponents attribute the Lake Wobegon Effect to *high stakes* testing despite the fact that high-stakes tests tend to have extremely high profiles, with their operations and results shone under a bright media glare and difficult to hide. They argue that education officials have more *motive* to cheat when the stakes are high for students. But, the motive to cheat is present whenever test scores are to be made public, and the *opportunity* to cheat is far greater when stakes are low and no one is watching. The critics' claims are most ironic given

Most other countries are increasing their use of standardized testing and high stakes. Most citizens of the work could not imagine an education system—or one that made any sense—without them.




that the Lake Wobegon Effect was a product of the system to which they advocate we return, one where local school officials maintained complete control over all information related to their own schools' performance.

Testing Outside the United States

Most foreign observers would likely be confused or bemused by the controversy over external, high-stakes standardized testing in the United States. Or, they might even be peeved, because most media coverage and anti-testing research in the United States implicitly assumes that no other country on earth has any experience with testing or is worth talking to about it.

To the contrary, the majority of the world's industrialized nations have administered large-scale high-stakes tests for decades. U.S. testing opponents caution incessantly that we should not develop high-stakes tests here until we know what will happen if we do this or that. More research is called for, presumably from those researchers who can be relied upon to produce negative results. It is rarely suggested that one need only make a few overseas telephone calls to learn what happens.



Testing's popularity remains steadfast today, despite an onslaught of one-sided media coverage, and despite the efforts of testing opponents (and naïve or sympathetic journalists) to drum-up "backlash."

Moreover, most other countries are increasing their use of standardized testing and high stakes. Most citizens of the world could not imagine an education system—or one that made any sense—without them. The dire "sky is falling" consequences of high-stakes testing predicted by opponents here should have befallen most of the globe by now, if they were accurate.

If anything, the United States, in its relative paucity of testing, exit, and entrance requirements, is the odd country out. Most standardized tests in other countries are fully high stakes—students who do not pass the examinations are denied a diploma, entry to the next level of education, or entry to the school or university of their choice. Moreover, they are generally timed tests, given only once a year, with few chances to retake, and cover a wide range of subjects. Finally, they can be grueling, lasting days, or even weeks. The average student in a dozen European countries surveyed in 1991 endured well over 40 hours of high-stakes exams during her elementary-secondary school career, and over 60 hours in all kinds of standardized testing.

Testing's Enduring Popularity

The U.S. public has often been asked how it feels about testing. The margins in favor have typically been huge, on the order of 70-point spreads between the percentage in favor of more testing and the percentage against.

Testing's popularity remains steadfast today, despite an onslaught of one-sided media coverage, and despite the efforts of testing opponents (and naïve or sympathetic journalists) to drum up a "backlash." Testing opponents know it will not work to reveal their self-interest in opposing testing; their opposition has usually been couched in claims of defending defenseless students, whom they claim are harmed by tests. Given that, testing's popularity has always been inconvenient. Nor does it work to claim that the public is ignorant about tests, as virtually every adult has completed at least ten years of education and has encountered many tests, for better or for worse.

So, some opposition groups have conducted their own opinion polls, with crudely worded, transparently-biased questions and atrociously low response rates. After the completion of one of these polls, the group's president wrote: "Any farmer knows that you do not fatten cattle simply by weighing them every day." Moreover, he added, "Standardized tests cannot reliably measure...honesty and perseverance." On that point, he's right. They cannot. Nor are they supposed to. Nor should regular classroom teachers try to; that's not their job. Standardized tests, however, do a terrific



job of measuring academic knowledge and skills, which is their purpose.

Even with that, however, he disagreed. He proclaimed tests to be: “unreliable, unrealistic, and unfair;” “imprecise;” and “invalid for...high-stakes decisions such as promotion or graduation.” They: “measure factoids” “distort” and “narrow” curriculum and “cannot reliably measure problem solving.” Their effect on instruction is to “narrow” it; turn teachers into “drill sergeants;” “taint the atmosphere;” and promote “memorization and repetition.”

This is rhetoric, of course. This advocate of honesty and perseverance shouts a plethora of bad-sounding words at standardized tests in hopes of overwhelming the reader’s disposition to consider another side of the story.

Research ... Lots and Lots of “Research”

But, the group’s poll is still “research” after all and, like much research conducted by testing’s opponents, whatever its quality, it serves two purposes: It gives testing opponents “evidence” to cite to support their position (while they ignore or deny the existence of higher quality, contradictory evidence); and it helps to obfuscate the issue, which is particularly helpful in dealing with journalists.


Testing opponents can also consult a vast store of in-house research expertise that is largely funded by us—taxpayers—at a charge of about a billion dollars a year. These tax dollars pay the salaries of education school faculty and for the work of dozens of federally funded education

If anything, the United States, in its relative paucity of testing, exit, and entrance requirements, is the odd country out. Most standardized tests in other countries are fully high stakes—students who do not pass the examinations are denied a diploma, entry to the next level of education, or entry to the school or university of their choice.

research centers, staffed largely by education school professors and graduates.

Given their occupational position and their self-selection to that occupation within the current industry structure, one should not be too surprised that many education professors might feel an inherent dislike of objective, external testing. The commonly accepted notion that all university professors are objective or that, at least in the aggregate, the mechanisms for open debate in academia work to achieve objectivity and truth, should be put into some context. The public probably can feel a high degree of confidence in the objectivity of the collective opinion of, say, astronomy professors in regard to their current views of the cosmos. In terms of the cosmos, there is no collective self-interest in shaping the debate.

But, what of astronomy professors’ opinions regarding government funding for astronomy research or requirements that university students take astronomy courses? These issues affect astronomers personally and, depending on how these issues are resolved, some of careers could be



The importance of standardized testing cannot be overstated. Without the data provided by fair, reliable, and valid tests we—students, parents, teachers, and policy makers—would lack ability to determine whether our schools and children are making the grade.

ruined. Even if there were an astronomer who genuinely felt that too much public money was spent on astronomy research, how willing would he be to speak up publicly, against the wishes of his colleagues, upon whom he depends for his own career advancement?

Ownership of the Means of Instruction

For education school professors, much, if not most, of the research they conduct has implications for the structure of the education enterprise in our country and, in turn, for their own livelihoods and those of their colleagues. How is high-stakes standardized testing a threat to the profession? It is capable of exposing flaws in the current system, thereby opening the system to criticism and inducing calls for change. Moreover, testing can publicly expose a pervasive practice of which most citizens remain blissfully unaware, but that could potentially outrage many of them: for many education professors and the education professionals they train,

academic achievement is only one of many, and far from the most important, goals they have in educating *our* children. From that perspective, standardized testing that focuses “only” on academic achievement can seem quite narrow, unfair, and distracting.

Probably largely for these reasons of self-interest, much research conducted by education professors on the topic arrives at emphatically anti-testing conclusions, with “external” and “high stakes” testing, in particular, attracting a cornucopia of invective. Much, if not most, of this research is then passed on to the public by journalists as if it were neutral, objective, unbiased research, like that which we might expect from astronomy professors in the topic of the origins of the cosmos.

It has not escaped the notice of some, however, that most of the alleged negative consequences of external standardized testing are entirely under educators’ own control. *Educators* tell us that high-stakes testing will induce *educators* to cheat. *Educators* tell us these tests will induce *educators* to “teach to the test” and “narrow the curriculum” (and that these are bad practices). *Educators* tell us that *educators* will dumb down the curriculum, practice rote instruction, and write meaningless test score reports for public consumption.

Testing opponents suggest that educators simply cannot be trusted to behave responsibly or ethically if the public insists on employing quality control in the public schools. Yet, we are to be expected to trust them without reservation if we leave them

free to spend our tax dollars and treat our children as they wish.

In Closing

The importance of standardized testing cannot be overstated. Without the data provided by fair, reliable, and valid tests *we*—students, parents, teachers, and policy makers—would lack ability

to determine whether our schools and children are making the grade. There is no dispute that standardized tests provide information in a reliable, accurate, objective and cost-effective manner. They are the best measures we have regarding student performance, and it's clear that they are messengers worth listening to—not killing.

This document excerpts from *Kill the Messenger: The War Against Standardized Testing*, by Richard P. Phelps, published by Transaction Publishers. Copyright © 2003 by Transaction Publishers, Rutgers University, 35 Berrue Circle, Piscataway, NJ 08854 ISBN 0-7658-0178-7 / cloth / 331 pp. / 2003 \$39.95.

Information about quantity discounts of *Kill the Messenger* is available from: Special Orders Department, Transaction Publishers Distribution, 390 Campus Drive, Somerset, NJ 08854 www.transactionpub.com

About the Author: Richard P. Phelps has devoted most of his working life to the world of education. His research work on testing has been published in *Evaluation Review*, *The Industrial Organizational Psychologist*, and *Educational and Psychological Measurement* among others.



ASSOCIATION OF AMERICAN PUBLISHERS, INC.

Association of American Publishers

50 F St., N.W., Suite 400

Washington, DC 20001

www.publishers.org