

STOPPING THE MISMEASUREMENT OF EDUCATIONAL QUALITY

W. James Popham
University of California, Los Angeles

America's educators are being obliged to participate in a contest they cannot win. The content is called *the score-boosting game* and its object is to have educators raise students' scores on standardized achievement tests. The score-boosting game rests on the assumption that a set of high scores by students equals more successful instruction by educators. That assumption, however, rests on a clear misapplication of educational measurement.

Unfortunately, many educators' current efforts to succeed in this accountability-induced contest are causing serious educational harm to the nation's students. Therefore, today's evaluation of schools on the basis of students' standardized test scores must be stopped. In this article, I'll suggest how this brand of educational mismeasurement might be reduced.

Let me be as candid as I can be. I am not writing this article because the topic is "interesting" or because I need to beef up my publication list. I am writing it specifically because *The School Administrator* reaches the nation's educational leadership. I want that leadership to do something. More accurately, I want *you* to do something. I want you to help halt the significant educational mischief now going on in our schools because of an ignorance-inspired approach to the evaluation of schooling.

Briefly, here's what you'll find in the following paragraphs. First, I'll indicate why standardized achievement tests do not yield an accurate picture of a school staff's instructional effectiveness. Second, I'll describe a few of the harmful educational practices that have been spurred by educators' sometimes mindless pursuit of higher test scores. Finally, I'll set forth a series of six specific action-options that the nation's educational leaders could follow to diminish or expunge this increasingly widespread misuse of standardized achievement tests. Along the way, I'll be trying to persuade *you* to implement one or more of these action-options.

In simple language, bad things are currently happening to children in many of our schools because of an unsound measurement-grounded accountability strategy. The strategy was installed by educational policymakers who, albeit well-intentioned, mistakenly believed that students' scores on a standardized achievement test tell us how well those students have been taught. But I do not fault the architects of these well-intentioned but misguided testing programs. Those people were attempting to help children on the basis of flawed knowledge about educational assessment. Rather, *I fault us*. That is, I fault the education professionals (myself included) who sat on their hands while the score-

* This essay appeared in the December 2000 issue of *The School Administrator*, Vol. 57, No. 11, pp. 12-15.

boosting game was born and multiplied. Our apathy in allowing this mismeasurement of educational quality to flourish has led to a situation that puts both teachers and administrators in an untenable position. Even worse are its adverse educational consequences for our students.

Measurement Misuses

A standardized achievement test is any examination that is administered and scored in a standard, predetermined manner. The most commonly encountered standardized achievement tests are the five nationally distributed ones, namely, the *California Achievement Tests*, *Comprehensive Tests of Basic Skills*, *Iowa Tests of Basic Skills*, *Metropolitan Achievement Tests*, and *Stanford Achievement Tests*. But, in a number of states, we also see standardized achievement tests that have been built to mesh better with a state's curricular preferences.

Today's standardized achievement tests are modeled after the *Army Alpha*, built in 1918 to help identify suitable officer candidates for our World War I forces. The *Alpha* was a group-administrable intelligence test measuring an Army recruit's *aptitude* for success in an officer training program. The *Alpha's* comparative assessment strategy allowed recruits to be compared with a norm group (of previous Alpha test-takers) in order to isolate each recruit's relative aptitude.

Today's standardized *achievement* tests fundamentally follow the *Alpha* assessment strategy. And, therefore, today's standardized achievement tests must produce a sufficient degree of *score-spread* so that fine-grained comparisons among students can be made. I have no quarrel with standardized achievement tests if they're used properly. Both educators and parents can be helped by learning a child scored at the 86th percentile in mathematics but only at the 30th percentile in language arts. That's a good use of such tests. But standardized achievement tests do *not* tell us how well students have been taught. Here are three reasons why that's so.

First, because the makers of these tests must market their exams on the basis of a one-size-fits-all curriculum, in many localities what is taught does not match what is tested. Some studies suggest that more than fifty percent of the content on many standardized achievement tests is not taught, or not even supposed to be taught in a particular school district.

Second, because of the tests' relentless quest for score-spread, test-makers dare not include many items that fail to discriminate among examinees. Items that many students answer correctly, because such items do not contribute to score-spread, tend to be excised from these oft-revised tests. Yet, items on which students score well often turn out to deal with content that, because of its importance, teachers stressed. Thus, items covering the most important things

that teachers teach will tend to be jettisoned from standardized achievement tests.

Finally, if you spend any *serious* time reviewing the actual items on standardized achievement tests, you'll find that there are at least three distinguishable sorts of items. Some items measure the kinds of things that are taught in school. (Although, as suggested above, in a given school setting there may be mismatches between what's tested and what's taught.) A second kind of item consists of those linked to students' socioeconomic status (SES). These items will be answered correctly more often by students from higher-SES families than by students from lower-SES families. Finally, some items will be answered correctly more frequently by students who were fortunate enough to inherit higher levels of academic aptitudes, that is, verbal, quantitative, and spatial capacities. The SES-linked and aptitude-linked items do a wonderful job in producing the kind of score-spread so essential if standardized achievement tests are to fulfill their *Alpha*-like assessment mission. Yet, as a consequence of the inclusion of many such items, (and there are *many* such items), to a large extent standardized achievement tests measure what students bring to school, not what they learn there.

I hope you can see, then, that students' scores on nationally standardized achievement tests do not accurately represent how well a school's staff has taught. Unfortunately, the foregoing three problems also exist for many state-customized achievement tests which, not surprisingly, are often built by the same *Alpha*-oriented companies that create and sell the five nationally standardized achievement tests. If you're a superintendent in a district serving mostly high-SES kids, your test scores will almost always be high. If you're working with low-SES kids, however, your district is destined to get low scores. And that's not your fault, or the fault of your staff. It is the necessary consequence of using the wrong test to evaluate instructional quality.

Harming Children

Given the requirement to use the wrong tests, and the immense pressure to boost students' scores, a good many educators have begun to employ practices that clearly are not in children's best interests. Important curricular content is often eliminated because it is not addressed by whatever high-stakes test is used locally. Children are subjected to relentless drilling on the content apt to be measured by the high-stakes test. Children are given interminably long practice sessions with items mirroring those on the test or, worse, with actual items surreptitiously copied from the high-stakes test.

The message sent to children in such settings is clear: "It's not learning that's important; it's higher test scores. And we are out to get them any way we can!" Every week, or so it seems, we read a handful of horror stories about teachers or administrators who have been caught cheating on high-stakes tests.

They cheat (and I'm not excusing it) because they simply can't play the score-boosting game and win. We cannot allow this absurd score-boosting game to continue. It is a contest in which the chief indicator of success is flat-out wrong!

A Menu of Action-Options

What I'm suggesting, of course, is that if you are a school leader who allows the score-boosting game to continue as is, you really cannot succeed. If you serve low-SES children, you'll either be on the receiving end of poor student test-scores or you'll be obliged to endorse (or overlook) instructional practices that harm children. If you happen to be serving high-SES children, then you run the risk of allowing your district's *undeserved* success to mask instructional practices that need improving. Moreover, if the parents in a high-SES district ever discover that the true reason your district's test scores are so high is simply that you drew an advantaged set of students, you're in trouble. Accordingly, I urge you to consider the following six action-options, described here only briefly. Several follow-up resources are cited at the article's close for those who wish to consider these options more carefully.

- *Option 1: Provide an intense, extensive assessment-literacy program for teachers and administrators.* If educators do not understand why certain high-stakes tests yield invalid estimates of instructional quality, they cannot inform parents or policymakers of these problems. Moreover, assessment-illiterate educators cannot describe more appropriate evidence by which to evaluate instructional quality. Today, an assessment-illiterate staff of educators represents a clear liability for any educational leader.
- *Option 2: Offer carefully structured briefing sessions to educational policymakers regarding appropriate/inappropriate ways to evaluate schooling.* Concise explanatory sessions for board members, legislators can be planned to explain why some kinds of tests yield faulty pictures of instructional quality. The use of actual items from currently used tests is especially helpful in getting policymakers to understand why certain kinds of high-stakes tests provide invalid estimates of educational quality.
- *Option 3: Implement a meaningful assessment literacy program for parents.* As soon as a staff-focused assessment literacy program has been concluded, provide an outreach program tailored to the interests of parents. Parents will be supporters of what's good for children if they truly understand the key measurement concepts involved.
- *Option 4: Foster establishment of autonomous parent-action groups.* Unfortunately, if educators protest the misuse of high-stakes testing programs, they will typically be regarded as self-serving, defensive,

and thoroughly unbelievable. However, if *nonpartisan* parent groups protest poorly conceived high-stakes testing programs, the views of those parents will be given serious consideration by policymakers. I have great confidence in *autonomous* parent groups if the members of those groups are assessment-literate.

- *Option 5: Conduct security-monitored reviews of the items in the high-stakes test(s) being used.* There are enormous insights to be gained by having educators and noneducators carefully analyze, one item at a time, the actual items in national or state standardized achievement tests. The protocols for such item-reviews must be carefully designed, of course, but the results of such rigorous item reviews can be remarkably illuminating.
- *Option 6: Devise and implement more valid, credible evaluative schemes.* There is nothing wrong with accountability-oriented programs to evaluate schooling if the appropriate kinds of evaluative evidence are incorporated in the evaluations. To reject evaluative programs based on the wrong data without replacing those evaluative programs with defensible data would clearly be both professionally and politically improper. (I described an example of one kind of credible evidence that can be collected to reflect instructional quality in the previous month's issue of *The School Administrator*.)

One of the nice things about a menu of options is that one need not choose only one item. Personally, when I find myself in a great cafeteria, my tray is soon laden with far more than one choice. (I find the use of *two* trays to be far less constraining.) If you really want to do something serious to stop the mismeasurement of educational quality, I encourage you to implement more than one option. Indeed, why not go for all six?

Fomenting Time

To make up for our heretofore apathetic acceptance of the wrong sorts of high-stakes tests to measure educational quality, I suggest that the nation's educational leadership set out sans delay to foment a meaningful revolution. The erroneous, and educationally harmful, appraisal of instructional quality via standardized achievement tests must be stopped. And *you* can help stop it. Or don't you care?

W. James Popham is an emeritus professor in the UCLA Graduate School of Education and Information Studies. E-mail: wpopham@ucla.edu.