

USES AND MISUSES OF STANDARDIZED ACHIEVEMENT TESTS*

W. James Popham
University of California, Los Angeles

The *Law of the Hammer* reflects a well-known truth that if you give a hammer to a child, the child will soon identify an enormous number of things in need of hammering. American educators most likely believe that there is an analogous *Law of the Standardized Achievement Test* because such tests are currently being used for an almost unlimited number of purposes. If an important educational decision needs to be made, and test data are considered relevant, then it's more than likely someone will try to rely on the results of standardized achievement tests. But tests, as is true with hammers, can sometimes hit the wrong targets.

Focusing on Five Tests

In the following analysis, I want to consider the appropriate as well as the inappropriate uses of standardized achievement tests. I refer specifically to those educational tests that, designed to assess students' skills and knowledge in particular subject fields, are to be administered and interpreted in a standard, predetermined manner. I realize that there are nationally standardized achievement tests as well as state-specific standardized achievement tests being used by educators these days. Those state-level tests, however, are not only linked to a particular state's curriculum, they can vary substantially in the way they are constructed.

Accordingly, I'm going to focus only on the five nationally standardized tests now widely used in the nation's elementary schools, namely, the *California Achievement Tests* and *Comprehensive Tests of Basic Skills* published by CTB/McGraw-Hill, the *Iowa Tests of Basic Skills* published by Riverside, and the *Metropolitan Achievement Tests* and *Stanford Achievement Tests* published by Harcourt Educational Measurement.

As you see, in the U.S. there are only three publishers who create and distribute the five nationally standardized achievement tests. Because I don't want to be accused of sniping at any one of these three test-publishing firms, I'll simply refer to them as the "Big Three" whenever I cite something they've written about the proper or improper use of their own tests. I suppose, given one's attitudes toward educational tests, the three publishers might be thought of either as the "blessed trinity" or the "terrible trio." But I'll simply use "Big Three" and, when presenting something that one of them has presented in print, I will simply cite it as emanating from a Big Three Member. (If someone is truly desperate to get specific citations, I have such information available.)

I will be dealing only with *achievement* tests, not *aptitude* tests. Aptitude tests, of course, are intended to predict a test-taker's likely success in a future situation, typically

* This essay appeared in the February 2001 issue of the *NASSP Bulletin*, Vol. 85, No. 622, pp. 24-31.

a future academic setting such as when the SAT or ACT are employed to predict how successfully high school students will perform after they get to college. The proper and improper uses of aptitude tests is another important, but different topic.

All right, having identified the tests under consideration and the way I'm going to cite some of the test publishers' words, let's turn to the appropriate and inappropriate uses of standardized achievement tests. We can start with ways that standardized achievement tests can be used properly.

Appropriate Uses of Standardized Achievement Tests

There are four major uses of standardized achievement tests I regard as hunky-dory. Taken in concert, these four uses thoroughly justify the existence of standardized achievement tests. I am *not* opposed to such tests. I think they should exist, and even flourish. I just don't think they should be misused. But I'll deal with such misuses in a moment. Here, then, are four uses of standardized achievement tests I think are altogether appropriate.

Informing parents about their children's relative achievements. One important use of nationally standardized achievement tests is to give parents an idea about how their children stack up, in various subjects, against the performance of a national comparison group. The *norm-referenced* interpretations yielded by such tests can be quite useful to Marty's mom and dad who discover that Marty scores at the 89th percentile in reading but only at the 34th percentile in mathematics.

From their earliest beginnings, standardized achievement tests have been conceptualized and constructed so they provide accurate and fine-grained discriminations among test-takers. The discriminating efficiency of such tests provides parents with a meaningful fix on how their youngster compares with a national norm-group in responding to a test's items in, say, language arts or social studies. Such comparative interpretations "enable home and school to work together in the student's best interests."^{*}

Informing teachers about their students' relative achievements. In much the same way that a parent can gain insights about a child's relative standing in different subject areas, so too can teachers profit from identifying their own students' comparative strengths and weaknesses. A fifth-grade teacher who discovers that Lee scores super in science but miserably in math can begin to think how to *add* to Lee's math insights and, thereby, *subtract* from Lee's math weaknesses.

I believe that this use of students' relative strengths and weaknesses ought to be limited either to the subject areas themselves (such as science or reading) or to two or three fairly large chunks of content in a particular subject area. I don't think more fine-

* Big Three Member, *Educator's Guide*, current edition.

grained breakdowns, e.g., per sub-skill breakdowns, should be used—for reasons I'll spell out later.

Selecting students for special programs. Because the chief mission of standardized achievement tests is to provide norm-referenced comparisons among students, such tests are ideally suited for reaching decisions in *fixed-quota* settings where there are more applicants than openings for a special program. The special program might be an enrichment activity for gifted children or a remedial activity for low-performing children. Because standardized achievement tests can accurately identify who performs *best* or *worst* in responding to a set of items, these tests are quite appropriate whenever choices must be made from an excessively large applicant pool.

Allocating supplemental resources. A fourth appropriate role for standardized achievement test arises when state or district educational policymakers are trying to decide how to distribute discretionary resources such as special funds for additional staff-development support or, perhaps, supplemental monies to support after-school tutorial sessions for low-performing students.

It seems to me that, other things being equal, it makes more sense to assign such supplemental resources to the schools (or in the case of a state to the districts) where students' performances on standardized achievement tests indicates that serious instructional action must be taken. I am not suggesting that the schools whose students perform poorly on standardized achievement tests are *instructionally ineffective*. On the contrary, the students in those schools may be receiving first rate instruction, but come to school from homes that fail to supply the academic support so often reflected in high test scores. But, whatever the cause (and it surely *might* be poor teaching), the allocation of supplementary resources to low-performing schools seems more sensible than spreading finite supplemental resources so thinly that neither high-performing nor low-performing schools benefit.

Four defensible uses. In review, then, I suggest that the following four uses of nationally standardized achievement tests are wholly appropriate:

Appropriate Uses of Standardized Achievement Tests

1. Informing parents about their children's relative achievements
2. Informing teachers about their students' relative achievements
3. Selecting students for special programs
4. Allocating supplemental resources

If used in one or more of these four ways, standardized achievement tests will be educationally useful assessment tools. I'm sure that there are other suitable uses for standardized achievement tests, but the four I've identified here are, in my view, the most important uses to which such tests should be put. Let's turn, now, to the flip-side of standardized achievement test usage.

Inappropriate Uses of Standardized Achievement Tests

To keep this analysis even-handed, at least for appearance's sake, I will now identify four inappropriate uses of standardized achievement tests to match the four appropriate uses just treated. To be candid, however, because I think some of the following four misuses of standardized tests are particularly serious, I tend to become more worked up about them—as you will see.

Evaluating schools. Because so many people, including most educators, believe that schools should be judged chiefly on the basis of students' scores on standardized achievement tests, I need to tackle this misuse first. I certainly concur that a school staff's effectiveness should be dominantly evaluated on the basis of students' measured achievements. But those achievements should not be measured by standardized achievement tests. Yet, we find that the publishers of standardized achievement tests, although dispensing politically required rhetoric about the need for other evidence of school quality, continue to tout their tests as meaningful contributors to the evaluation of educational effectiveness:

. . . a standardized achievement test can provide valuable information about the progress of individuals and groups *and the effectiveness of educational programs* (emphasis added).*

Yet, educators need to remember what the primary purpose of standardized achievement tests really is, namely, to detect sufficient differences among test-takers so that sensitive norm-referenced comparisons can be made. To pull off that mission, the developers of such tests sometimes include items apt to be answered correctly by students who either (1) come from advantaged socioeconomic backgrounds or (2) were fortunate enough to inherit above-average academic aptitudes such as verbal or quantitative capacities. Test items based chiefly on these two factors almost always produce the desired spread of student scores. That's because both socioeconomic status and inherited academic aptitudes reflect what children bring to school, not what they learn there.

In addition, the publishers of standardized achievement tests must, for marketing purposes, base their tests on the content representing a common-denominator “national curriculum.” Thus, test publishers create a one-size-fits-all test covering the content most apt to be emphasized throughout the nation. Yet there's a high likelihood the curricular preferences of educators in a given locale will not mesh all that well with the content of a standardized achievement test.

A school's educators ought to collect test-based evidence that students are making meaningful achievement gains, but other sorts of tests should be used for this

* Big Three Member, *Teacher's Guide*, current edition.

important purpose. Standardized achievement tests should not be employed to evaluate schools, and every educator needs to understand why.*

Evaluating teachers. If standardized achievement tests ought not be used to evaluate a *group* of teachers in a school, it is surely unsound to employ such tests to appraise an individual teacher. Anyone who's had anything to do with schools realizes that the caliber of a given teacher's students can vary dramatically from year to year. To evaluate teachers based on year-to-year changes in their students' performances on a standardized achievement test is downright laughable—given the potential for shifting ability-levels of a teacher's students. This year's group of whiz kids may be replaced next year by a collection of chowderheads. Yet, some naïve educational policymakers are advocating precisely this sort of teacher evaluation model, that is, a simple subtraction of last year's students' average test scores from this year's students' average test scores. Simple it is; smart it isn't.

Fortunately, the publishers of standardized achievement tests recognize this misuse of their wares, and some publishers even go on record to discourage it. After pointing out a series of difficulties in judging teachers with standardized achievement tests, one publisher notes "it is strongly recommended that student test results *not* be used as a criterion for evaluating student performance."^{**}

Promoting or grading students. A third inappropriate use of standardized achievement tests arises when teachers use results of such tests to *grade* students or when test scores are used to *promote* or *retain* students. Test publishers point out that standardized achievement tests *sample* the knowledge and skills at any grade level, and "are not intended to be 'end of course' tests."^{***} As one test-publisher points out, "if a retention decision is to be made, classroom assessment data gathered by the teacher over a period of months is likely to form a highly relevant and accurate basis for making such a decision."^{***}

For the same reasons, especially the probable mismatch between a standardized achievement test's "national" content and a particular teacher's instructionally emphasized content, use of scores from such tests for *grading* purposes is unwise. Again, given the timing of a standardized achievement test's administration and the delay in the return of students' scores, use of these tests for grading purposes is fairly rare. But it's still wrong.

* As might be guessed, I have railed more than a few times about this particular misuse of standardized achievement tests. Educators will find a more extended treatment of the issue in Popham, W. James, *Modern Educational Measurement: Practical Guidelines for Educational Leaders*, (3rd ed.). Boston: Allyn & Bacon, 2000. Parents will find the misuse of tests treated in Popham, W. James, *Testing! Testing! What Every Parent Should Know About Schools Tests*. Boston: Allyn & Bacon, 2000.

** Big Three Member, *Planning Guide*, current edition.

*** Big Three Member, *Planning Guide*, current edition.

**** Big Three Member, *Educators' Guide*, current edition.

Making classroom instructional decisions. A fourth and final misuse of standardized achievement tests arises when teachers really try to employ the tests or students' test scores in connection with day-to-day classroom decisions. One publisher argues that a test's content should *not* be used "to decide which instructional objectives should be taught at a given grade level" because the test's questions constitute only a small sample of the potential questions that might be asked.*

Yet, a page earlier in its printed materials, the very same publisher extols the instructional dividends obtainable from one of the publisher's standardized achievement tests. Here, for example, are some of the specific purposes the publisher contends that the tests were designed to serve:

1. To help teachers determine the extent to which individual students in their classes have the knowledge and skills needed to deal successfully with the academic aspects of an instructional program the teacher has planned;
2. To estimate the general developmental level of students so that teachers may adapt materials and instructional procedures to meet individual needs;
3. To identify each student's areas of greatest and least development to use in planning individual instructional goals and approaches;
4. To establish a baseline of achievement information so that the monitoring of year-to-year developmental changes may begin.*

But the chief obstacle to using standardized achievement tests for purposes of teachers' instructional decision-making is insuperable. Put simply, these tests do not provide teachers with sufficiently clear descriptions of the knowledge and skills represented by the test's items. If the descriptions of the test-represented knowledge and skills aren't sufficiently clear for teachers to focus their instruction in an attempt to promote students' mastery of such knowledge and skills, then the teacher's instruction is almost certain to be off-target.

Presented below is a description of one reading objective, along with its subskills, from a widely used standardized achievement test. It's really all a teacher receives in the way of descriptive information regarding what this objective really means:

Evaluate and Extend Meaning

Subskills: generalize; fact/opinion; author—purpose/point of view/tone/bias; predict/hypothesize; extend/apply meaning; critical assessment.**

*Big Three Member, *Educators' Guide*, current edition.

** Big Three Member, *Planning Guide*, current edition.

Now I want you to imagine that you're a teacher who has the responsibility to boost test scores that are, at least in part, based on this objective and its subskills. Don't react generally. Think *specifically* about what you'd need to incorporate in a lesson plan for your class. I'm sure you'll see that, although the subskills (listed above) give teachers a somewhat better idea of what's being tested than the unadorned objective, there is still an enormous amount of slack in the description of the test-

represented content for this objective. Let's say you decided to zero in on one of the subskills, for instance, "generalize." Now what does this subskill really mean? How can teachers plan lessons if they don't understand what they're supposed to be teaching toward? Is it any wonder that some teachers, dismayed by such assessment obfuscation, end up by teaching toward a test's actual items?

And anyone who tells you that classroom teachers can get an accurate idea about their students' progress in objective-mastery or, worse, in subskills-mastery, just hasn't been counting the items-per-objective or items-per-subskill in most standardized achievement tests. One publisher calculates subskill scores using as few as three items per subskill. One can only hope all three items are rather good. They're being asked to carry out a serious psychometric mission.

So, despite reams of promotional rhetoric from publishers about the instructional dividends derivative from their standardized achievement tests, I just don't see any. I regard the use of such tests to make classroom instructional decisions as clearly inappropriate. Are standardized achievement tests more helpful to a teacher's instructional decision-making than no tests at all? Of course. But, I fear, not all that much more helpful.

All right, then, I've trotted out the following four misuses of standardized achievement tests:

Inappropriate Uses of Standardized Achievement Tests

1. Evaluating schools
2. Evaluating teachers
3. Promoting or grading students
4. Making classroom instructional decisions

Only one of these misuses, *evaluating schools*, is so widespread that I think emergency action is required to head it off. I think that the fourth misuse, *making classroom instructional decisions*, is pretty misguided, but because most teachers can't figure out how to get meaningful instructional insights from standardized achievement tests or from students' scores on those tests, it is a fairly feckless misuse. So, my major quarrel is with the first misuse I described, that is, the evaluation of schools.

Medicine for Measurement Misusage

With all the incredible advances being made these days in the pharmaceutical field, it's now possible to get one or more medications for a great many ailments that, in earlier years, were essentially untreatable. But what sort of medicine should we use if we want to counter the inappropriate uses of standardized achievement tests and to increase the appropriate uses of these assessment devices? My answer is quite simple. We need to bolster every relevant constituency's *assessment literacy*. I refer not only to educators—heaven knows they need it—but also to educational policymakers, media representatives, and garden variety citizens.

Today's educational tests are far too important to be misused by those who are supposed to employ them properly. Today's educational tests are far too important to allow such misuses to continue. Although a solid dose of assessment literacy, all by itself, won't work educational miracles. It's a sensible place to start the treatment.