Customized Criterion–Referenced Tests

W. James Popham

Who should develop criterion-referenced tests—local districts, commercial publishers, or organizations that produce customized tests on a contractual basis?

An interesting yet troubling situation is developing in the field of educational measurement with respect to the creation of criterion-referenced tests. Increasingly, educators are becoming dissatisfied with traditional standardized achievement tests as indicators of educational quality. Because of substantial deficiencies in such norm-referenced measures, many educational evaluators are espousing the benefits of criterion-referenced tests.

Norm-referenced tests are designed to determine an examinee's status in relationship to that of a normative group ("Johnny scored at the 76th percentile"), whereas criterion-referenced tests attempt to identify whether an examinee can display a clearly defined class of behaviors. A criterion-referenced test might be used to assess a pupil's mastery of a clearly defined competency in a subject field, for example, reading, mathematics, or history. The more lucidly a criterion-referenced test defines the class of examinee behaviors it measures, the more accurately it can be used by evaluators and instructional designers.

Criterion-referenced tests, in part because of the greater likelihood that they can be selected so as to coincide with a local curriculum, offer school people a chance to "look good" for a change. After years of consistent and frustrating failure with norm-referenced tests, it is no wonder that many educators are currently clutching at criterion-referenced straws.

Commercial Test Development

But where, we might ask, will decent criterion-referenced tests come from? Will they be mysteriously born, as are most of their norm-referenced cousins, in prestigious places such as Princeton, New Jersey, or Monterey, California? Will, in other words, commercial test publishers be able to satisfy the growing appetite of American educators for high-quality criterion-referenced assessment devices?

My answer to this important question, after several years of pondering the issue, is—NO. I am convinced that commercial publishers of criterion-referenced tests face an insoluble problem arising from two facts:

Fact One: If a criterion-referenced test is going to be worth its salt (or pepper), it must *unequivocally* describe the domain(s) of examinee behavior it assesses.

Fact Two: Except, perhaps, in reading and mathematics, the curricular emphases in U.S. school systems are so diverse as to render only a small proportion of their curricula appropriately assessable by a *single* criterion-referenced test.

Couple these two facts and, if you're a test publisher, prepare for instant insomnia. If you do a decent job of creating a criterion-referenced achievement test, a test with explicit behavior

and content specifications, you reduce your potential sales market dramatically. Many potential test purchasers will, having considered the test, decline to buy it because it fails to coincide with local instructional emphases. If you fuzz up your criterion-referenced test descriptors so that the tests will be more attractive to many would-be purchasers, the tests lose the advantages that come with clarified descriptions. 'Tis a situation where there are many ways to lose, and no ways to win.

Local Test Development

Because commercial testing agencies will be unable to supply off-the-shelf criterion-referenced tests that both satisfy local curricular preferences and simultaneously incorporate clarified performance descriptors, it seems to me that local districts will have to arrange for local test development.

But local test development poses some serious problems that must be overcome. In the first place, well constructed criterion-referenced tests cannot be whomped up by tyros. Merely labeling a test "criterion-referenced" does not assure us that the test possesses the qualities it needs. Some of the worst criterion-referenced tests ever created have been produced in the past few years by well-meaning local school people who were raw amateurs at test development. Nor will a "quickie" workshop in the "fundamentals" of criterion-referenced test construction automatically bestow expertise on novice test-developers. Moreover, local school people who have been hired to be teachers or administrators often do not possess those qualities (largely masochistic) necessary to be effective generators of criterion-referenced tests.

Contractual Test Development

Faced with this dilemma, my guess is that local school people will turn to testing organizations which, on a contractual basis, can create customized criterion-referenced measures suitable for a particular school system. Although the local school staff will determine in general terms what should be tested and will need to monitor the in-progress work of the test development agency, the bulk of the actual test construction will be carried out by the contractor.

If individual school systems adopt this strategy for test acquisition, several problems must be recognized. The most substantial, of course, is the cost factor. Producing a first-rate criterion-referenced test, that is, one with a clear descriptive scheme and a sufficiently large pool of items congruent with that descriptor, is costly business. Could local districts afford such developmental expenses?

Well, there are a few ways to keep these kinds of costs at a manageable figure. First, if state departments of education can be persuaded to underwrite the creation of criterion-referenced tests to measure more fundamental and widely accepted competencies in reading and mathematics, local districts can share in the use of such tests. Second, as commercial testing outfits become more heavily involved in this type of customized test-development work, they will surely be able to offer local educators at least some choices among already created criterion-referenced tests, thereby reducing the developmental costs. Finally, by limiting its measurement focus to a modest number of highly significant competencies, rather than trying to tap every pupil skill worth promoting, the district may be able to afford the creation of criterion-referenced tests that really do the measurement job it wants done.

Whether the customized, contractual development of criterion-referenced tests will become a common practice in the next few years remains to be seen. The major deterrent would seem to be the financial requirements involved. Yet, astute local educators will surely recognize the enormous costs they are currently paying as a result of public disapproval based on the wrong kinds of test performance. A citizenry that believes its schools are doing an effective job may be more apt to finance those demonstrably successful schools at an appropriate level. ☆



W. James Popham is Professor of Education, University of California, Los Angeles.

Copyright © 1977 by the Association for Supervision and Curriculum Development. All rights reserved.