

# 2

## Extensions

### *Three Primary Purposes of Educational Testing*

#### Chapter 2's Assessment-Related Understanding

*Purposeful Educational Testing:* The construction and evaluation of educational tests should be profoundly influenced by one of the three primary purposes of such testing, namely, (1) comparisons among test-takers, (2) improvement of ongoing instruction and learning, or (3) evaluation of instruction.

#### BETTER UNDERSTANDING AN UNDERSTANDING

The most important thing about this chapter's understanding is its effort to split educational testing's primary purposes into three mutually exclusive categories. If this division of educational assessment holds up, then all of us will have an easier time when thinking about the dominant measurement mission of a particular educational test. Please note that

two key assessment operations should be “profoundly influenced” by this three-way division among educational testing’s primary purposes. Those two operations are the *construction* and *evaluation* of educational tests. That is, the way we build educational tests and—after they’ve been built—the way we measure how good they are. Given that the building and using of educational tests pretty much captures what’s significant about educational testing, then this chapter’s assessment-related understanding should be regarded as a significant one. A test’s chief purpose, you see, should influence almost everything else in the educational assessment arena.

Chapter 2 discusses three chief uses of these tests, namely, (1) comparisons among test-takers, (2) improvement of ongoing instruction and learning, and (3) evaluation of instruction. There may be lesser purposes of educational assessments, such as satisfying some sort of governmental requirement to administer certain kinds of tests or, perhaps, satisfying the curiosity of a skeptical school-board member. And the virtues of these sorts of purposes should not be trivialized. They can sometimes be quite important. Yet, they are not a test’s *primary* purpose. There is only one *primary* (most significant) purpose of an educational test.

Those who wish to become truly knowledgeable about educational testing must possess a crisp grasp of the tripartite distinction among purposes described in Chapter 2. One determination in the crispness of one’s grasp of these three primary purposes rests on an important distinction between two measurement notions, an *assessment purpose* and a *results application*.

Let’s consider one way of distinguishing between *assessment purpose* and *results application*. An educational test whose primary purpose is to compare test-takers will, if the test is successful, provide scores that allow the performances of test-takers to be contrasted with one another. For instance, regardless of the kind of scoring system being used, one test-taker’s score can be (1) the same as the scores of certain test-takers, (2) better than the scores of certain test-takers, or (3) worse than the scores of certain test-takers.

One of the most common ways of reporting student's performances on an educational test is to indicate how many score-points have been earned by a student as a "raw score" and then translate this raw score into a *percentile*. A percentile indicates what percentage of the test-takers were outscored by a particular student. For instance, a student whose raw score was equal to the 89th percentile, would have performed better than 89 percent of the other test-takers. If an educational test whose primary purpose is to make comparisons among test-takers produces a flock of different scores so such comparisons can be accurately made, then this test has fulfilled its primary purpose.

But how do we translate raw scores into *applications* leading to actual decisions? For example, we might rely on a set of percentiles to decide which students should be given awards for high-level mastery of whatever the test was measuring. Or, in contrast, we might use students' contrastable raw scores to identify a group of students who are then eligible for intensified remedial assistance because of their relatively low performances on a comparison-focused educational test.

Putting it differently, the test's *purpose* is to provide a set of contrastable performances by test-takers. This purpose is unarguably comparative in its intended use. However, the *applications* of the resultant comparative scores might vary substantially. All of those applications, however, hinge on this educational test's having accomplished its primary purpose, to provide performance-comparisons among test-takers.

Another primary purpose of educational testing is the improvement of ongoing instruction and learning. Let's see how a contrast between purpose and application plays out. If an educational test's primary purpose is to supply information that's usable in the improvement of instruction and learning, then such information must be provided in a consumer-usable fashion. Otherwise, if those who want to rely on a test's results can't actually employ those results, what good are such results in accomplishing a test's instructional-improvement purpose? For instance, let's say that a test has been constructed in such a way that the results it yields are at a "grain-size" that's far too broad to be of any practical

value—either to teachers who wish to improve their ongoing instruction or to students who wish to improve their ongoing learning. Because the test’s primary purpose has not been accomplished, then its most important application, in this example the enhancement of instruction and learning, is destined to flop. Had the test’s purpose been better satisfied—had more usable grain-size reporting of results been produced, then there would have been a more positive application result.

As a final illustration, let’s consider the third primary purpose of educational testing, the evaluation of instruction. One application of tests with such a primary purpose is to identify truly terrific instructional programs so that those programs can be replicated elsewhere. Another obvious application of the results of an evaluatively focused test is to help root out ineffective instruction so it can be replaced with more effective instruction. However, a number of other applications of an educational test’s results are surely possible. None of these applications of a test’s results, however, are apt to be successful unless it has been demonstrated—with persuasive evidence—that the test itself is capable of distinguishing between well-taught and poorly-taught students. Putting it differently, any successful application of this third primary purpose of educational testing requires the provision of sufficient evidence that the test itself is *instructionally sensitive*.

Wrapping up this look at Chapter 2’s assessment-related understanding from a somewhat different perspective, then, it becomes clear that those who would rely on the chapter’s three-way division of educational testing’s primary purposes need to keep distinct the overriding *purpose* of a test and the resultant *application* of the results yielded by the test.

## COLLEGIAL CONJECTURING

Please consider the contents of the brief, charmingly boxed e-mail presented below. It was supposedly sent to you by a close colleague. If you were in a mood to reply to this make-believe buddy, what would your own e-mail say?

Remember, if you are carrying out this activity as part of a group, then different participants' responses can be collaboratively examined by the group's members. If, however, you are sauntering through this *Online Supplement* solo, you can still find it useful to consider the *strengths* and *weaknesses* of your e-mail response to this colleague.

**TO: THE READER OF A BOOK ABOUT EDUCATIONAL TESTING FROM: A MAKE-BELIEVE CLOSE COLLEAGUE**  
**SUBJECT: WHEN IS TEST-USAGE NOT TEST-USAGE?**

Hi:

I know, from what you told me last weekend at Lester's party, you've been recently reading up a storm about educational testing. Why you would want to do such a thing escapes me. However, we should not let such silliness go to waste! Accordingly, I need your help.

A guy who works with me, Clyde, portrays himself as an expert—on just about everything. I think he is quite confused about one topic related to educational assessment, but I am not certain about how to handle him. Maybe you can draw on your recent reading to give me some ideas.

Clyde claims that, as he puts it, "A test is a test is a test." What he means is that if an educational test is *properly* constructed and *carefully* evaluated, then it can be effectively employed for a whole host of assessment purposes. For example, if a national standardized achievement test has been built by a major educational testing firm, odds are that this test can be properly used to support teachers' instruction, help school boards evaluate a school's instructional success, and compare the current achievement levels of students.

Clyde holds the view that well-made educational tests can successfully carry out a variety of missions. Do you think what he says makes sense? Is Clyde right or wrong? I'll be interacting with him later this week, and I'd really benefit from your thinking on this issue.

Thanks,  
Lee

## THOUGHT-PROVOCATION QUERIES

Look over one or more of the following three queries related to this chapter's assessment-related understanding. Having done so, if a query provokes even a smidgeon of thought on your part, then please consider how you might answer the query (or queries) you chose.

**Query 1.** Although we have experienced roughly a full century of large-scale, standardized testing in this nation, in recent decades the significance of the purposes to which large-scale educational tests have been used has become much more "societally significant." Which of the three primary purposes of educational testing treated in Chapter 2 do you regard as the test-usage most contributory to our current "high-stakes" testing? Why?

**Query 2.** If you were trying to explain to a friend why it is that the 2014 revision of the *Standards for Educational and Psychological Testing* (by AERA, APA, and NCME) have become so influential on the way today's educational tests are both built and appraised, what sort of explanation would you provide?

**Query 3.** It is argued in Chapter 2 that one of the most effective ways to isolate the primary purpose of an educational test is to identify the decision(s) riding on the test's results. Do you agree with this contention? Why or why not?

## A REAL-WORLD APPLICATION

This particular Extension activity is less "real-world" than it is "pretend-world." However, as a group activity it can sharpen participants' conversance with a critical distinction between a test's *purpose* and the *application* of the test's results. As a group activity, it is intended to be used in settings where a collection of individuals is collaboratively probing the content of

*The ABCs.* As indicated earlier, a reader who has no structured interactions with others about the book can still benefit from mentally isolating the chief factors invoked during the sub-group exercise.

### **PURPOSE OR APPLICATION: WHICH ONE?**

#### **(A SUB-GROUP EXERCISE)**

*For this exercise, it is necessary to have at least two sub-groups, although additional sub-groups can be easily accommodated. The essence of the exercise is for members of each sub-group to caucus individually (and quietly), then work up descriptions of an educational assessment situation that clearly represents either a test's primary purpose or an application of a test's results. Ideally, each description will exemplify a purpose or an application, but will not be blindingly obvious to others. In other words, each sub-group might cause its colleagues to think harder about the content of this exercise by making their descriptions accurate but not embarrassingly blatant. Usually, 15–20 minutes is allocated for this activity.*

*To illustrate, suppose a group of a dozen parents is split into three sub-groups of four parents each. Sitting apart from one another, and keeping voices down, each sub-group then constructs three descriptions of either (1) a test's primary purpose or (2) an application of a test's results. The three descriptions can be all purpose-focused, all application-focused, or mixed. These descriptions should be accurate exemplars, therefore, of either test-purpose or results-application.*

*Each sub-group should generate its three descriptions so that they can subsequently be read aloud to the other sub-groups. Each description should be designated—in writing—by the originating sub-group as either a purpose or an application. (Think of this in-writing requirement as an obligatory answer key whose existence will tend to preclude belated answer key changes by those who contrived a description.)*

*Each sub-group then reads each of its descriptions, allowing other subgroups to indicate (as individuals or as a total*

(Continued)

(Continued)

*sub-group) whether they believe the description depicts a test's primary purpose or an application of a test's results. The originating sub-group then indicates which of the two options it was trying to illustrate. Disagreements can be briefly discussed.*

*This process is repeated until all sub-groups have had an opportunity to (1) present their descriptions, (2) have other sub-groups react to each description, and (3) discuss any disagreements. A general review of the examples, along with a reconsideration of the fundamental distinction between test-purpose and results-application can conclude this exercise.*