

3

Extensions

Behind Standardized Testing's Cloudy Curtain

Chapter 3's Assessment-Related Understanding

Standardized Test Development: Essentially identical to the procedures used when teacher-made classroom tests are built, the development of standardized assessments relies on particularly careful test-building and more complete—yet plain language explainable—documentation of purpose determination, content selection, and item construction/revision.

BETTER UNDERSTANDING AN UNDERSTANDING

How to build and successfully use an interstellar rocket is a patently complicated enterprise. It's so complex, in fact, that very few people possess the smarts, training, and commitment necessary to pull off such an endeavor. Similarly, many people regard the standardized testing of our nation's students to

be an enterprise that's almost as off-putting as sending rockets into the dark beyond. As a consequence of the belief that standardized educational testing is somehow "beyond" them, too many individuals who should have a vital interest in the results of standardized tests simply avoid learning much about them, particularly how such tests came into existence. Most of today's citizens, who may have a rough ball-park notion of what standardized tests do, are frankly intimidated by the ways such tests were constructed and how they are then used.

Chapter 3 is written with an anti-intimidation mission clearly in mind. Today's standardized educational tests have become too important to leave them only to measurement specialists. Fortunately, as this chapter's assessment-related understanding asserts, the building of standardized educational tests is essentially no different from the way teachers crank out their own classroom tests every few days.

This chapter's understanding assumes that people who regard the creation of standardized educational tests as fundamentally unfathomable will usually be reluctant to look into the merits of such assessment instruments. On the other hand, those folks who recognize that the basics of standardized test-building are no different from the basics of test-building by classroom teachers will often be empowered to demand plain-language explanations of what went on when a specific standardized test was born. Such demands, though often warranted, are rarely registered.

Chapter 3's understanding indicates that three basic operations must always be undertaken when building any sort of educational test, whether the test is a nationally standardized achievement test or Miss Ballard's test of her students' mastery of punctuation. Those three test-development operations are (1) determination of the purpose for which a test is to be used, (2) selection of the knowledge or skills to be measured, and (3) generation of the test's items—and subsequent honing of those items if such honing is needed. This three-stage sequence will always be followed for any educational test worth its salt—or its pepper. It is always possible, of course, to

give short shrift to any of those three operations—but this treatment invariably reduces the quality of the resultant test. (Contrarily, there is no evidence that tall-shrift treatment of educational test appreciably improves them).

In the chapter's featured understanding, you'll find an important phrase that might easily be overlooked, namely, "plain-language explainable." It's a potent phrase because, once we realize that what's going on in the generation of a high-stakes educational test essentially travels the same three-stage trail that teachers follow as they gin up their own classroom tests, we can call for the creators of any significant standardized tests to spell out—in *language comprehensible to normal earthlings*—how each of the three fundamental steps of test-building actually took place.

Oh, sure, the explanations that are supplied of (1) purpose determination, (2) content selection, and (3) item construction/revision might to be dished up in language too technical for many. But an appropriate response from a listener at that point should be similar to a paraphrase of the following request: "Could you please explain that procedure again, but less technically?" If the person who's doing the explaining can't provide a clear and comprehensible, jargon-free explanation of how a test-development procedure took place, then the listener should accurately conclude that the explainer is simply not up to the job. *All* technical procedures associated with the building and sharpening of standardized educational tests can be explained in a way that regular individuals can intuitively comprehend. Moreover, once the nature of a test-development operation is understood, then the quality of that operation can be better judged.

To illustrate, in a July 24, 2016 opinion essay in the *New York Times*, Diane Ravitch, a prominent educational historian, renounced her previous endorsement of the Common Core State Standards, a set of curricular aims nurtured by the federal government. In her essay, Ravitch pointed out that the development of these Standards was funded "almost entirely" by the Bill and Melinda Gates Foundation. She also

characterized the development of them as “a rush job.” Two sets of federally funded nationally standardized tests measuring students’ mastery of the Common Core State Standards have since been developed and are now employed in various parts of the nation. Well, if you think back to the three basic test-building operations set forth in Chapter 2’s test-related understanding, you will see that “content selection” definitely took place when the Common Core’s curricular targets were identified. That particular phase of the Common Core’s creation has never been adequately described to the public, but it certainly could be revealed if sufficient pressure to do so were present. The chapter’s understanding indicates that all three stages of standardized test-building can be described in a transparent, comprehensible manner. All that’s needed is for a sufficiently potent demand calling for such explanations to be present.

Building educational tests rests on common sense, whether such common sense is seen in the generation of teachers’ classroom tests or in the construction of high-stakes standardized tests. Information regarding the quality with which the three chief test-building operations have been carried out can be, and should be, demanded by those who have a stake in the use of standardized educational tests.

COLLEGIAL CONJECTURING

Now, please read the boxed, totally make-believe e-mail from an imaginary friend below. To add a spot of verisimilitude, we’ve called him Philip. Your task is to consider what your friend Philip is concerned about in his communication, then come up with a reasonable reply to him. Your response to Philip, if you are tackling these Chapter Extensions with others, can be compared to the replies of other individuals. If you are using this *Online Supplement* all by yourself, it can still help to think through what kind of reply you might whomp up for your fictitious friend.

**TO: A READER OF THE ABCS FROM:
FERVENTLY FRIENDLY PHILIP SUBJECT:
OPACITY UNDER SCRUTINY**

Howdy:

I am so pleased that you're back from your annual vacation. I hope you really enjoyed the islands. If you did not bring back gobs of photos from your snorkeling sessions, I'll never forgive you. However, I am writing for another reason than an appreciation of underwater photography.

You'll remember I told you that I've been taking an active part in the "Parents Together" group at the school where Fred and Fiona are now in the 3rd and 5th grade. Something came up during this month's meeting, only a week ago, and it has me really perplexed. I know you've been reading that book about educational testing that you told me about when we were together a few weeks ago. I can't recall the title, or even the author, but I think he was named Popoff—or something like that. Anyway, here's my dilemma.

A month ago, the governing Board of our 22-school district announced tentative plans to spend a huge amount of money on the purchase of what they call "standardized interim tests," a battery of three-tests-per-year for use in grades three through six. These tests are supposed to be administered several times during the school year, and the results are supposed to help teachers better mesh their instructional activities with the learning levels of their students. Because of the size of the district's investment (Read that as spending "Flo's and my tax dollars."), the Board is sending a representative from the company that wants to sell these tests to each of the district's schools. This person will make a visit and, if a school requests it, a second visit to each school. We had our first visit last month, and we've asked for a second visit from the same person in two weeks. Here's where you come in.

At the first of our two meetings, a Dr. Jill Havens of the testing company described the tests and how they were supposed to be used by our school's teachers. She did a solid job in laying out the potential uses of the tests, and most of our Parents Together

(Continued)

(Continued)

group understood what she was talking about. However, one of our members, Joe Simpson, asked Dr. Havens to explain how the tests were actually developed, that is, where they came from and how did they get to us? Joe wanted to know "how well the tests were developed." Dr. Havens replied that her company's standardized interim tests had been developed in accord with "best professional practices" and she was confident that they were, as she put it, "first-rate interim tests." At that point, however, I thought she became a bit condescending, and indicated that if we wanted more detailed information about the construction of the tests, she would need to return for a second meeting with us. She indicated, however, that she did not think most of us would understand the technically sophisticated procedures that had been employed to create these interim tests.

My question to you is the following: Can members of our parent group understand the test-development procedures employed with these standardized interim tests or, as Dr. Havens indicates, are those procedures too complicated for us? I really need your take on this issue. If she's correct, then we'll need to leave the "buy/don't buy" decision to others. But if she's wrong, then I am going to join Joe in demanding plain-talk descriptions of how these tests were built.

Thanks, Philip

THOUGHT-PROVOCATION QUERIES

Please take a gander at the following three queries, and select one or more to which you'd like to supply an answer. Then, if you're willing, churn up an answer to any of the questions that captured your fancy.

Query 1. In Chapter 3 of *The ABCs*, standardized tests are defined as follows: "A standardized test is a test that's administered, scored, and interpreted in a consistent, predetermined manner." As you can see, three required factors are identified,

the administration, scoring, and interpretation of a test's results. But what about the original construction of such a test? Do you think that different standardized tests can be built in meaningfully different ways? Why or why not?

Query 2. According to Chapter 3's assessment-related understanding, the description of a standardized educational test's development procedures must be "plain language explainable." However, procedures that are able to be explained in plain language need not always be explained in such a way. Or should they be? Do you think that plain-language descriptions of the procedures employed in the construction of a standardized educational test must, without exception, actually be provided? Or is the chapter's understanding satisfied when a plain-language explanation can be provided if necessary?

Query 3. Do you believe that the parallel drawn in Chapter 3 between the construction of teacher-made classroom tests and the construction of standardized tests—particularly the kinds of tests employed for high-stakes educational decisions—if accurate, really makes any difference? Or, instead, even if fashioned according to an identical test-development strategy, are the two kinds of educational tests so fundamentally different that such comparisons are of little utility?

A REAL-WORLD APPLICATION

The final Extension activity for this chapter revolves around the fact that standardized educational tests are merely more careful and better documented implementations of the same sort of test-development procedures that the world's teachers have been using since the middle ages and much earlier.

Please, in a group or by yourself, tackle the sub-group exercise described below in italics. You'll find that the need to come up with a rationale in support of Chapter 3's understanding will help you internalize it.

ENTICING LEGISLATIVE LEARNING

(A SUB-GROUP EXERCISE)

After dividing a larger group into subgroups of 5–7 members, imagine that each subgroup is charged with the task of testifying to the members of a state legislature’s Select Committee appointed to study “The Role of High-Stakes Testing in Our State.” More specifically, each subgroup is to describe why it is that the Select Committee’s members should understand the basics of the way that high-stakes standardized tests are typically constructed and, after their development is finished, evaluated as to their quality. At least two of the legislators on the Select Committee have previously registered the opinion that the building and appraisal of educational tests—especially standardized tests being used for significant decisions—is a topic “best left to those measurement experts who really understand what’s going on.”

The task of each subgroup is to prepare an oral presentation of about five minutes’ duration to be presented to the Select Committee. First discuss, then choose, what your sub-group regards as the most persuasive reasons for a state legislator to dip into a topic often regarded as too technical for laypeople—or even for educators.

After a preparation period of 10–15 minutes, have one or two members of your sub-group present to the total group your defense of the proposition embodied in this chapter’s assessment-related understanding. Subgroups should take turns presenting their rationale in defense of the chapter’s understanding. A final, full-group discussion regarding the strengths and weaknesses of different sub-groups’ presentations should conclude this exercise.