

7

Extensions

Reporting Results: Where Rubber and Road Meet

Chapter 7's Assessment-Related Understanding

Score-Reporting Rudiments. Because test-based inferences about student groups or individual students are typically formulated from score reports, users of such reports should demand that the scores being reported are both easily interpretable and have been elicited by a test whose purpose coincides with those inferences.

BETTER UNDERSTANDING AN UNDERSTANDING

This chapter's assessment-related understanding addresses only two points. First, a plea is made for the *ease of interpretability* of score reports. Second, those who interpret such reports are urged to make sure that report-based inferences about test-takers reflect *consonance with a test's intended purpose*. Although readers will find a flock of other useful score-reporting

information in Chapter 7, the chapter's assessment-related understanding hinges dominantly on those two points.

Chapter 7 focuses on standardized educational tests rather than the sorts of test teachers churn out for their own classes. Interestingly, most teacher-made tests are readily interpreted because, after all, teachers need to make sense out of the tests they build themselves. In most instances, teachers crank out classroom tests that are altogether consonant with whatever purpose a teacher has in mind. Where we start getting into trouble with respect to this chapter's assessment-related understanding is with standardized tests, especially the sorts of tests whose use contributes to high-stakes decisions about the students taking those tests and/or the educators preparing those students for such testing.

The "trouble" that we start getting into with standardized tests stems directly from the two issues captured in the chapter's assessment-related understanding, namely, (1) the clarity with which a test's results are explained and (2) the degree to which a test's score-based interpretations are in accord with the purpose for which a test was built and evaluated. Putting it simply, a great many standardized educational tests supply score reports that are remarkably difficult to interpret. Similarly, a great many standardized educational tests report their results in a manner that fails to mesh properly with a test's primary purpose. Clearly it is difficult, and sometimes impossible, to make serious sense out of a standardized test's unclear results. Similarly, if those results are being put to a use inconsistent with the measurement mission for which a test was built, then such a test will contribute little to improving the quality of schooling. (And this, remember, is what most people regard as the underlying reason for educational testing's existence.)

With respect to the interpretability of a standardized test's results, this is an issue that has been addressed with vigour by members of the educational-assessment community for more than a quarter of a century. Prominent measurement specialists, such as Ronald Hambleton of the University of Massachusetts,

have led a concerted effort not only to report the results of standardized tests *accurately*, but also *understandably*.

Nevertheless, the difficulties of simultaneously attaining accuracy and understandability in reports of standardized-test results are formidable. As you read in Chapter 7, most standardized educational tests' results these days are reported in the form of "scale scores." For a number of statistical reasons, such scale-score reporting makes statistical sense because it can present equally difficult challenges to different test-takers. However, all by themselves, the scale scores in score reports from standardized tests often make no sense at all. Unless those who craft the score-reporting procedures from standardized tests get much more inventive than many of their score-reporting predecessors, what we end up with for standardized educational tests is a collection of remarkably uninterpretable results.

The second concern embodied in the chapter's assessment-related understanding is the need for persuasive evidence that reported standardized test results are based on a test that is demonstrably *fit for purpose*. Remember, if an educational test has been created, then evaluated, with an instructional-enhancement purpose clearly in mind, but that test's results are being dished up as though they were suitable for the evaluation of schools or teachers, then the test's results are blatantly out of whack with the test's purpose. Users of a standardized test's reported results must be constantly vigilant in making sure that test-results match test-purpose.

COLLEGIAL CONJECTURING

If you're willing, please give a bit of scrutiny to the fictitious e-mail that you'll find below. This is a supposed e-mail from a buddy asking you to comment on a testing-relevant issue he's recently encountered. Because your friend knows you have been voraciously ingesting the measurement contents of a nifty new book about the fundamentals of educational

testing, and because your opinion is apparently valued, please try to respond to your friend either as a formally composed reply or, if you are in a make-believe mood, only mentally.

**TO: MY FRIEND AND RECENT DABBLER IN
EDUCATIONAL TESTING FROM: YOUR AMIGO
SUBJECT: TEST-REPORT INCOMPREHENSIBILITY**

Buenos Dias:

I need your smarts or, putting it more plainly, I need your smarts about how the results of educational tests ought to be reported. I remember your telling me, sometime when we were last together, that you've been reading some sort of "can't put it down" book about educational tests, and something has come up with the parents of our local school that—based on your "newly acquired insights"—you can probably clarify. It has to do with the way that students' test results are reported to parents.

First off, you'll remember that almost a quarter of the kids in Jose's elementary school have Hispanic backgrounds. In fact, some of the families arrived in the district less than two years ago. Fortunately, the staff of Jose's school has made a serious effort to engage these Latino parents in a wide range of activities. Because our family is Hispanic-American, Imelda and I have been asked to take part in some of these activities. It was during one of these meetings last week that the issue bubbled up about which I need your thinking.

In our school district, on three occasions during the school year, students in grades 3–5 are given commercially developed standardized tests that the district refers to as "interim tests." Parents receive reports of their children's performances on these tests within a week after the tests have been taken. The problem is that none of the parents—particularly Hispanic-American parents—have any idea what to do with these test reports. When several of the Latino parents asked during last week's meeting what these standardized tests were for, our school's principal replied that, "They are to be used for instructional improvement at school and at home." After the meeting, as we were walking to our cars,

at least three of the other parents came to me and asked in various ways, "How can we use the results of these tests if we can't make sense out of them?"

And this is where I hope you can help. Can the results of standardized tests be made understandable to parents—even parents whose first language is not English? Or is the reporting of standardized tests' results simply something that parents, especially Latino parents, aren't supposed to understand?

I'll appreciate any insights you can toss my way.

Javier

THOUGHT-PROVOCATION QUERIES

Please consider this set of three questions intended to evoke several dollops of thought from you. If you find yourself enticed by one or more of these queries, please fashion—mentally or in writing—a response. Sharing your responses with other individuals is not formally prohibited by the United Nations' Charter, hence might be a smart thing to do.

Query 1. Why do you suppose that the educational measurement community has had such a difficult time in coming up with score-reporting procedures that can simultaneously provide accurate representations of a test-taker's performance, yet be compatible with the primary purpose for which a test was created?

Query 2. One of the concepts advocated for in the development and use of educational tests is that they cleave to the idea of "universal design." This testing approach requires attention to be given to assessment needs of *all* test-takers from the moment an educational test is first contemplated—all the way until that test is administered and its results interpreted. How, though, should this quest for fairness in testing be portrayed, if at all, when test results are reported? For

instance, if meaningful—but appropriate—accommodations were made in the design and delivery of a test for certain subgroups of students, should those students' scores be accompanied by an explanation describing the nature of such accommodations?

Query 3. Chapter 7's assessment-related understanding calls for users of a test's score-reports, particularly reports from standardized testing, to make certain that a test's results were elicited by a test whose purpose coincides with the kinds of score-based inferences being drawn. In practical terms, if you were a parent of a student taking a high-stakes standardized test, how might you determine the degree to which a test's score-based interpretations coincide with the primary purpose for which the test was built?

A REAL-WORLD APPLICATION

Although this chapter's assessment-related understanding targets only two important insights, Chapter 7 did deal with a number of other test-reporting procedures. These procedures contribute to the likelihood of satisfying the recommendations of the chapter's understanding. For the sub-group activity (4–8 members) presented below, you and your colleagues are asked to generate 3–5 *reporting-related* statements such as the following example: "A score-report based on *percentiles* indicates the proportion of a group of test-takers who have out-performed the student whose percentile is being reported." These should be declarative sentences that are definitively correct or incorrect. (The preceding example, two sentences ago, would be incorrect because it describes the *inverse* of what a percentile actually tells us.)

After your sub-group creates a handful of such reporting-related statements, some accurate and some inaccurate, you should read each statement aloud to the remainder of the group, and then ask the individuals in the larger group to indicate whether each statement is *Right* or *Wrong*. Supply the

correct answer for each statement and, if necessary, defend your group's choice about the rightness or wrongness of the statement. Other groups should do the same.

REPORTING RIDDLES: RIGHT OR WRONG?

(A SUB-GROUP EXERCISE)

After having assembled in sub-groups of 4–8 members, each sub-group should generate a set of three to five statements about reporting of test-takers' results based on the contents of Chapter 7. Some of the statements should be accurate; some should be inaccurate. That is, certain statements should "sound like" accurate chapter-based assertions regarding the reporting of standardized tests' results, yet contradict what's presented in Chapter 7.

After allowing, say, 10–15 minutes for this preparation of "ammunition" for the exercise, sub-groups then take turns presenting their statements, one at a time but read each statement aloud twice, then those in the larger group (other than the statements-originating sub-group) should indicate whether each statement is Right or Wrong. The sub-group presenting the statements then indicates whether the statement was Right or Wrong. Discussion can follow any substantial disagreement about the rightness or wrongness of a given statement.

This exercise is entitled Reporting Riddles: Right or Wrong? The exercise, of course, could have been described as a True-or-False endeavor. However, by so doing, the exercise's name would have abandoned its alliterative allure.