

6

Extensions

Fairness in Testing: A Long Time Coming

Chapter 6's Assessment-Related Understanding

Fairness in Testing. Fairness in educational testing, now seen to be as important as validity and reliability in the construction and evaluation of tests, must be carefully documented—employing both judgmental and empirical procedures—to maximally minimize assessment bias.

BETTER UNDERSTANDING AN UNDERSTANDING

This chapter's assessment-related understanding contains a 16-word phrase, properly punctuated by commas at either of its ends. This 16-worder is much more significant than it seems at first glance. Here then, suitably shorn of its commas, is this potentially impactful phrase that focuses on fairness in testing: "now seen to be as important as validity and reliability in the

construction and evaluation of tests.” And why, you might ask, is this collection of 16 words so very important? This would be a good ask.

To discern why fairness in testing has recently become so important, you’ll really need to check on the two commodities that fairness in testing now matches in importance, namely, *validity* and *reliability*. Validity and reliability have been the hands-down heavy hitters in educational testing for eons, so if fairness in testing is currently occupying a comparably lofty position, then it is a level of significance not to be ignored.

The assignment of significance to the notion of fairness in testing was not made by a small gaggle of educators or a few measurement mavens. No, the significance of fairness in testing flows from the 2014 *Standards for Educational and Psychological Testing*. Moreover, we can safely predict that fairness in testing will become *even more important* as time goes by. Not only was fairness in testing given its own chapter alongside validity and reliability chapters in the 2014 revision of the joint *Standards*, but fairness in testing is now touted as equally important to those more traditional measurement constructs.

Note, too, in Chapter 6’s assessment-related understanding that fairness in assessment is to be accorded great significance in both the *construction* of educational tests as well as in the *evaluation* of those tests. Actually, in the chapter itself we are urged to employ fairness “from the initial building of a test all the way through its evaluation, administration, scoring, and interpretation.” We can reasonably foresee an ever-increasing—across the board—attention to fairness in educational testing.

Finally, the chapter’s assessment-related understanding identifies the two procedures currently employed to enhance fairness in testing, that is, *judgmental* and *empirical* approaches. Clearly, a person who fully grasps the meaning of Chapter 6’s assessment-related understanding must become reasonably knowledgeable regarding how these two strategies attempt to minimize assessment bias. Several of this Chapter’s Extensions deal with those two distinctive but complementary techniques for squeezing unfairness out of educational testing.

COLLEGIAL CONJECTURING

Please consider what is being said below in the imagined e-mail from Chris, a long-time friend of yours. It deals with an assessment issue that often pops up whenever the topic of fairness in educational testing is being considered. After reviewing what your pal has written to you, please decide whether you agree or disagree with the major point made in the e-mail. Then, having landed positively or negatively on what your friend has written, try to generate (mentally or in writing) a response to Chris's electronic communication.

TO: A READER OF THE ABCS
FROM: CHRIS SUBJECT: WHO GOOFED?

Hi:

I've been remiss in not writing to you for the past few weeks, but I've been staggeringly busy at work, and haven't had any time to relax or to contact friends.

I'm writing to get your reaction to a point of view about students' performances on significant educational tests—such as the state-wide “accountability” tests taken a month ago throughout our state. I remember that you were reading a new book about educational testing, and your reading seems to bear directly on an incident that came up just yesterday at a dinner party for the residents of our subdivision. One of my neighbors, Floyd Jones, was complaining bitterly about the quality of tests being dispensed by our state's Department of Education. He indicated that those tests are “flagrantly” biased against certain sorts of youngsters, particularly students of color and students from lower socioeconomic backgrounds. Floyd is the parent of three African-American children, all of whom are enrolled in our local schools. Here's how he backed up his contention that our state's tests are, as he said, “bristling with assessment bias.”

As you may know, our state administers standardized tests in grades 3–10 each spring in mathematics and reading. Students'

(Continued)

(Continued)

performances on those tests are released 10 weeks later for every district and report key demographic strata such as gender, race, and family income (as reflected by whether children receive any state or federal funding for at-school lunch or breakfast). The test results for the last school year were made public about a week ago. At every grade level where the tests were given, black and brown students performed significantly lower than their white classmates. To illustrate, with very few exceptions state-wide, the average percent correct on their grade-level's tests was almost 10 percentage points higher for white students than the percent correct earned by both Hispanic and African-American students. Those achievement gaps garnered the most headlines when our local media reported on the state-test results.

Because these particular performance disparities were seen in almost every corner of the state, Floyd is convinced that *the tests themselves* are biased against children of color. Precisely the same differences between white and minority students seen at the state level were also definitely present in our own school district.

Floyd was telling all those who would listen that the school district's parents need to band together and demand that the district leadership make available descriptions of the procedures taken during the state tests' development to reduce assessment bias. Thereafter, if appropriate, a group of district parents can send state officials—all the way up to the governor—a formal request to replace the state's biased standardized tests.

Based on your reading in that book about testing, what do you think of Floyd's recommendations?

I'll really appreciate any time you can give to this. Floyd is such a straight shooter that I'd like to come up with a sensible response to his concerns. Thanks.

Chris

THOUGHT-PROVOCATION QUERIES

By looking over the following four queries in this section of the Chapter 6 Extensions, you can determine whether any

of them are of sufficient interest to warrant a response—written or only mental—from you. Remember, if you are tackling these activities along with others, comparing your responses with the responses of others can be useful in isolating, then clarifying, the nuances associated with each query.

Query 1. If, for financial or practical reasons, you were obliged to put all of your fairness-promoting efforts behind *either* an empirical or a judgmental bias-detection strategy, which one of those two would you choose? And, having made your decision, please wrestle with the obligatory *why* that underlay your choice.

Query 2. One of the reasons educational measurement specialists attempt to minimize assessment bias is that, when present, assessment bias contributes to “construct-irrelevant variance.” If you were attempting to explain the nature of “construct-irrelevant variance,” what would your explanation be? That is, try to fashion an accurate and understandable description of “construct-irrelevant variance” for a *layperson*—not an assessment-knowledgeable person.

Query 3. If, as Chapter 6 in *The ABCs* asserts, attention to the reduction of assessment bias should be present during the beginnings of test development up to and including the actual administration of an educational test, how can such attention realistically be fostered? And, if fostered, how can such attention be effectively documented?

Query 4. Given the long-standing attention to assessment validation and test reliability, what are some potentially effective ways of engendering more serious attention to the elimination of assessment bias? How can this be done, not only among those who build large-scale educational tests, but also among regular classroom teachers and school administrators?

A REAL-WORLD APPLICATION

Please form small sub-groups, then spend about 15 minutes discussing the nuts and bolts, that is, the detailed particulars, of *one* of the four bias-reduction options presented below. What your sub-group's members should be trying to do is come up with a really sound, flaw-free procedure capable of withstanding the slings and arrows of skeptics. If you are tackling these Chapter Extensions on your own, you might still follow the italicized directions in the exercise, but do so mentally—and just for you.

GETTING SPECIFIC: ONE OF FOUR BIAS-BLASTING PROCEDURES (A SUB-GROUP EXERCISE)

If you are working with a collection of others on this activity, then you should join together with 4–8 participants. What your sub-group is being asked to do is come up with a step-by-step description of one of the following four procedures—each of which calls for the application of either a judgmental or an empirical approach to minimizing assessment bias:

Employing judgmental bias-reduction tactics while constructing an important educational test

Using empirical bias-reduction techniques while constructing a high-stakes educational exam

Utilizing judgmental bias-reduction ploys while evaluating the quality of a significant educational assessment

Relying exclusively on judgmental bias-reduction procedures while evaluating the worthiness of an important educational examination

After choosing one of the above operations, in 10–15 minutes or so devise what your group believes is an essentially unassailable plan to accomplish either test-development or test-evaluation, then present your plan to the remainder of the larger group. Those listening to this presentation should provide constructive feedback regarding the described procedures.

If time permits, the entire group should engage in a discussion regarding the possibility of arriving at a defensible mix of judgmental and empirical procedures when building and appraising important educational tests.