

Forum 3: Norm-referenced test interpretation and evaluation

Item 1

Suppose you have administered a 30-item, rather difficult, test that yields a mean score of 20 with a standard deviation of 5. You decide that a mean of 20 should be considered a passing score of 80. Furthermore, to spread-out the scores, you decide that you want the standard deviation to be 10. In other words you would like to convert the set of *raw* scores on the test to a set of *scale* scores with a mean of 80 and a standard deviation of 10. (This, incidentally, is what DPI does with raw scores on the ABC tests to convert them to scale scores.)

Compute the *scale* scores for the following *raw* scores: 10, 15, 20, 25, and 30.

Item 2

Messick, in his 1989 chapter in the 3^{ed} edition of *Educational Measurement*, building upon Cronbach's 1971 chapter in the 2nd edition of *Educational Measurement*, expanded on the meaning of test *validity*. In presenting what he termed a facited *unified theory of validity*, Messick introduced the notion of *consequential* validity, in which he considered the implications and consequences of test use and interpretation. These are important to contemporary understanding of test validity. In your own words, and as briefly as you can, give a description of consequential validity that you could share with your teachers. (I realize that consequential validity is a complex notion, but try to "boil" it down as best you can. Your goal is to describe it in such a way that teachers will understand why it is important. Consequential validity has important implications for classroom assessment, also.)

Item 3

An argument can be made that a parent (or a teacher, or an administrator) should not be overly concerned about a student's scoring at, say, the 27th percentile rank on a nationally (or state) normed assessment. Can you make this argument? Give it a try.

Item 4

By now you should have at least attempted to compute *coefficient alpha* for the *Motivations for Reading* instrument. What did you find? Would you consider this a reliable instrument?

Item 5

In the Iowa document, four categories of errors that could potentially affect students' scores on an assessment were discussed. The article also discussed three types of assessment interpretation that one could make depending upon the level of reliability. In North Carolina, increasing attention is being focused on the use of EVAAS for evaluating student achievement growth. EVAAS uses scores from previous tests to estimate how a student should perform on a current or future assessment. Considering what you now know about reliability and the levels of reliability needed to interpret test performances in different situations, what can you say about the use of EVAAS for evaluating student performance? This is a wide-open question. I am particularly interested in your opinion.