

MEASURING WITH A “RUBBER RULER”

Randall E. Schumacker, Ph.D.

THE PROBLEM

You have finished your training, test day has arrived, and it is now time to click the “start test button” on your computer screen or report to your designated testing center. Through many hours of study you have carefully and confidently prepared yourself to perform well. Will you pass? Probably.

Yet, your success does not depend solely on your level of knowledge. You will probably be measured with a “rubber ruler” certification exam. Measurements with these exams fluctuate depending on changes in the so called “norming group”. Whether you pass or fail will depend on the person sitting next to you at the testing center, the person who logged-on to the test just ahead of you or most certainly some cohort group who previously took the test.

Does this seem arbitrary and a bit unfair? Well, before you call your association director, your lawyer, the ACLU, or your local congressman, realize that test designers have been struggling with the “rubber ruler” problem for decades, since before most of us were born. Your local psychometrician who lives down the street will be the first one to tell you it is a very difficult issue and no malice on anyone’s part is intended.

THE SOLUTION

The problem is simple to formulate, but difficult to solve. Here it is in a nutshell. The items on your test (and consequently your test as a whole) are considered difficult or easy depending on the knowledge level of the people in the “norming group”. Get the picture? The required pass score will be lower if low ability people were used to establish the pass score, and higher if highly knowledgeable people were used. Simply stated, was the group to whom you are being compared those colleagues who think 100 on an I.Q. test is perfect and 70 is passing, or those who routinely discuss the theory of parallel universes over coffee and, even worse, understand the concept. It will make a big difference in how well you measure up.

However, be of good cheer. Psychometricians are hard at work on the problem. They have carefully calculated an item “discrimination” statistic before placing any question on a test. This step enables you to feel confident that each item on the test performs in a satisfactory way with respect to the total test score. That is, the statistic guarantees that high scoring persons (as opposed to low scoring persons) will exhibit a greater probability of getting each item correct. Now if you don’t follow the logic of why such a statistic is important, psychometricians do. On the other hand, if you do follow the logic of the discrimination statistic, you will probably still have some concerns. This statistic, while indicating the degree of internal test consistency, doesn’t do much to solve the basic problem, namely, how smart is the group to whom you are being compared?

But take heart, life under classical test theory construction is not all bad. Many good tests have been constructed using classical test theory models. Every day people pass such certification exams, even though they are “sample dependent”. Consequently, it

can't be an insurmountable problem, look how many certified people we encounter every day. Also, realize help is on the way. Over the past 25 years, test theorists have developed new models (called item response theory models) that promise some relief from sample and item dependent tests, but don't wait for your association to adopt an item response theory exam. You may reach retirement age before becoming certified.

Just study hard for your exam and the short-comings of classical test theory models will probably not be a problem, but be aware of the fact that the person sitting next to you at the testing center may be helping to determine whether you pass or fail. So, be nice to them. You are both in the same situation - victims of certification exams developed using classical test theory models.