MEASUREMENT STANDARDS AND MODELS FOR CERTIFICATION EXAMINATIONS

Randall E. Schumacker, Ph.D.

Measurement standards

Standards for judging certification examinations are well known. Effective examinations should (a) produce linear measures so that results are amendable to mathematical operations, (b) consist of items that have been calibrated in a manner that does not depend on the examinee (sample) whose responses were used to calculate the item parameter estimates (sample-free estimates), (c) produce person scores that are independent of which test items they happen to take, and (d) provide test parameters that are invariant. While these standards are quite straightforward, developing a test that meets these standards can be a formidable task.

Classical test theory

In building tests, test developers have used classical test theory for more than 60 years. The classical test theory model outputs such statistics as item difficulty, item discrimination, guessing estimates, item-test intercorrelation, and various test distribution statistics. These statistics are used to judge the quality of the items that make up a test. While many good tests have been developed using classical models, shortcomings in the methodology are widely recognized. A serious limitation is the fact that classical test theory model item statistics – item difficulty and item discrimination – depend on the samples from which they are obtained. Thus, they are sample-dependent and are only useful when constructing tests for examinee populations that are similar to the sample of examinees from which the item statistics were obtained. Additionally, examinee scores are dependent on the choice of test items that reflect the difficulty level of the test. Tests with more difficult items produce lower proficiency scores, while tests containing a large number of easy items produce higher examinee scores. It is important that test scores not be a function of the particular items (test free) or the group (sample free) who took the items on the test.

Item response theory

Many of the problems associated with classical test theory can be avoided by using an item response theory model, which was developed during the last quarter of the twentieth century. Modern approaches to test construction widely use latent trait theory and have gained in popularity with the trend toward item banking, computer-assisted testing, and computer adaptive testing. Many professional associations that produce objective certification examinations are currently using one-, two-, or three-parameter item response theory models. The item response theory models tend to produce sample-invariant item parameters and utilize goodness-of-fit criteria to detect items that do not fit

the specified item response theory model. Most importantly, the property of sample invariance means that test developers can calibrate items and develop item banks useful in measuring examinee knowledge and establish objective standards of performance. The resulting person scores represent test free and sample free measures. We can assist professional associations in selecting and implementing either classical or item response theory models. The psychometric services are designed to accommodate the specific needs of each professional association.