CLASSICAL TEST ANALYSIS

Randall E. Schumacker, Ph.D.

Classical Test Analysis

Classical test theory (CTT) is an emanation of the early 20th century approaches to measuring individual differences. CTT introduces three basic measurement concepts (1) test score or observed score, (2) true score, and (3) error score. Classical test analysis postulates linking the observed test score (X) to the sum of the true score (latent unobservable score) and error score: X = T + E. The following assumptions underlie CTT: (a) true scores and error scores are uncorrelated, (b) the average error score in the population of examinees is zero, and (c) error scores on parallel tests are uncorrelated. Classical test analysis utilizes traditional item and sample dependent statistics. These include item difficulty and item discrimination estimates, distractor analyses, item-test intercorrelations, and a variety of related statistics. Most of the psychometric analyses have focused on examinee assessment at the test score level, rather than at the item level. Classical test analysis also typically includes a measure for the reliability of scores (i.e., Cronbach Alpha) and difficulty of the test.

Advantages of Classical Test Analysis

Several benefits are obtainable through the application of good instructional objectives and item writing using classical test analysis. First, when compared to item response theory models, analyses can be performed with smaller representative samples of examinees. This is particularly important when field-testing a measuring instrument. Secondly, classical test analysis employs relative simple mathematical procedures and model parameter estimations are conceptually straightforward. Thirdly, classical test analysis is often referred to as "weak models" because the assumptions are easily met by traditional testing procedures.

Disadvantages of Classical Test Analysis

While classical models have proven very useful in test development, they have several important limitations. The two statistics that form the cornerstones of most classical test theory, item difficulty and item discrimination, are both sample dependent. Higher item difficulty values are obtained from examinee samples of lower-average knowledge while lower item difficulty values occur from examinee samples of above-average knowledge. In terms of discrimination indices, higher values tend to be obtained from hetrogeneous examinee samples, and lower values are associated with homogeneous samples. Such sample dependency relationships reduce the overall utility of these statistics.

Classical test theory applications are also test dependent or "test-based". Test difficulty directly affects the resultant test scores. Higher knowledge scores are associated with tests composed of relatively easy items, and low knowledge scores can be a function of tests composed of items that are more difficult. The true-score model, upon which much of classical test theory is based, permits no consideration of examinee responses to any specific item. Thus, no basis exists to predict how a given examinee will perform on a particular test item.

Conclusion

Classical test analysis is still useful when constructing tests, although the trend is definitely toward item response theory (IRT) that provides for sample-free and item-free measurement. As professional associations become more familiar with IRT models, the utility of measurement interpretations provided by these models will undoubtedly become more widely appreciated, resulting in more IRT-based objective measurement and assessment of training. Additionally, as more professional associations move toward item banking and computer adaptive testing, IRT-based psychometric models will be required to implement meaningful measurement and assessment. We are committed to assisting our customers who need to implement and maintain classical test analysis of measurement data.