

Please use the following citation when referencing this work:

McGill, R. J. (2018). Psychometrics. In B. B. Frey (Ed.), *The SAGE encyclopedia of educational research, measurement, and evaluation* (pp. 1328-1329). Thousand Oaks, CA: SAGE Publications.

## **PSYCHOMETRICS**

Psychometrics is a branch of scientific psychology that is concerned with the theory and technique of psychological and educational assessment and measurement and the construction of instruments that are developed to appraise psychological and educational constructs (e.g., multidimensional achievement batteries, intelligence tests, and behavior rating scales). As a result, psychometric techniques are routinely employed throughout the corpus of quantitative educational research.

### **Historical Development**

Although the genesis of these methods has been debated extensively within the technical literature, Sir Francis Galton, has been called the “father of psychometrics,” and is credited by many as the first to apply early versions of these techniques when attempting to measure individual differences during the Victorian Era (1850-1900). As a result of these experiments, Galton developed the correlation coefficient that later served as the focal point for Charles Spearman’s research on intelligence and the subsequent discovery of the general intelligence factor (*g*) in 1904, a finding that Arthur Jensen argued in 1998 to be one of the greatest discoveries in the history of the social sciences.

### **Theoretical Approaches and Fundamental Concepts in Measurement**

As a general principle, measurement consists of rules for assigning symbols to objects so as to 1) represent objects numerically (scaling) and 2) determine whether objects fall within a particular

category (classification). While the vast majority of psychometric research has traditionally been devoted to the task of scaling, classification research has intensified as psychometric techniques have been applied more extensively to appraise the quality of diagnostic and other decision-making models in a variety of educational contexts including but not limited to: providing a diagnosis of a learning disorder, determining whether a student is at-risk for educational failure, and appraising whether a student has met *a priori* standards for educational attainment within a high-stakes accountability model.

The field of psychometrics is bifurcated by two divergent theoretical approaches to measurement: classical test theory (CTT) and modern test theory. The CTT model posits that any observed score is the produced from two hypothetical components expressed in the form

$$X = T + E$$

where  $X$  represents the observed score;  $T$  reflects the hypothetical true score for that construct; and  $E$  denoting a random error term. This model provides the foundation for estimating the reliability of a measure. Reliability refers to the degree to which differences in the observed score are consistent with differences in true scores and not the product of measurement error. Reliable measurement is also a necessary but not sufficient for establishing the validity of a measure, or, the degree to which a test measures the construct of interest. Validity is a multidimensional concept that requires analyzing elements such as the internal structure of a test and relationships between test scores and external criteria. In 1995, Samuel Messick argued that validity is of more importance than reliability, as it provides the bases for how a measure should be interpreted in clinical practice.

Modern test theory, also known as item response theory (IRT), is a psychometric approach emphasizing that an individual's response on a test item is influenced by their standing

on the construct being sampled as well as the degree of difficulty that item samples that particular construct. IRT techniques are commonly used in education and psychology to document test bias and to develop computerized adaptive tests.

*Ryan J. McGill*

**See also** Classical Test Theory, Confirmatory Factor Analysis, Exploratory Factor Analysis, Item Response Theory, Reliability, Standards for Educational and Psychological Testing, Validity

### **FUTHUR READINGS**

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *American Psychologist*, *61*, 27-41. doi: 10.1037/0003-066X.61.1.27

Furr, R. M., & Bacharach, V. R. (2014). *Psychometrics: An introduction*. Thousand, Oaks, CA: Sage.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*, 741-749. doi: 10.1037/0003-066X.50.9.741

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3<sup>rd</sup> ed.). New York: McGraw-Hill.