

Encyclopedia of Research Design

Predictive Validity

Contributors: Jessica Lynn Mislevy & André A. Rupp

Editors: Neil J. Salkind

Book Title: Encyclopedia of Research Design

Chapter Title: "Predictive Validity"

Pub. Date: 2010

Access Date: October 14, 2013

Publishing Company: SAGE Publications, Inc.

City: Thousand Oaks

Print ISBN: 9781412961271

Online ISBN: 9781412961288

DOI: <http://dx.doi.org/10.4135/9781412961288.n328>

Print pages: 1077-1079

This PDF has been generated from SAGE Research Methods. Please note that the pagination of the online version will vary from the pagination of the print book.

<http://dx.doi.org/10.4135/9781412961288.n328>

Validity refers to the degree to which a measure accurately measures the specific construct that it claims to be measuring. Criterion-related validity is concerned with the relationship between individuals' performance on two measures used to assess the same construct. It specifically measures how closely scores on a new measure are related to scores from an accepted criterion measure. There are two forms of criterion-related validity: predictive validity and concurrent validity.

Concurrent validity focuses on the extent to which scores on a new measure are related to scores from a criterion measure administered at the same point in time, whereas predictive validity uses the scores from the new measure to predict performance on a criterion measure administered at a later point in time.

Examples of contexts where predictive validity is relevant include the following:

The primary reason that predictive validity is of interest to users is that a concurrent criterion measure may not be available at the point in time at which decisions must be made. For example, it is not possible to evaluate a student's first-year college success at the time he or she is submitting college applications. Therefore, a measure that is able to correctly identify individuals who are likely to succeed in college at the time of application is a highly desirable tool for admissions counselors.

Before users can make decisions based on scores from a new measure designed to predict future outcomes reliably, they must have evidence that there is a strong relationship between the scores on the measure and the ultimate performance of interest. Such evidence can be obtained through a *predictive validation study*. In such a study, the new measure is administered to a sample of individuals [p. 1077 ↓] that is representative of the group for whom the measure is intended to be used. Next, researchers must allow enough time to pass for the behavior being predicted to occur. Once it has occurred, an already existing criterion measure is administered to the sample. The strength of the relationship between scores on the new measure and the scores on the criterion measure indicates the degree of predictive validity of the new measure.

The results of a predictive validation study are typically evaluated in one of two ways depending on the level of measurement of the scores from the two measures. In the case when both sets of scores are *continuous*, the degree of predictive validity is established via a *correlation coefficient*, usually the Pearson product-moment correlation coefficient. The correlation coefficient between the two sets of scores is also known as the *validity coefficient*. The validity coefficient can range from -1 to $+1$; large coefficients close to 1 in absolute value indicate high predictive validity of the new measure.

Figure 1 displays hypothetical results of a predictive validation study reflecting a validity coefficient of .93. The predictive validity of the aptitude measure is quite satisfactory because the aptitude measure scores correlate highly with the final exam scores collected at the end of the program; simply put, individuals scoring well on the aptitude measure later score well on the final exam.

Figure 1 Scatterplot of Scores on a Foreign Language Aptitude Exam and Language Fluency Exam

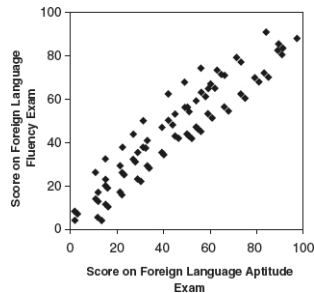


Table 1 Cross Classification of Classification From an Employment Measure and Performance Rating

Initial Classification From Employment Measure	End-of-Quarter Job Performance Rating From Supervisor	
	Unsatisfactory	Satisfactory
Unsatisfactory	14	3
Satisfactory	7	86

In the case when the outcomes on both measures are classifications of individuals, *coefficients of classification agreement* are typically used, which are variations of correlation coefficients for *categorical data*. Evidence of high predictive validity is obtained when the classifications based on the new measure tend to agree with classifications based on the criterion measure.

Table 1 displays hypothetical data for a predictive validation study reflecting a classification consistency of 91%. The predictive validity of the employment measure is high because the resulting classification aligns well with the supervisor's job performance rating in almost all cases; simply put, the outcome of the measure predicts well the rating of the supervisor.

When determining the predictive validity of a new measure, the selection of a valid criterion measure is critical. Ideally, as noted by Robert M. Thorndike, criterion measures should be relevant to the desired decisions, free from bias, and reliable. In other words, they should already possess all the ideal measurement conditions that the new measure should possess also. Specifically, criterion measures should be

If the criterion against which the new measure is compared is invalid because it fails to meet these quality standards, the results of a predictive validation study will be compromised. Put differently, the results from a predictive validation study are only as useful as the quality of the criterion measure that is used in it. It is thus key to select the criterion measure properly to ensure that a lack of a relationship between the scores from the new measure and the scores from the criterion measure is truly due to problems with the new measure and not due to problems with the criterion measure.

Of course, the selection of an appropriate criterion measure will also be influenced by the availability or cost of the measure. Thus, the practical limitations associated with criterion measures that are inconvenient, expensive, or highly impractical to obtain may outweigh other desirable qualities of these measures.

Jessica Lynn Mislévy and André A. Rupp

<http://dx.doi.org/10.4135/9781412961288.n328>

See also

Further Readings

Crocker, L., & Algina, J. (1986). Introduction to classical & modern test theory. Belmont, CA: Wadsworth.

Thorndike, R. M. (2005). Measurement and evaluation in psychology and education (7th ed.). Upper Saddle River, NJ: Pearson.