

VALIDITY IN MEASUREMENT

Construct Validity

The **validity** of a measurement method is the extent to which it measures what it is supposed to measure. For example, does an intelligence test really measure intelligence? Does a self-esteem scale really measure self-esteem? Does whether or not someone throws his soda can into a recycling bin indicate his level of environmental consciousness? This overall validity of a measurement method is sometimes referred to as its **construct validity**.

Evaluating Validity

There are some extremely important points to remember about the way that psychologists evaluate the validity of a measurement method. First, this process requires **empirical evidence**. A measurement method cannot be declared valid or invalid before it has ever been used and the resulting scores have been thoroughly analyzed.

Second, it is an ongoing process. The conclusion that a measurement method is valid generally depends on the results of many studies done over a period of years. In fact, every new study using that measurement method provides additional evidence for or against its validity.

Third, validity is not an all-or-none property of a measurement method. It is possible for a measurement method to be judged "somewhat valid" or for one measure to be considered "more valid" than another. It is also the case that a measurement method might be valid for one subject population or in one context, but not in another. For example, it would be fine to conclude that an English-language achievement test is valid for children who are native English speakers but not for children who are still in the process of learning English.

Types of Validity

Construct validity is often broken down into separate types. The best way to think of these types is that each one concerns a different kind of evidence that a measurement method is measuring what it is supposed to. Here are some of the different types.

Face Validity

Face validity is the extent to which the measurement method appears "on its face" to measure the construct of interest. For instance, does the Rosenberg Self-Esteem scale appear to be measuring self-esteem? This is a very weak kind of evidence. The simple fact that a test *seems* to be measuring a particular construct is no guarantee that it is. However, face validity can be important because it can affect people's attitudes toward a test. For example, people might have negative reactions to an intelligence test that did not appear to them to be measuring their intelligence.

Content Validity

Content validity is the extent to which the measurement method covers the entire range of relevant behaviors, thoughts, and feelings that define the construct being measured. For example, one's attitude toward an object is considered to consist of thoughts about the object, feelings about the object, and behaviors toward the object. Therefore, a test to assess one's attitude toward taxes should include items about thoughts, feelings, and behaviors. If test anxiety is thought to include both nervous feelings and negative thoughts, then any measure of test anxiety should cover both of these aspects. A course exam has good content validity if it covers all the material that students are supposed to learn and poor content validity if it does not.

Criterion Validity

Criterion validity is the extent to which people's scores are correlated with other variables or **criteria** that reflect the same construct. For example, an IQ test should correlate positively with school performance. An occupational aptitude test should correlate positively with work performance. A new measure of self-esteem should correlate positively with an old established measure. When the criterion is something that will happen or be assessed in the future, this is called **predictive validity**, as when SAT scores are shown to be correlated with a students' eventual college grades. When the criterion is something that is happening or being assessed at the same time as the construct of interest, it is called **concurrent validity**, as when scores on a new self-esteem test are shown to be correlated with scores on an existing test taken at the same time.

Discriminant Validity

Discriminant validity is the extent to which people's scores are *not* correlated with other variables that reflect distinct constructs. Imagine, for example, that a researcher with a new measure of self-esteem claims that self-esteem is independent of mood; a person with high self-esteem can be in either a good mood or a bad mood (and a person with low self-esteem can too). Then this researcher should be able to show that his self-esteem measure is not correlated (or only weakly

correlated) with a valid measure of mood. If these two measures were highly correlated, then we would wonder whether his new measure really reflected self-esteem as opposed to mood.

As a another example, there is a [Need for Cognition Scale](#) that measures the extent to which people like to think and value thinking, which is supposed to be largely independent of people's intelligence. Just because someone is intelligent does not mean that he or she has a high need for cognition, and just because someone is less intelligent does not mean that he or she has a low need for cognition. In this case, one would expect to see that need for cognition scores and intelligence test scores were not highly correlated. Otherwise it would look like need for cognition was just another measure of intelligence.

Other Types: [External and Internal Validity](#)

© 2006 California State University, Fresno | Dr. Price & Dr. Oswald | Updated 07/21/2008 | Disclaimer | Contact Web Master