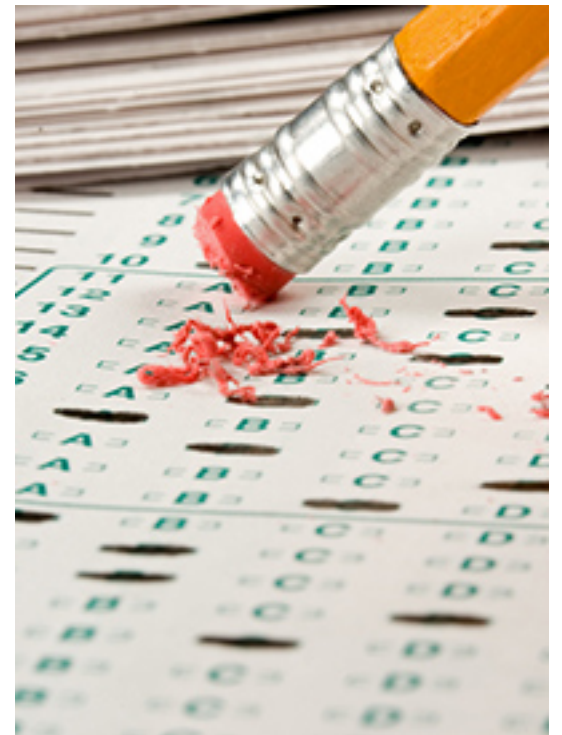


Chapter 7.3 Test Validity

[Dr. Christopher L. Heffner](https://allpsych.com/researchmethods/validityreliability/)

<https://allpsych.com/researchmethods/validityreliability/>

Whenever a test or other measuring device is used as part of the data collection process, the validity and reliability of that test is important. Just as we would not use a math test to assess verbal skills, we would not want to use a measuring device for research that was not truly measuring what we purport it to measure. After all, we are relying on the results to show support or a lack of support for our theory and if the data collection methods are erroneous, the data we analyze will also be erroneous.



Test Validity

Validity refers to the degree in which our test or other measuring device is truly measuring what we intended it to measure. The test question “ $1 + 1 = \underline{\hspace{1cm}}$ ” is certainly a valid basic addition question because it is truly measuring a student’s ability to perform basic addition. It becomes less valid as a measurement of advanced addition because as it addresses some required knowledge for addition, it does not represent all of knowledge required for an advanced understanding of addition. On a test designed to measure knowledge of American History, this question becomes completely invalid. The ability to add two single digits has nothing do with history.

For many constructs, or variables that are artificial or difficult to measure, the concept of validity becomes more complex. Most of us agree that “ $1 + 1 =$

_____” would represent basic addition, but does this question also represent the construct of intelligence? Other constructs include motivation, depression, anger, and practically any human emotion or trait. If we have a difficult time defining the construct, we are going to have an even more difficult time measuring it. **Construct validity** is the term given to a test that measures a construct accurately and there are different types of construct validity that we should be concerned with. Three of these, concurrent validity, content validity, and predictive validity are discussed below.

Concurrent Validity. Concurrent Validity refers to a measurement device’s ability to vary directly with a measure of the same construct or indirectly with a measure of an opposite construct. It allows you to show that your test is valid by comparing it with an already valid test. A new test of adult intelligence, for example, would have concurrent validity if it had a high positive correlation with the Wechsler Adult Intelligence Scale since the Wechsler is an accepted measure of the construct we call intelligence. An obvious concern relates to the validity of the test against which you are comparing your test. Some assumptions must be made because there are many who argue the Wechsler scales, for example, are not good measures of intelligence.

Content Validity. Content validity is concerned with a test’s ability to include or represent all of the content of a particular construct. The question “ $1 + 1 = \underline{\quad}$ ” may be a valid basic addition question. Would it represent all of the content that makes up the study of mathematics? It may be included on a scale of intelligence, but does it represent all of intelligence? The answer to these questions is obviously no. To develop a valid test of intelligence, not only must there be questions on math, but also questions on verbal reasoning, analytical ability, and every other aspect of the construct we call intelligence. There is no easy way to determine content validity aside from expert opinion.

Predictive Validity. In order for a test to be a valid screening device for some

future behavior, it must have predictive validity. The SAT is used by college screening committees as one way to predict college grades. The GMAT is used to predict success in business school. And the LSAT is used as a means to predict law school performance. The main concern with these, and many other predictive measures is predictive validity because without it, they would be worthless.

We determine predictive validity by computing a correlational coefficient comparing SAT scores, for example, and college grades. If they are directly related, then we can make a prediction regarding college grades based on SAT score. We can show that students who score high on the SAT tend to receive high grades in college.