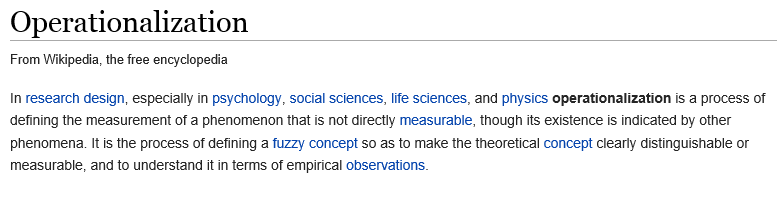
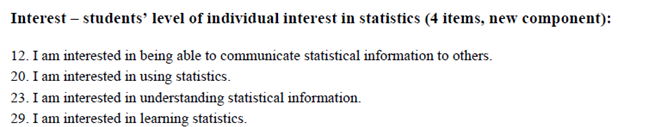
In this handout, we will discuss different types of and methods for establishing *validity*. Recall that this concept was defined in an earlier handout as follows.

|  |
| --- |
| Definition |
| **Validity** – This is the extent to which survey questions measure what they are supposed to measure. |

In order for survey results to be useful, the survey must demonstrate **validity**. To better understand this concept, it may help to also consider the concept of **operationalization**. Wikipedia.org defines this as follows:  
  


For example, in the previous handout we considered measuring students’ interest in statistics using the SATS (Survey of Attitudes Toward Statistics). Note that the construct of *Interest* is a theoretical and rather vague concept – there is no clear or obvious way to measure this. So, the creators of this survey defined their own measure of *Interest* using these four questions:  
  


The resulting sub-score for *Interest* is their operationalization of this construct. Will the resulting score really measure *Interest*? Will the resulting scores for the other constructs on the SATS really measure what the researchers intend? These are the questions we seek to answer when establishing the **construct validity** of this survey.

In the remainder of this handout, we will introduce various types of construct validity and briefly discuss how survey instruments are shown to be valid.

**TYPES OF CONSTRUCT VALIDITY**

When designing survey questionnaires, researchers may consider one or more of the following types of **construct validity**.

|  |
| --- |
| Types of Construct Validity |
| **Face Validity** –An operationalization has face validity when it appears to observers that it truly measures the construct it is intended to measure. |
| **Content Validity** – An operationalization has construct validity when it adequately covers the range of meanings included in the construct it is intended to measure. |
| **Criterion-Related Validity** – This is assessed by investigating the relationship between the operationalization and other variables. The following are all specific types of criterion-related validity:   * **Predictive Validity** – This is established if the operationalization is able to predict another variable that it should theoretically be able to predict. * **Concurrent Validity** – This is established if the operationalization is well correlated with other variables measured at the same time to which it should theoretically be related. * **Convergent Validity** – This is established if the operationalization is well correlated with other variables to which it should theoretically be related. * **Discriminant Validity** – This is established if the operationalization is shown to be dissimilar from other variables that it theoretically should not be related to. |

Ways to establish these types of validity are discussed in more detail below.

**Face Validity**

Face Validity can’t be established with any sort of statistical analysis. Instead, it’s based on a subjective judgment call (which makes it one of the weaker ways to establish construct validity). The best approach for establishing face validity is to assemble a panel of experts to report on whether or not they feel an operationalization appears to be a good measure of the construct of interest.

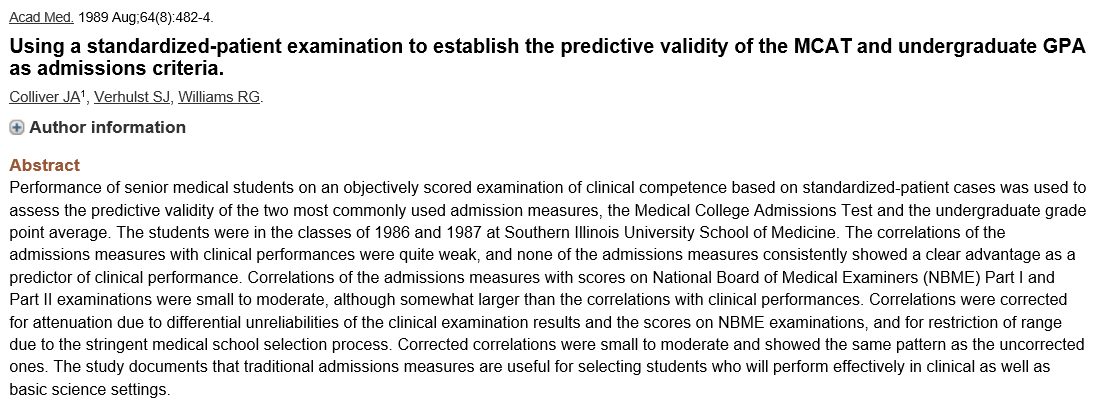
|  |  |  |
| --- | --- | --- |
| **Survey Questions** |  | **Concept** |

**Content Validity**

Like face validity, this is also best established by assembling a panel of experts. The researcher will describe the content domain for their construct (e.g., “My goal is to measure students’ interest in statistics.”) The experts will then be asked to judge how well the operationalization covers all of the criteria that constitute the content domain.

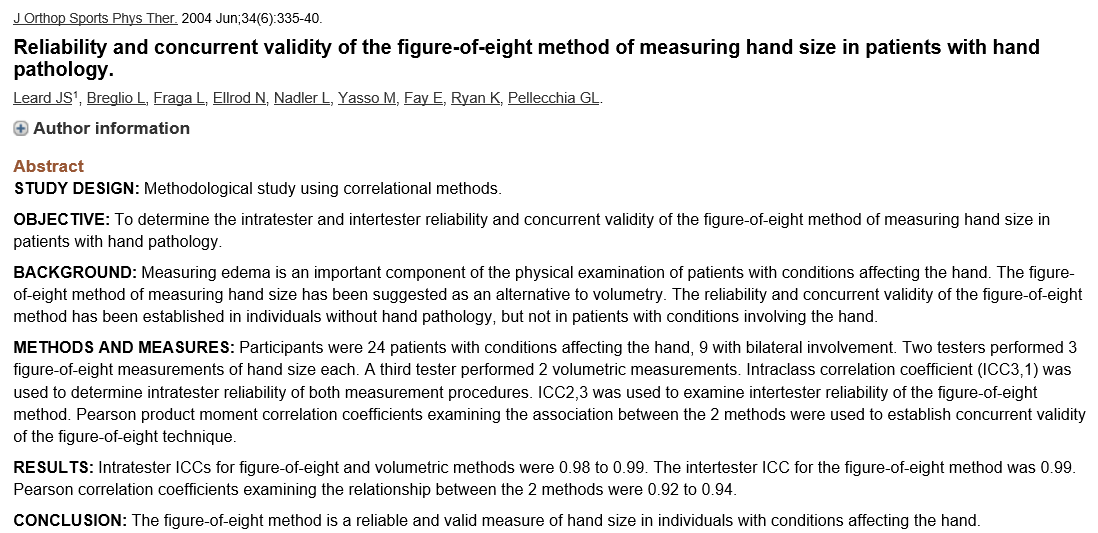
|  |  |  |
| --- | --- | --- |
| **What does Interest mean (in statistics)?** |  | **Covers aspects of Interest** |

**Predictive Validity**

As mentioned above, this involves assessing the ability of the operationalization to predict something it should theoretically be able to predict. Typically, this involves the computation of correlation coefficients. For example, the abstract below describes a study conducted to assess the predictive validity of the Medical College Admissions Test (MCAT).  
  


**Concurrent Validity**

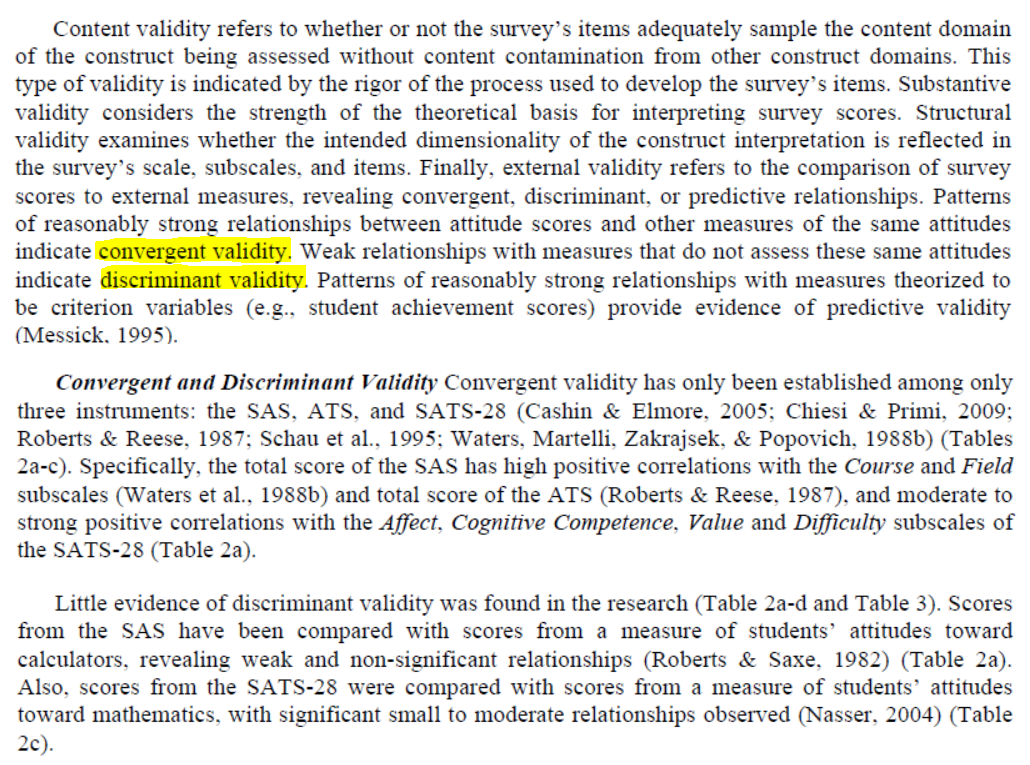
This involves assessing the strength of the relationship between the operationalization of interest and other variables measured *at the same time* to which the operationalization should theoretically be related. Once again, this typically this involves the computation of correlation coefficients. For example, consider the following abstract.



**Convergent and Divergent Validity**

As stated above, we establish **convergent** validity if the operationalization of interest is well correlated with other variables to which it theoretically should be related; alternatively, we establish **divergent** validity if the operationalization of interest is *not* well correlated with other variables to which it theoretically should not be related.

For example, consider the following excerpt from a paper titled “Surveys Assessing Students’ Attitudes Toward Statistics: A Systematic Review of Validity and Reliability” written by Meaghan M. Nolan et al. This was published in the *Statistics Education Research Journal*, 11 (2), pp. 103-123 (November 2012).



Note that to estimate the degree to which any two measures are related to each other, we typically use the correlation coefficient.

**A MORE MODERN CONCEPT OF CONSTRUCT VALIDITY**

In the 1990s, Samuel Messick proposed a new, more unified framework for the concept of validity. This framework describes six distinguishable aspects of validity. He argues that these six aspects should be viewed as interdependent and complementary to one another.

| Modern Concept of Construct Validity |
| --- |
| **Content Validity** – An operationalization has content validity when it adequately covers the range of meanings included in the construct it is intended to measure (i.e., it is representative of the construct). When assessing this, one should also consider the content relevance of each item and its technical quality. |
| **Substantive Validity** – This considers the strength of the theoretical rationales for interpreting the survey scores. For example, consider the following excerpt from the aforementioned paper by Nolan et al. |
| **Structural Validity** – This is assessed by investigating the degree to which the operationalization adequately reflects the dimensionality of the construct to be measured. For example, a researcher may conduct a factor analysis using the observed scores for a survey. If the factor analysis reveals the same number of factors as there were constructs measured on the survey (and the factor loadings show that the right questions are grouped together within factors), then structural validity is established. Such an analysis is discussed in the following excerpt from the Nolan et al. paper. |
| **Generalizability** – This examines the extent to which scores generalize across different population groups, different situations or settings, different time periods, and/or to other operationalizations representative of the construct domain. |
| **External Validity** – This refers to the comparison of the operationalization of interest to external measures (see the earlier discussions of convergent, divergent, and predictive validity). |
| **Consequential Validity** – This includes gathering evidence and rationales for evaluating the consequences of score interpretations from a survey. Researchers should accrue evidence of positive consequences and evidence that adverse consequences are minimal. |

With this more modern framework, the process of establishing construct validity could potentially involve the accumulation of all six of the aforementioned forms of validity evidence. Note, however, that a “compelling argument” for validity can still be made even if some of these aren’t addressed.

Source: Messick, Samuel. (1995). Validity of psychological assessment: Validation of inferences from persons’ responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

**CONSIDERING VALIDITY AND RELIABILITY SIMULTANEOUSLY**

Often, validity and reliability are viewed as completely separate ideas. To think about how the two are related, we can use a “target” analogy. Let the center of the target represent the construct you intend to measure. For each subject that responds to your survey questionnaire, you take a shot at the target. If you measure the concept perfectly for a person, you hit the center of that target. The figure below shows four possible situations.

|  |  |  |  |
| --- | --- | --- | --- |
| High Reliability High Validity | High Reliability  Not Valid | Low Reliability Low Validity | Low Reliability Not Valid |
|  |  |  |  |
| Source: <https://www.socialresearchmethods.net/kb/relandval.php> | | | |