About Validity and Reliability
<table>
<thead>
<tr>
<th>TABLE OF CONTENTS</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>About Validity and Reliability</td>
<td>1</td>
</tr>
<tr>
<td>What is VALIDITY?</td>
<td>1</td>
</tr>
<tr>
<td>What is RELIABILITY?</td>
<td>1</td>
</tr>
<tr>
<td>Types of Validity</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Face Validity</td>
<td>1</td>
</tr>
<tr>
<td>Content Validity</td>
<td>2</td>
</tr>
<tr>
<td>Criterion Validity</td>
<td>2</td>
</tr>
<tr>
<td>Concurrent Validity</td>
<td>2</td>
</tr>
<tr>
<td>Predictive Validity</td>
<td>2</td>
</tr>
<tr>
<td>Construct Validity</td>
<td>2</td>
</tr>
<tr>
<td>Types of Reliability</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>3</td>
</tr>
<tr>
<td>Test-Retest Reliability</td>
<td>3</td>
</tr>
<tr>
<td>Multiple Forms</td>
<td>3</td>
</tr>
<tr>
<td>Internal Consistency Approaches</td>
<td>4</td>
</tr>
<tr>
<td>Spearman-Brown and Cronbach’s Alpha</td>
<td>4</td>
</tr>
<tr>
<td>What is the Spearman-Brown formula?</td>
<td>4</td>
</tr>
<tr>
<td>What is Cronbach’s Alpha?</td>
<td>5</td>
</tr>
<tr>
<td>Internal and External Validity</td>
<td></td>
</tr>
<tr>
<td>What is Internal Validity?</td>
<td>5</td>
</tr>
<tr>
<td>History</td>
<td>6</td>
</tr>
<tr>
<td>Maturation</td>
<td>6</td>
</tr>
<tr>
<td>Testing</td>
<td>6</td>
</tr>
<tr>
<td>Instrumentation</td>
<td>6</td>
</tr>
<tr>
<td>Selection</td>
<td>6</td>
</tr>
<tr>
<td>Experimental Attrition</td>
<td>6</td>
</tr>
<tr>
<td>What is External Validity?</td>
<td>6</td>
</tr>
<tr>
<td>Reactive Effects of Testing</td>
<td>7</td>
</tr>
<tr>
<td>Unrepresentative samples</td>
<td>7</td>
</tr>
<tr>
<td>Reactive Settings</td>
<td>7</td>
</tr>
<tr>
<td>Multiple-Treatment Interference</td>
<td>7</td>
</tr>
<tr>
<td>Improving Validity and Reliability</td>
<td>8</td>
</tr>
<tr>
<td>How Can We Improve Validity and Reliability?</td>
<td>8</td>
</tr>
<tr>
<td>Improved training of those who apply the measure</td>
<td>8</td>
</tr>
<tr>
<td>Interview subjects who have experienced the measuring device</td>
<td>8</td>
</tr>
<tr>
<td>Assess each item on a multiple-item questionnaire</td>
<td>8</td>
</tr>
<tr>
<td>Improve testing on minority populations</td>
<td>8</td>
</tr>
<tr>
<td>Glossary</td>
<td>9</td>
</tr>
<tr>
<td>References</td>
<td>11</td>
</tr>
</tbody>
</table>
About Validity and Reliability

What is VALIDITY?

• Validity is the degree to which a test measures what it claims to measure. If the test is not valid, then results cannot be accurately interpreted and applied.

What is RELIABILITY?

• Reliability refers to the consistency of a measure's results for each application of the measure. Reliable measures demonstrate variation that is strictly attributable to manipulation of the variables under investigation.

Types of Validity

Introduction

• Validity can be broken over the following sub-types:

  1) Face validity
  2) Content validity
  3) Criterion validity
  4) Concurrent validity
  5) Predictive validity
  6) Construct validity

Face Validity

• A subjective form of validity measure, which associates the variable of interest with the proposed study variable, by relying heavily on logic and common sense.

• Given its nature, face validity should only be considered as a starting point.
**Content Validity**

- A less subjective form of validity measure than face validity, although it does extend from face validity, which relies on an assessment of whether the proposed measure incorporates all content of a particular construct.

**Criterion Validity**

- A much more objective measure than both face validity and content validity, criterion validity relies on comparison between the proposed measure and a measure already established to be valid, as concerns the variable of interest.

**Concurrent Validity**

- In applying your proposed measure and a measure with established validity to the same data set, concurrent validity refers to strong correlation between the measures' results.

- A correlation of $r = 0.50$ is the minimum acceptable for declaring concurrent validity.

**Predictive Validity**

- A valuation of your proposed measure's ability to predict future events.

- Unlike concurrent validity, predictive validity first requires a collection of scores on the measure, followed by a collection of scores on the criterion measure at a later time.

**Construct Validity**

- A complex validity measure, first identified by Lee J. Cronbach and Paul E. Meehl in 1995, construct validity evaluates whether your proposed measure correlates with all concepts related to the theory under investigation.
Types of Reliability

Introduction

- Reliability can be broken over the following sub-types:
  1) Test-Retest
  2) Multiple Forms
  3) Internal Consistency Approaches

Test-Retest Reliability

- The most common form of reliability measure, test-retest reliability refers to correlating multiple applications of a single measure, conducted at different times.

- Generally, $r \geq 0.80$ is indicative of a reliable measure.

Multiple Forms

- Similar to test-retest reliability, multiple forms reliability refers to the application of equivalent forms of a measure (e.g. the same measure, with differently worded questions), and their correlation.

- As before, $r \geq 0.80$ is indicative of a reliable measure.

- Multiple forms reliability has an advantage over test-retest reliability, in that it can be applied in a single session of testing. In this circumstance, subjects are given both forms of the measure concurrently.

  Unfortunately, subjects may come to realize that certain questions are measuring similar concepts, and modify their answers accordingly. This phenomenon is called "multiple testing effects," and is seen to occur in test-retest reliability as well.
Internal Consistency Approaches

- We will discuss the most common form of internal consistency, Split-half Reliability, which refers to a single administration of a measure, correlated with itself. Essentially, the measure is divided equally, with the halves having a comparable focus. Then, as with test-retest and multiple forms, the halves are correlated.

- As before, \( r \geq 0.80 \) is indicative of a reliable measure, with 8 to 10 items per half being considered a minimum requirement.

- An intuitive disadvantage of the split-half reliability procedure is that the correlation coefficient may not represent an accurate measure of the reliability. This is due to the fact that a single scale is being split into two scales, decreasing the reliability of the measure as a whole. As such, longer scales are preferable to shorter ones.

To solve this problem, researchers often apply the Spearman-Brown formula.

Spearman-Brown and Cronbach’s Alpha

What is the Spearman-Brown formula?

- The Spearman-Brown formula adjusts the correlation coefficient, preventing a decrease in the observed reliability of a measure when that measure is split into multiple parts.

- The formula is as follows:

\[
r = \frac{2r_i}{1 + r_i}
\]

where \( r_i \) = uncorrected correlation coefficient

\( r \) = corrected correlation coefficient
(also called the reliability coefficient)
**What is Cronbach’s Alpha?**

- Cronbach's alpha is another internal consistency approach, used to overcome disadvantages seen with the split-half reliability approach.
- Cronbach's alpha is, in essence, the average of all possible split-half correlations within a measure.
- Calculating Cronbach's alpha is beyond the scope of this tutorial, and can be easily done using statistical software packages like SPSS.
- Cronbach's alpha is simply another technique used to establish reliability for a measurement scale.

**Internal and External Validity**

**What is Internal Validity?**

- Internal validity seeks to answer the question: Is the observed effect of the independent variable on the dependent variable an actuality?

  Put another way, internal validity classifies the effect of the stimulus (independent variable) on the measured variable (dependent variable), attributing it to extraneous sources or not.

- If the observed effect is determined to be due to the stimulus, then the measure is said to be internally valid.

- **Threats to internal validity** can be broken over the following list:

  1) History
  2) Maturation
  3) Testing
  4) Instrumentation
  5) Selection
  6) Experimental Attrition
**History**

- Most prevalent in field experiments, history refers to events external to the study, that may affect the dependent variable.

**Maturation**

- Refers to changes in an experimental environment strictly attributable to the passage of time.

**Testing**

- Refers to threats created by a subject having multiple exposures to the same measure.

**Instrumentation**

- Refers to a threat to internal validity, which occurs due to a researcher recording observations differently at the start of a study vs. the end of the same study.

**Selection**

- This threat to internal validity refers to the recruitment of different types of people for different groups within a study, which makes comparisons between the groups largely uninformative.

**Experimental Attrition**

- This threat refers to differing drop-out rates between groups within a study. As with selection, this results in comparison between the groups being mostly useless.
What is External Validity?

- Unlike internal validity, external validity considers environments external to the experimental setting.
- This form of validity is concerned with the generalizability of observed results.
- Four major threats to external validity are as follows:
  1) Reactive Effects of Testing
  2) Unrepresentative samples
  3) Reactive Settings
  4) Multiple-Treatment Interference

Reactive Effects of Testing

- The setting may affect the subjects', or even the experimenters', behaviour.

Unrepresentative samples

- Samples may not be representative of the general population, having a detrimental effect on the generalizability of the results. This largely stems from the volunteer nature of experiments.

Reactive Settings

- The setting may affect the subjects', or even the experimenters', behaviour.

Multiple-Treatment Interference

- Often, more than one independent variable will be under investigation. These variables will be presented in a particular order, which may not be representative of the external environment, affecting the generalizability of the results.
Improving Validity and Reliability

How Can We Improve Validity and Reliability?

● When a measure has insufficient validity or reliability, researchers often attempt to redesign the measure, in an effort to reach acceptable levels.

Improved training of those who apply the measure

● Especially useful for subjective measures, researchers can train users of the measure to detect and avoid the introduction of bias.

Interview subjects who have experienced the measuring device

● Receiving feedback from unbiased subjects can illuminate previously unperceived fallbacks of the measure. For instance, perhaps an item the researcher felt was crystal clear, was viewed as being ambiguous by the subject.

Assess each item on a multiple-item questionnaire

● Taking this step may allow for redundant or faulty items to be identified and removed.

Improve testing on minority populations

● Often, researchers do not consider the uniqueness of minority populations in the development of a measure. Proper translation, if applicable, is a must.
Glossary

Validity: Validity is the degree to which a test measures what it claims to measure.

Reliability: Reliability refers to the consistency of a measure's results for each application of the measure.

Face validity: A subjective form of validity measure, which associates the variable of interest with the proposed study variable, by relying heavily on logic and common sense.

Content Validity: A less subjective form of validity measure than face validity, although it does extend from face validity, which relies on an assessment of whether the proposed measure incorporates all content of a particular construct.

Criterion Validity: A much more objective measure than both face validity and content validity, criterion validity relies on comparison between the proposed measure and a measure already established to be valid, as concerns the variable of interest.

Concurrent Validity: In applying your proposed measure and a measure with established validity to the same data set, concurrent validity refers to strong correlation between the measures' results.

Predictive Validity: A valuation of your proposed measure's ability to predict future events.
Construct Validity: A complex validity measure, first identified by Lee J. Cronbach and Paul E. Meehl in 1995, construct validity evaluates whether your proposed measure correlates with all concepts related to the theory under investigation.

Test-Retest Reliability: The most common form of reliability measure, test-retest reliability refers to correlating multiple applications of a single measure, conducted at different times.

Multiple Forms Reliability: Similar to test-retest reliability, multiple forms reliability refers to the application of equivalent forms of a measure (e.g. the same measure, with differently worded questions), and their correlation.

Split-half Reliability: refers to a single administration of a measure, correlated with itself.

Spearman-Brown Formula: adjusts the correlation coefficient, preventing a decrease in the observed reliability of a measure when that measure is split into multiple parts.

Cronbach's alpha: another internal consistency approach, used to overcome disadvantages seen with the split-half reliability approach, which is, in essence, the average of all possible split-half correlations within a measure.

Internal Validity: If the observed effect is determined to be due to the stimulus, then the measure is said to be internally valid.

External Validity: This form of validity is concerned with the generalizability of observed results.
References