SHAZAM! SELECTING INSTRUMENTS TO USE IN RESEARCH STUDIES

Patricia Hart, PhD, RN
Nancy Ballard, MSN, RN, NEA-BC
Objectives

Participants will be able to:

• Describe the purpose of instrument for measurement.
• Describe the steps of evaluating a measurement instrument for research.
• Describe the different types of validity for a measurement instrument.
• Describe the different types of reliability for a measurement instrument.
• Identify resources to locate measurement instruments.
Purpose of Quantitative Research

Gaining Knowledge by Measuring Variables

- Quantify presence - Descriptive
- Quantify strength of association – Correlation
- Quantify effect of an intervention – Experimental, Quasi-experimental (Cause and Effect)
Variables

- Physiological variables - Direct measurement (BP, height, weight, hemoglobin, $O_2$ Sat.).
- Other phenomenon – Indirect Measurement (Depression, stress, grief, self-care agency, practice environment).

Instruments used to measure both types of variables. All instruments must be valid and reliable.
Choosing an Instrument

• Population & variables of interest
• History of selected tool use
• Appropriateness
  • Language
  • Reading level
  • Pictures (visual scale vs or words)
  • Ethnicity

All important to Validity & Reliability
Validity-Defined

• The property of an assessment tool that indicates that the tool does what it says it does. If it does, the test scores have meaning (Salkind, 2006)

• Accuracy of inferences from test scores; Soundness and relevance of proposed interpretation of scores (Cronbach, 1984)

• Degree to which an individual possesses some hypothetical trait or quality presumed to be reflected by performance on a measure (Waltz et al, 1991)
Types of Instrument Validity

Face Validity

Content Validity

Construct Validity

Criterion Related Validity
Face Validity- Reviewed by Experts

- An assessment of instrument by individuals whose characteristics are measured.
  - Accuracy: perception that an item correctly measures the intended construct
  - Acceptance: perception that item is appropriate to stated intent of measure
  - Relevance: degree to which item matches an individual’s situation

Content Validity – Numerical Rating

• Do the items sampled for inclusion on the measure adequately represent the domain of content addressed by the measure. Is largely a function of how a measure is developed (Waltz et al., 1991).

• Each item is scored \[ CVR = \frac{n_e - N/2}{N/2} \]
  
  • CVR = the content validity ratio
  • \( n_e \) = the number of judges who selected the item as essential
  • \( N \) = total number of judges
  • Score of < 0 indicates item is not essential to the instrument.
Construct Validity

• Extent to which measure relates to other measures through hypothesis testing
• Evidence of support is gathered piece by piece
  • Known or contrasted groups approach
  • Factor analysis (Factorial Validity)
  • Multitrait-Multimethod Matrix (less common)
Criterion Related Validity

- When wish to infer from a measure an individual’s probable standing on some other variable or criterion
  - Predictive: extent to which future level of performance on criterion can be predicted from knowledge of performance on a prior measure. Example: SAT Score
  - Concurrent: extent to which measure may be used to estimate present standing on the criterion. Example: Self Efficacy Scale
<table>
<thead>
<tr>
<th>Type of Validity</th>
<th>When You Use It</th>
<th>How You Do It</th>
<th>What you can say about it.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Content Validity</strong></td>
<td>Want to know whether a sample of items truly reflects an entire universe of items or a certain topic.</td>
<td>Examine content for accuracy of what you want to test.</td>
<td>“My set of questions fairly assesses the knowledge of content of interest.”</td>
</tr>
<tr>
<td><strong>Construct Validity</strong></td>
<td>When you want to know if a test measures some underlying psychological construct</td>
<td>Correlate the scores with some theorized outcome that reflects the construct for which the test was designed.</td>
<td>“It is true- men who participate in contact sports w/bodily contact score higher on the ______ test for aggression.”</td>
</tr>
<tr>
<td><strong>Criterion Validity</strong></td>
<td>Want to know if test scores are systematically related to other criteria indicating the test taker is competent in or demonstrates the characteristic in a certain area of interest.</td>
<td>Correlate the scores from the test with some other measure that is already valid and assesses the same set of abilities.</td>
<td>The EATS (test of culinary skills) has been shown to be correlated with being a fine chef 2 years after culinary school (predictive validity)</td>
</tr>
</tbody>
</table>

Reliability

• When a *Measurement Procedure* yields **consistent scores when the phenomenon being measured is not changing.**

• Degree to which scores are free of “measurement error”

• Consistency of measurement
Reliability

- Concept of reliability
  - Do items allegedly belonging to a scale actually assess a single construct
    - Pain
    - Hope
    - Quality of life
  - Do scales measuring a single construct produce consistent estimates of that construct across multiple measurements
Correlation Coefficients

- Measures the strength of the relationship between two variables
- Values range from -1.00 for a perfect negative relationship, through zero for no relationship, to +1.00 for a perfect positive correlation.
- The closer the correlation is to +1 or -1, the more closely the two variables are related.
- If the correlation is positive, it means that as one variable gets larger the other variable gets larger (often called “direct” correlation).
- If the correlation is negative it means that as one variable gets larger, the other variable gets smaller (often called an "inverse" correlation).
<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal Consistency Reliability</td>
<td>The extent to which items on the questionnaire are measuring the same thing.</td>
</tr>
<tr>
<td>Equivalence Reliability</td>
<td>The extent of consistency of scores obtained from the data collected by use of questionnaires.</td>
</tr>
<tr>
<td>Stability</td>
<td>The extent to which the same or similar scores are obtained with repeated testing with the same group of respondents. In other words, the scores are consistent from one time to the next.</td>
</tr>
</tbody>
</table>
# Internal Consistency Reliability

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Values</th>
</tr>
</thead>
</table>
| **Split-half**           | Divide the items to form two questionnaires. Give both questionnaires and compare the halves using correlation coefficients.  
• Even/odd questions  
• First half/last half  
• Randomly selected items | A common rule of thumb is 0.80 or high for adequate reliability and 0.90 or higher for good reliability. correlation coefficient. |
<p>| <strong>Cronbach’s alpha</strong>    | Most widely used method for estimating internal consistency. It is a function of the intercorrelations of items and the number of items in the scale. | Should be at least 0.70 or higher to retain an item in an &quot;adequate&quot; scale; and many researchers require a cut-off of 0.80 for a &quot;good scale.&quot; |
| <strong>Kuder-Richardson Coefficient</strong> | Used with dichotomous items instead of Cronbach’s alpha | Should be at least 0.70 or higher to retain an item in an &quot;adequate&quot; scale; and many researchers require a cut-off of 0.80 for a &quot;good scale.&quot; |</p>
<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parallel or Alternate Forms</td>
<td>Refers to the extent to which two different versions of the same questionnaire yield the same score when administered to the same people. The two versions of the instrument may be administered sequentially in one testing session of at different times.</td>
<td>Generally values of at least 0.80 are considered acceptable</td>
</tr>
<tr>
<td>Intra-rater Reliability</td>
<td>Refers to the extent to which one researcher scores the same data in the same way two or more times.</td>
<td>Generally values of 0.70 or higher are considered acceptable</td>
</tr>
<tr>
<td>Inter-rater Reliability</td>
<td>Refers to the extent to which two or more researchers independently score the same data in the same way.</td>
<td>Generally values of 0.70 or higher are considered acceptable</td>
</tr>
</tbody>
</table>
# Stability Reliability

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test-retest</td>
<td>Administration of a questionnaire to the same research participants two different times.</td>
<td>Generally values of $\geq 0.80$ is considered as a reliable determinant</td>
</tr>
</tbody>
</table>
Finding Research Instruments

- **Online Instrument Indexes**
  - HaPI (Health and Psychosocial Instruments)
  - Mental Measurement Yearbooks
  - Tests in Print

- **Full Text Instruments in Books**
  - The Measurement of Nursing Outcomes Series (3 Volumes)
    - Volume 1: Measuring Nursing Performance in Practice, Education, and Research
    - Volume 2: Client Outcomes and Quality of Care
    - Volume 3: Self Care and Coping
  - Instruments for Clinical Health Care Research (Marilyn Frank-Stromborg, Sharon J. Olsen)
  - Assessing & Measuring Caring in Nursing and Health Sciences (Jean Watson)
  - Measurements for Long-term Care: A Guidebook for Nurses (Sarah R. Beaton, Susan A. Voge)

- **Articles**

- **Dissertations and Theses**
Article Resources


Questions