FCE 3900
EDUCATIONAL RESEARCH
LECTURE 9
INSTRUMENT, VALIDITY & RELIABILITY
OBJECTIVE

- To understand how instruments are developed.
- To understand the concept of measurement scales.
MEASUREMENT

• Measurement: The process of assigning value or quantitative number to a variable so that the variable may be analysed

• Measurement scales are likely to produce new information of the object being measured (example: level of tension pre and post exam weeks among undergraduates of different status)

• The analytical operations that can be done on the numbers or values depend on the type of measurement used.
Nominal Scale

Categories that describe traits or characteristics that participants can check (categorical) (e.g., male/female)

Ordinal scale

Participants rank order a characteristic, trait, or attribute (categorical). The difference in scores between the ranks would not necessarily be the same. (e.g.: rank order student scores in biology)

Interval scale

Provides “continuous” response possibilities to questions with assumed equal distance on the scale (continuous) (e.g., perception of adequacy, level of attainment, temperature)

Ratio scale

A scale with a true zero and equal distances among units (continuous) (e.g., height, body measurement)
Rating Scales

4. How would you rate this product?

☐ Excellent
☐ Good
☐ Fair
☐ Poor

5. On a scale where "10" means you have a great amount of interest in a subject and "1" means you have none at all, how would you rate your interest in each of the following topics?

Domestic politics ... __
Foreign Affairs ...... __
Science & Health .. __
Business ................ __

Agreement Scale

6. How much do you agree with each of the following statements:

<table>
<thead>
<tr>
<th>Statement</th>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>My manager provides constructive criticism</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>Our medical plan provides adequate coverage</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
<tr>
<td>I would prefer to work longer hours on fewer days</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
<td>☐</td>
</tr>
</tbody>
</table>
INSTRUMENT

- Questionnaire
- self checklist
- Attitude scales
- Personality inventories
- Achievement/aptitude test
- Performance test
- Projective device (vague stimulus that allows respondents to project their anxieties, prejudices, needs etc.)
- Sociometric devices- peer rating indicating making choices amongst their peers
- Time and motion logs
- Tally sheet
GUIDE TO BUILD INSTRUMENT

- Research Objective driven.
- Must be based on operational definitions.
- Consider the type of measurement
- Ascertaining the validity
- Consider the length, difficulty, clarity of questionnaire.
- Check on item discriminant.
- Check on reliability
SUMMARY – STEP BY STEP TO DEVELOP INSTRUMENT

1. Review literature in the domain which you wish to measure (i.e., "computer attitudes").

2. Develop a list of categories (subscales [i.e., constructs]) that you wish to sample from the domain. The domain may be "Computer Attitudes" and the categories (constructs) might be "Ease of Use of Computers" and "Usefulness in Education".

3. Write 8 to 10 items/statements (operational definitions) for each category (i.e., "Computers will help students learn material faster."). Avoid common survey pitfalls when writing your statements.
SUMMARY – STEP BY STEP TO DEVELOP INSTRUMENT

4. Give the items to at least 5 experts for classification (Content Validity). The panel of experts will attempt to match the operational definitions with their appropriate categories within the domain. The statements should be mixed (in other words, don't put all the statements about a given construct together).

5. Develop an instrument with the successfully classified items. Use a Likert scale to design your instrument. You may wish to rewrite some of the items that were not successfully classified. You may remove some of the original items that the content experts found difficult to categorize.
6. Run a factor analysis (exploratory) on the field test responses. More advanced students may wish to do a confirmatory factory analysis.

7. Name each factor (category) based on the items which loaded on it (> .40)

8. Review whether each item conceptually belongs with its factor (subscale) and remove those which do not.

9. Run Cronbach's Alpha Reliability for each factor/category (subscale) to investigate internal consistency reliability.

10. Modify and retest the instrument if necessary (alpha < .70).
Here are some specific things to **avoid:**

**double-barrelled questions,** for example: **Do you like school and do you study regularly?**

**An established pattern of positively or negatively constructed questions** which are targeting opposite values. There is a tendency for respondents to want to comply, to agree. This is one reason questionnaires should feature a balance between positively and negatively constructed questions,

**leading or suggestive questions,** such as: **If you agree that education is important, would you spend 30%-40% of your pay on education?**
Here are some specific things to **avoid:**

**ideologically-loaded expressions**, called *red flag words* (using terms such as ‘communist government’ to elicit negative responses which are then interpreted as indicating negative feelings about socialistic values)

**long, complex questions**; aim at short, clear and direct questions

putting the **main idea at the beginning** of the question; respondents will tend to expect *end-focus* of the key information you are seeking
LOCATING OR DEVELOPING AN INSTRUMENT FOR DATA COLLECTION

- Look in published journal articles
- Run an ERIC search and a descriptor for the instrument you want in an online search to see if there are articles that contain instruments
- Search Internet
- Develop your own instrument
VALIDITY

• Validity: appropriateness, meaningfulness and usefulness of the inferences a researcher makes from the results obtained from the instrument.

• Reliability: consistency of answers or scores from one administration of the instrument to another.
VALIDITY

• Researchers want information they obtain through the use of an instrument to serve their research objectives. They want to be able to make meaningful, useful and correct inferences from the data collected.

• It is the validity of the inferences from the data collected through the use of the instrument that is at stake, not the instrument itself.
VALIDITY

• Validity depends on the amount and type of evidence there is to support the interpretation researchers wish to make concerning the data.
TYPES OF VALIDITY

• Content related validity (content validity) : refers to content and format of instrument. Content and format must be consistent with the operational definition of the variables and sample of subjects.
• E.g. how many questions and what type of questions would you give to test on students’ understanding of mathematical probability.
• Criterion validity: refers to the relationship between scores obtained using the instrument and scores obtained using another or more other instruments. How strong is the relationship between music aptitude score and course grades in orchestral performance?
• Construct validity: the nature of the psychological construct being measured by the instrument. How well does a measure of the construct explain differences in the behavior of individuals?
• Predictions and constructs made from theory
• Three steps in obtaining construct validity: (i) clarify operational definition
• (ii) formulation of hypotheses: (people with low scores will behave/perform differently from those with higher scores)
• Test of hypotheses
RELIABILITY

• Refers to consistency of the scores obtained from one administration of the instrument to another and from one set of items to another.
• Reliable instruments MAY NOT be valid.
• Both reliability and validity may change according to context.
• We need to target for high reliability and high validity.
ERRORS IN RELIABILITY OF MEASUREMENT

• When estimating the reliability of instrument, errors may occur when:
• An instrument is administered to the same group more than once. The longer the time interval, the lower the reliability coefficient. Therefore test-re-test method to measure reliability must happen between 2-3 month period.
ERROR IN RELIABILITY OF MEASUREMENT

• When two different forms of an instrument are used, or from one part of an instrument to another (equivalent form method). If the two sets of scores are high, then the two forms are reliable, measuring the same thing. The problem arises when one set of score is higher than the other.
ANALYSING RELIABILITY OF MEASUREMENT

- Internal consistency method for single administration of an instrument
- Split half procedure: get correlation between two halves of a test (Spearman Brown prophecy), then calculate reliability
- Alpha coefficient
<table>
<thead>
<tr>
<th>Cronbach's Alpha Based on Standardized Items</th>
<th>N of Items</th>
<th>Cronbach's Alpha</th>
<th>Scale Mean if Item Deleted</th>
<th>Scale Variance if Item Deleted</th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach's Alpha if Item Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>.769</td>
<td>.805</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Item</th>
<th>Scale Mean if Item Deleted</th>
<th>Scale Variance if Item Deleted</th>
<th>Corrected Item-Total Correlation</th>
<th>Squared Multiple Correlation</th>
<th>Cronbach's Alpha if Item Deleted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val1</td>
<td>74.80</td>
<td>49.065</td>
<td>.509</td>
<td>.576</td>
<td>.751</td>
</tr>
<tr>
<td>Val2</td>
<td>75.01</td>
<td>49.416</td>
<td>.300</td>
<td>.254</td>
<td>.762</td>
</tr>
<tr>
<td>Val3</td>
<td>74.88</td>
<td>49.405</td>
<td>.390</td>
<td>.404</td>
<td>.757</td>
</tr>
<tr>
<td>Org4</td>
<td>74.78</td>
<td>49.012</td>
<td>.567</td>
<td>.624</td>
<td>.750</td>
</tr>
<tr>
<td>Org5</td>
<td>74.92</td>
<td>48.992</td>
<td>.410</td>
<td>.301</td>
<td>.755</td>
</tr>
<tr>
<td>Org6</td>
<td>75.15</td>
<td>47.867</td>
<td>.414</td>
<td>.216</td>
<td>.754</td>
</tr>
<tr>
<td>UpS7</td>
<td>75.12</td>
<td>48.839</td>
<td>.389</td>
<td>.470</td>
<td>.756</td>
</tr>
<tr>
<td>UpS8</td>
<td>75.07</td>
<td>48.747</td>
<td>.434</td>
<td>.500</td>
<td>.754</td>
</tr>
<tr>
<td>UpS9R</td>
<td>76.96</td>
<td>53.438</td>
<td>-.042</td>
<td>.134</td>
<td>.796</td>
</tr>
<tr>
<td>Soc10</td>
<td>75.18</td>
<td>47.868</td>
<td>.450</td>
<td>.325</td>
<td>.751</td>
</tr>
<tr>
<td>Soc11</td>
<td>75.69</td>
<td>47.795</td>
<td>.373</td>
<td>.326</td>
<td>.757</td>
</tr>
<tr>
<td>Soc12</td>
<td>75.23</td>
<td>47.544</td>
<td>.482</td>
<td>.383</td>
<td>.749</td>
</tr>
<tr>
<td>Conv13</td>
<td>75.05</td>
<td>48.146</td>
<td>.477</td>
<td>.312</td>
<td>.750</td>
</tr>
<tr>
<td>Conv14</td>
<td>74.83</td>
<td>49.052</td>
<td>.514</td>
<td>.419</td>
<td>.751</td>
</tr>
<tr>
<td>Conv15</td>
<td>75.20</td>
<td>49.173</td>
<td>.338</td>
<td>.181</td>
<td>.760</td>
</tr>
<tr>
<td>Conv16</td>
<td>74.87</td>
<td>49.332</td>
<td>.444</td>
<td>.363</td>
<td>.754</td>
</tr>
<tr>
<td>Earn17</td>
<td>75.83</td>
<td>50.126</td>
<td>.194</td>
<td>.347</td>
<td>.772</td>
</tr>
<tr>
<td>Earn18</td>
<td>75.86</td>
<td>50.250</td>
<td>.181</td>
<td>.121</td>
<td>.774</td>
</tr>
<tr>
<td>Earn19</td>
<td>75.85</td>
<td>50.616</td>
<td>.159</td>
<td>.357</td>
<td>.775</td>
</tr>
</tbody>
</table>