Experimental Measures and Non-Experimental Quantitative Research Design

Katie Rommel-Esham
Education 504

Criterion-Referenced Interpretation

- Individual's scores are compared to a predetermined standard of performance (the criterion), not with the scores of others
- Score is usually expressed as a percentage or pass/fail
- Focuses on what individuals are able to do
- Most result in highly skewed distribution*
- Good for diagnosis

Norm-Referenced Interpretation

- Scores indicate how an individual compares with the norming population (i.e. individuals' scores are compared with the scores of others)
- Interpretation relies on relative standing with less emphasis on absolute amount of knowledge or skill
- Instrument needs to be able to make distinctions among individuals
Norm-Referenced Interpretation (con’t)

- Extremely easy and extremely difficult items are not included, though items tend toward the difficult end of the spectrum
- Score is generally expressed as a standard score or grade equivalent
- Need to look very carefully at the norming sample when interpreting results

Paper-and Pencil Tests

- A situation in which a standard set of questions is presented to each subject.
- The focus is on some cognitive task: what the person knows (achievement), is able to learn (ability or aptitude), chooses or selects (interests, attitudes, or values), or is able to do (skills).
- All tests measure current performance.

Standardized Tests

- Provide uniform procedures for administration and scoring, which may include timing, a script, specified testing conditions, materials allowed, whether or not students’ questions may be answered
- Scoring is usually objective
- Most have been given to some norming group
- Most are prepared commercially which helps to provide some degree of insurance regarding norms, reliability, and validity
- May not be specific enough to provide as sensitive a measure as is needed (as might be provided by a diagnostic instrument).
Achievement Tests

- Measure present knowledge and skills of a sample of relevant content
- Have restricted coverage, and are generally tied closely to school subjects
- Emphasis is on recent learning
- Types include diagnostic, survey battery, single subject achievement
- May be either norm- or criterion-referenced, though most are norm-referenced

Standards-Based Tests

- Standardized achievement tests with criterion-referenced interpretations
- Based on established standards
- Influenced by what students learn in school as well as by what they learn at home and in the community
- Students are generally judged as “proficient” or “non-proficient”
- Generally high stakes tests (for promotion, graduation, accreditation, for example)

Aptitude Tests

- Used to predict future performance on a criterion prior to instruction, placement, or training
- The test itself, not individual items, is considered to be predictive
- The terms “intelligence test” and “ability test” are often used interchangeably with aptitude
Standard Scores

Most common is the z-score:

\[ z = \frac{x_i - \bar{x}}{\sigma} \]

where \( x_i \) is an individual score, \( \bar{x} \) is the mean, and \( \sigma \) is the standard deviation

z- Scores

- The z-score provides an indication of how far a given score is from the mean in standard deviation units
- It allows comparisons across non-equivalent testing situations
- Scores below the mean will have a negative z-score, while those above the mean will have a positive z-score
- A z-score is not a score that would be recorded in a grade book, however it provides useful information

Measures of Noncognitive Traits

Personality, Attitude, Value, and Interest Inventories
Questionnaires

- Most widely used technique to gather information from subjects
- They are economical, anonymous, and consistent
- Subjects respond to something written

Guidelines for Questionnaires

- Items should be clear
- Questions should be limited to a single idea or concept (not “double barreled”)
- Respondents should be competent to answer (refers to timing, area of expertise, etc.)
- Questions should be relevant
- Items should be short and simple
- There should be no negative items
- There should be no biased items

Types of Items Used on Questionnaires

- Open Form
  - Subjects write in whatever response they choose
  - Exert little control over the responses
  - Often used in the development phase to determine responses for a closed form instrument

- Closed Form
  - Subjects choose from among pre-determined responses
  - May limit accuracy and variability of responses
Scaled Questionnaire Items

**Likert Scale**

- Question or statement followed by a scale of potential responses
- May be tailored to fit the nature of the question, but may also be misleading
- Science is very important

<table>
<thead>
<tr>
<th>strongly agree</th>
<th>agree</th>
<th>neutral</th>
<th>disagree</th>
<th>strongly disagree</th>
</tr>
</thead>
</table>

Scaled Questionnaire Items

**Semantic Differential**

- Variation of the *Likert Scale*
- Uses adjective pairs, each of which is used as an anchor for a continuum
  
  Math  
  
  Like ___ ___ ___ ___ ___ Dislike  
  Easy ___ ___ ___ ___ ___ Tough  

Ranked Questionnaire Items

- Allows for respondents to prioritize
- For example, a respondent might indicate that five items are "very important," which is of limited usefulness without a rank ordering of the items
Checklists as Questionnaire Items

- Provides respondents with a number of items from which to choose
- May also be used to ask students to reply yes/no, or choose the category to which they belong
- For categorical responses, respondents can be placed in exactly one category

Problems with Questionnaires

- Response set: includes selecting all positive responses or all negative responses (regardless of content), guessing, and sacrificing speed for accuracy
- Faking: includes responding in socially desirable ways
- Reliability is generally lower than in cognitive tests
- Construct validity is difficult to establish
- Because there are generally not “correct” answers, the nature of the comparison group is particularly important

Observation Instruments

- Rely on a researcher seeing and hearing things and recording these observations rather than relying on subjects’ self-report responses to questions or statements
- In quantitative research, the observer acts as a complete observer
- Observations may be high inference or low inference
Problems with Observations: *Observer Effects*
- *Observer Bias*: due to preconceived notions of the observer
- *Contamination*: result of observer's knowledge of the purpose of the study
- *Halo Effect*: based on initial impressions

Interviews as “Instruments”
- Interviews are essentially vocal questionnaires
- They are flexible and adaptable, and may be used with those who are not capable of completing a questionnaire (those who are illiterate or too young to read and write, for example)
- Responses may be followed up, questions clarified
- Result in a higher response rate than written questionnaires

Disadvantages of Interviews
- Potential for subjectivity and bias
- Costly
- Time consuming
- Lack of anonymity
- Respondent may be uncomfortable and not relay accurate information
- Interviewer may ask leading questions
- Sample may be smaller (resources)
Types of Interview Questions

- **Structured**: “Would you say the program has been highly effective, somewhat effective or not effective at all?”
- **Semi-structured**: “What has been the most effective aspect of your teacher development program?”
- **Unstructured**: “Tell me about your mentoring program.”
- **Leading**: “Given the expense of the new reading program, should we make the adoption this year?”

Interviewer Effects

- **Bias**
- **Contamination**
- **Halo Effect**

Unobtrusive Measures

- A type of measure in which participants are asked or required to do nothing out of the ordinary
- Provide data that are uninfluenced by an awareness of the subjects that they are participants
- Include things like physical traces, including worn floors, books, or computers; also documents, letters, and reports