Assessment, Evaluation, Testing, & Statistics: A Brief Overview

LESSEP 2016
Florida International University
Miami, FL

Steven J. Condly, PhD
United States Military Academy at West Point
scondly@gmail.com

1

Definitions

- Measurement (*What do you know?*)
  - Assigning numbers to things, events, people, actions, etc.
- Assessment (*How do you know?*)
  - Measurements, actions, processes, data that answer the question.
- Evaluation (*How are we doing?*)
  - Comparing results and observations with goals and objectives (implied or otherwise).

2

Methods of Assessment

- Assessment is data collection and analysis for the purpose of demonstrating/verifying
- Really only two ways to do it:
  - Question
  - Observe
- Raw data have to be processed (statistics)

3

Upon Construction and Use of an LO

- This was a good (bad) LO
- Assessment: How do you know?
  - They told me they liked (hated) it.
  - They did really well (poorly) on the quiz.
  - There was a lot of discussion in class on loops afterwards.
  - Marked (Slight) improvement from pre- to post-test.
  - Experts approve (disapprove)
  - It covers (fails to cover) the basics of the information

4

Upon Construction and Use of an LO

- This LO was worth it (not worth it).
- Evaluation: Compared to what standard?
  - It was free.
  - It didn't take much time to construct.
  - We didn't get the kind of gain we were hoping for or expecting
- And the implications of the evaluation are?

5

Research Design

- Get help from the Internal Evaluator, Debra Davis
- Get help from education or psychology professors at your institution

6

Testing

- According to instructional design theory, you start curricular design from the *tests* (i.e., you teach to the test)
- Upon interaction with a given LO, the SWBAT….
- Make it behavioral, not cognitive
  - SWBAT trace code
  - SWBAT enjoy programming

7

Testing

- Test items, collectively, should maximize variance (i.e., distinguish between students of differing ability)
- The more advanced the subject, the more difficult the items should become
  - e.g., All options are correct; choose the most correct
- No surprises on the test (in terms of content and format)

8

Statistics

- Select a good comparison criterion or group
- Standard statistical techniques **are** acceptable for Likert-scaled survey data
- **Don't** use NHST ($p \leq .05$)
  - Strongly influenced by sample size
    - Small p does not necessarily indicate a stronger relationship or effect, or practical significance
  - What people think it is: $P(H_0=0|\text{sample})$
  - What it actually is: $P(\text{sample}|H_0=0)$
  - How much *there* is there?

9

Effect Size Statistics

- For Likert or interval-level data, when comparing two groups, use Cohen's *d*
  - $M_1 - M_2 / [(s_1 + s_2) / 2]$
- For ordinal data, when comparing two groups, use Probability of Superiority
  - $MWU / (n_1 n_2)$
- For correlations between two groups, use r^2
  - $(r)\ (r)$ x 100 gives % of variance explained

10

Websites

- http://oerl.sri.com/ccli_resources.html
- www.socialresearchmethods.net/kb/contents.php
- http://www.uccs.edu/~lbecker/
- Grissom, R. J. (1994). The probability of the superior outcome of one treatment over another. *Journal of Applied Psychology, 79*(2), 314-316.