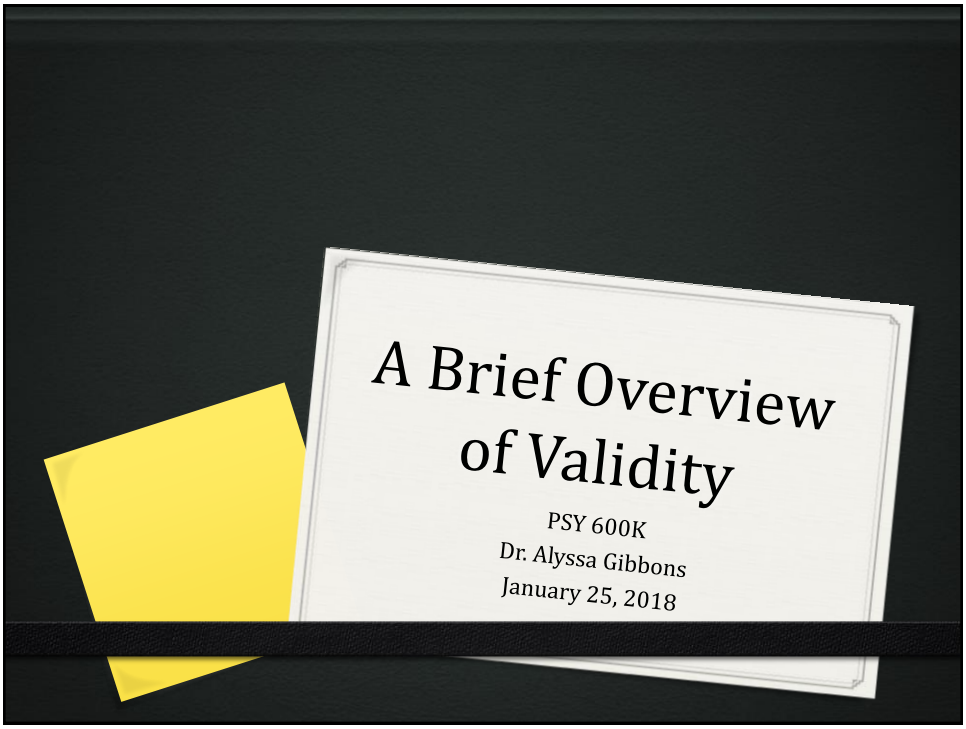


Agenda

- A general definition of validity
- The traditional view
  - 2 key papers
  - Content, criterion, & construct validity
- The modern view
  - 5 kinds of evidence
  - Validity as a unitary concept
- Why are we talking about this now?

Validity

- Lots of different definitions out there!
- Current best accepted definition:
  - "the degree to which accumulated evidence and theory support specific interpretations of test scores entailed by proposed uses of a test"

    - *Standards for Educational & Psychological Testing*

    AERA et al., 1999, p. 184
- A test is not valid or invalid in itself.
  - The inferences we make about the test scores may be more or less valid.
  - Validity can change depending on the purpose and the situation.

Validity & Purpose

- Does that make validity subjective?
- How could a test be valid in one situation and not valid in another?

- Things to consider:
  - Population
  - Inference made about the person from the test score
  - Intended use of that inference

Validity Over Time

- We think about validity differently now than we used to.
- But there is so much history behind the old view that you will still see it everywhere.
  - Some people still use it.
  - Others retain the language even when they support the modern view.
- When measurement researchers first became concerned with validity, they defined it in one way:
  - **Predicting stuff.**
  - A valid measure is one that predicts an outcome.

2 Key Papers

- Cronbach & Meehl (1955):
  - Prediction isn't the only purpose of measurement.
  - Sometimes we are interested in things for which there is no appropriate criterion (e.g., personality).
  - Suggested other ways to support interpretations of test data.
- Campbell & Fiske (1959):
  - Outlined the nomological net – relationships with other theoretically relevant constructs as evidence for particular interpretations.

The Traditional View

- Three types of validity:
  - Content
  - Criterion
  - Construct
- Content Validity = how representative are the items of the construct domain?
  - Established through logic and (usually) subject matter experts.
  - Do the items really reflect the construct as it is defined in this particular case?
  - A priori, before any data is collected.

Criterion & Construct Validity

- Criterion Validity = does the scale correlate with some outcome measure?
  - Can be *predictive* or *concurrent*
- Construct Validity = does the scale correlate as we expect it to with other relevant constructs?
  - *Convergent* validity = high correlations with constructs that should be similar.
  - *Discriminant* validity = low or zero correlations with constructs that should not be similar… OR moderate relationships with constructs that are not exactly the same…

The Modern View

- Arose in the late 1980s – 1990s
- Proponents: Messick (1989, 1995); Binning & Barrett (1989), others.
- Adopted & encouraged in the *Standards for Educational & Psychological Testing.*
  - American Educational Research Association
  - American Psychological Association
  - National Council on Measurement in Education
- Also promoted in the *SIOP Principles*.

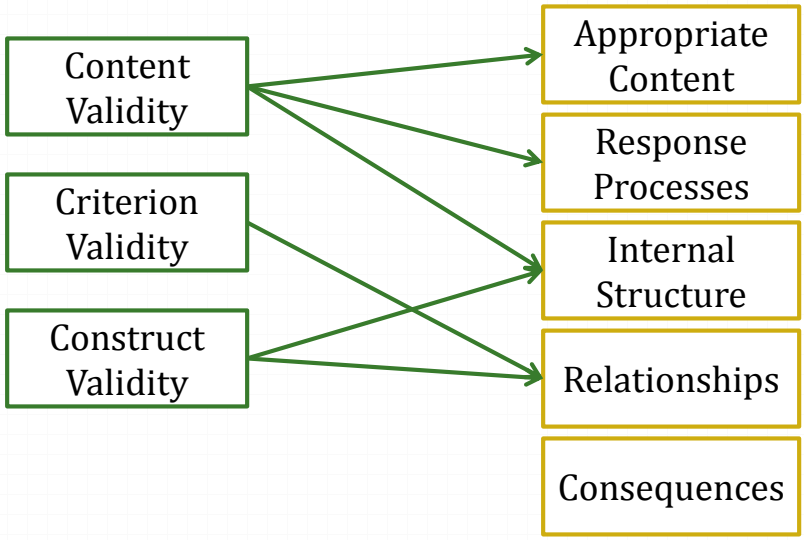
Traditional vs. Modern Views

- Traditional view implied "pick one" – if you had one of the three types, you were done.
  - Which type you focused on (sometimes) depended on your purpose.
- Modern view: that doesn't make sense.
  - All types of validity evidence ought to point to the same conclusion.
  - All types of evidence really address the same question:
    - Can I justify making this inference from these test scores?
  - There are other types of evidence that are relevant, but don't get the attention they should.

Modern View

- Five major categories of evidence:
  - Relationships with other variables
  - Appropriate content
  - Internal structure
  - Response processes
  - Consequences of testing
- "It is not the case that each of these five sources is an alternative approach to establishing job relatedness. Rather, each provides information that may be highly relevant to some proposed interpretations of scores, and less relevant, or even irrelevant to others."

Comparison



1. Appropriate Content

- Are the test items representative of the construct domain?
- Essentially identical to traditional view of content validity.
  - Based on **expert judgment** (and by experts, we don't just mean the test developers!).
  - Need people who understand the theoretical domain of the construct – what it is and what it isn't.

2. Response Processes

- Are test takers doing the things we expect them to do when answering the questions?
- Ruling out alternative explanations for test scores:
  - Knowledge of history vs. knowledge of test-taking strategies.
  - Skill in troubleshooting vs. eventually reaching the right answer.
- How do we get this information?
- Provides great information about how respondents view the construct.

3. Internal Structure

- Do the **items** relate to one another in the way we expect them to?
  - Does the empirical structure of the measure match the theoretical structure of the construct?
- Used to be considered under construct validity.
- Especially important for multidimensional constructs if you will be using the dimension scores.

4. Relationships with Other Variables

- Does the test as a whole relate to other constructs in the way we theoretically expect it to?
- Encompasses some aspects of *both* criterion & construct approaches.
  - Acknowledges that a correlation with a relevant variable is a correlation with a relevant variable, no matter what kind of variable it is.

5. Consequences

- What are the benefits of using the test, to users, test takers, and society? What are the costs?
- How can we assess this?
- Practically, often focused on issues of bias / group differences.
  - But there are other issues too.
- Is this really evidence about *validity*?

Validity is Unitary

- So have we gone from 3 types to 5 types?
- *SIOP Principles:* "different sources of evidence contribut[e] to an understanding of the inferences that can be drawn from a selection procedure."
- Sources should converge…
- … they may also provide information about the limits of our interpretations.
  - Example: employee engagement.
- You may not need all 5 types in a particular scenario.
- But over time, we build our understanding of a measure (and a construct) by accumulating and sifting this evidence.

Why Are We Talking About This Now?

- Validity is our goal in measurement.
  - Everything we do fits into this bigger picture.
- You can't develop a test without knowing what you want to use it for and what kind(s) of evidence you need to support your argument.
  - Well, technically you *can*, but…
- Practically… thinking ahead for your project.
  - We're going to collect evidence about appropriate content, internal structure and relationships w/other variables.
  - Start thinking about what those other variables should be.
    - Big Five or… ???

Questions?

Lab Friday: Fun with Excel

For next time:
Items & Item Writing
Read: DeVellis pp. 73 – 85; Burisch (1984)
Reading Response #2 on Canvas
Work on your Project Plan!