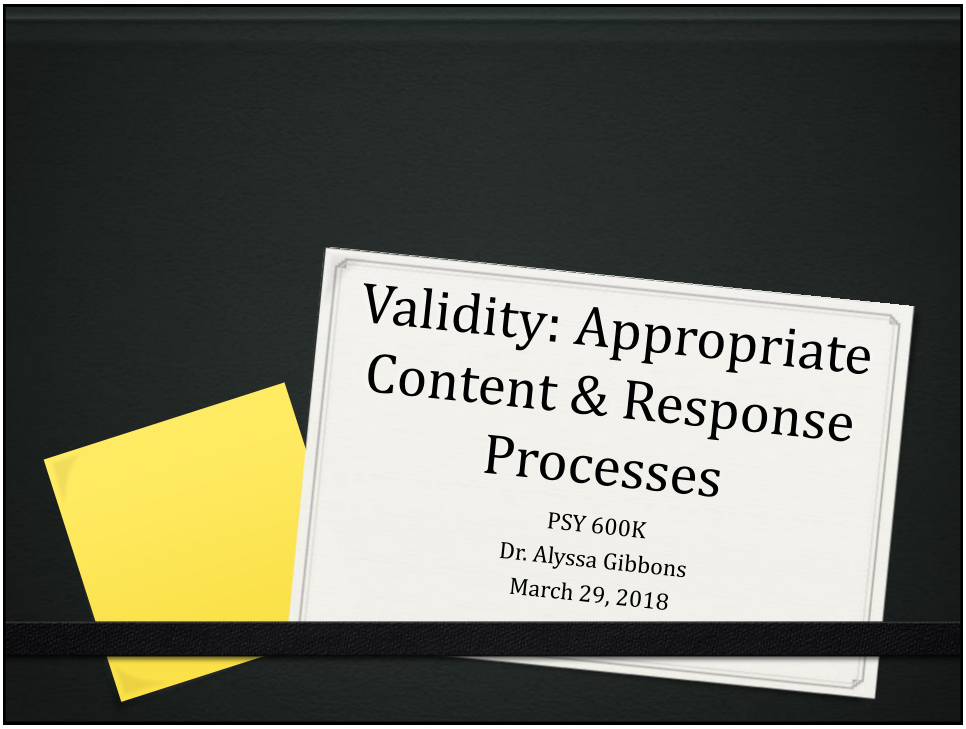


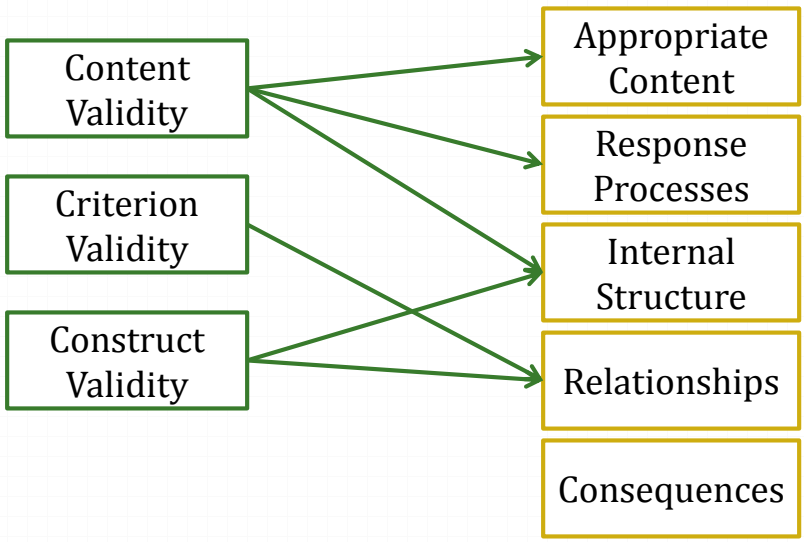
Agenda

- Getting back into validity mode.
- Evidence of appropriate content.
  - Test specifications.
  - Systematic processes.
  - External input & the content validity ratio.
  - Face validity.
- Evidence of appropriate response processes.

Quick Recap

- Validation is about **accumulating evidence** that supports the inferences we wish to make from our test scores.
- R & M point out that we want **both**:
  - **Multiple** lines of evidence that support our inferences **and**
  - Evidence that "**alternative** inferences are less well supported."
  - In other words, validation should include testing competing explanations.

Remember:



Evidence of Appropriate Content

- Goal: Items should be representative of the entire construct domain.
- Much of this evidence comes from our construct development process.
  - We still need to document it!
  - What is our definition? Why are we defining the construct this way and not another way?
  - What is the structure of the construct? How do we define the construct domain?
  - What reason do we have to believe that our items address this construct?
  - Etc.

The *Evidence* Part

- Although construct development is important, we can't really claim content validation on the strength of a lit review.
- Need some more "objective" support for our arguments about what the content should be:
  - Systematic process
  - External input
  - Or both!

Systematic Process

- In IO, content validation is usually linked to **job analysis**.
  - Careful, systematic way to identify the knowledge, skills, abilities, and other characteristics (KSAOs) needed to perform the job.
  - Can include interviews with SMEs, questionnaires, direct observations of behavior, review of job descriptions or manuals, etc.
    - All carefully documented – what information comes from where?
  - Might cover all aspects of the job or focus on just a few.
    - E.g., for a personality test we would focus on personality characteristics and exclude job knowledge.
- For knowledge tests (e.g., education, certification) content validation is often linked to curriculum.

Test Specifications / Test Blueprint

- Detailed outline of what your test should include.
  - What kinds of items should be included.
  - How many items of each kind.
    - What is most important?
    - What do you want to measure most precisely?
  - Definitions of **everything**.
  - Sometimes, what kinds of items should *not* be included.
  - Requirements about length, format, administration, etc.
  - For example:
- Allows you to justify/explain why your choices are appropriate.

Items Aren't the Only Content!

- You should also consider and justify choices about:
  - Targeted difficulty level
  - Specificity of the test content
  - Response formats and scale types
  - Reading level / type of language used
  - Instructions & administration procedures
  - Etc.
- All of these things are part of the content of the test!

External Input

- After item writing – ask knowledgeable individuals outside the research team to evaluate items.
- Appropriate SMEs depend on your purpose:
  - Job incumbents?
  - Supervisors?
  - Psychologists?

- Provide your definitions, ask SMEs to evaluate the extent to which the items address the construct and are **essential**.

Content Validity Ratio

- Lawshe (1975) proposed a way to "quantify" content validity:
- Ask SMEs to rate each item:
  - 3 = Essential, 2 = Useful but not essential, 1 = Not necessary
- Add up the number of experts who think each item is essential.
  - Ignore the "useful but not essential" items!

$$\frac{(\#\text{ of "essentials"}) - (\#\text{ of experts} / 2)}{(\#\text{ of experts} / 2)}$$

- Compare to predetermined minimum values – discard items where there isn't enough agreement that the item is essential.

Face Validity

- Content validity ≠ face validity!
- **Face validity** = degree to which test-takers perceive the test as measuring the construct.
- A test can be highly face valid but not content valid… or vice versa.
  - Examples?
- Face validity is sometimes desirable, sometimes undesirable.
  - Examples?
- Important to be sure your SMEs are not just evaluating face validity!

Response Processes

- If we want to make claims about the process test-takers use to answer items, we need to demonstrate those processes.
  - For example… customer service, troubleshooting, even personality items are all areas where we might care about process as much as the final answer.
  - Note that sometimes the final answer is all that matters!
  - Also relevant to issues of faking, cheating, etc.
    - Why?
- How can we get at this?

Response Process Evidence

- Ask the test-takers! 
  - Think-aloud protocols, debriefing interviews, room for comments on surveys.
- Ask SMEs
- Direct observation
  - Has major practical limitations.
- Experimental manipulation.
  - Administer test under differing instructions, etc.
  - Test competing hypotheses.
    - E.g., people can raise their scores by faking.

Response Process Evidence

- Inferences from test process data.
  - E.g., response times.
- Assess process steps directly.
  - E.g., computerized CPA exam, multipart writing questions, show your work, "choose-your-own-adventure" SJTs, assessment centers.

Questions?

For next time: Relationships with Criteria
Read: R & M 8.4
Reading Response #10
Lab Friday: Multiple Factor Models & More Model Comparisons
(we'll do item analysis next week)