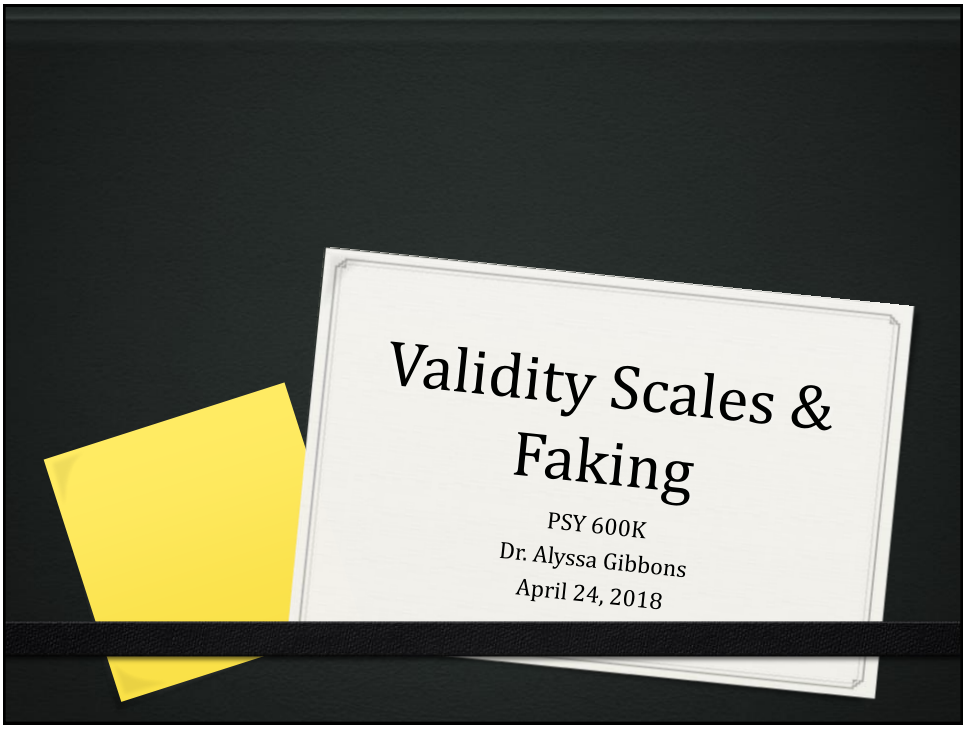

Agenda

- Reasons to distrust test scores
- Cheating
  - Particularly in online environments
- Faking
  - Is it a problem? Can people fake? Do people fake?
  - What can we do about it?
    - Social desirability (and similar) scales
- Aberrant or invalid responding

Can We Trust Our Test Scores?

- If we are relying on people to provide us information about themselves, there are many reasons it may not be advisable to take that information at face value:
  - Faking
  - Cheating
  - Lack of self-knowledge
  - Self-presentation
  - Lack of motivation
  - Misunderstanding of instructions
  - Etc.
- Can we do anything about it?

Faking & Cheating

- Both imply that the test taker is consciously doing something to obtain a test score that is not consistent with his or her true score.
  - Lots of scenarios in which this can occur – can you name some?
- Cheating usually refers to cognitive measures, faking usually refers to non-cognitive ones.
- Different processes involved:
  - Cheating requires **getting** the right answers.
  - Faking requires **identifying** the right answers.

Cheating

- Common solution: **test security**.
  - Restrict access to the test items, proctor the exams.
  - But sometimes this is not feasible, or conflicts with other priorities.
- **Adaptive** testing is another solution.
  - Use a lot of pre-knowledge about the items to tailor items to a test-taker's estimated ability level.
  - So different test-takers get different items – item pre-knowledge doesn't help so much.
- We can sometimes use IRT methods to detect aberrant response patterns.
  - i.e., getting hard questions right but missing easy questions.

Cheating Online

- One big current issue: **unproctored internet testing.**
  - Allowing test-takers to take a test from any location.
  - What pros and cons do you see to this?
- Current solution:
  - Use the unproctored test as a screen.
    - As with personality – low scores can probably be trusted, be skeptical of high ones.
  - Invite those who pass to take a proctored confirmation test.
    - Using IRT, tailored to their estimated ability level based on the first test.

Faking

- Quite a bit of controversy about how big a problem faking might be.
  - Can people fake?
  - Do they fake (outside the lab)?
  - How many people fake?
- Can people fake?
  - In the lab, absolutely.
  - But this is probably an exaggerated effect – *everyone* is trying to fake, and in real life there will be differences in motivation.
  - There are also individual differences in the ability to successfully fake.

Do People Fake?

- Common research paradigm: compare people who do and do not have an incentive to fake.
  - E.g., job applicants vs. job incumbents.
  - If applicants' scores are higher, we infer that faking occurred.
- This paradigm has produced a long history of mixed results.
  - Some groups of applicants appear to fake, some don't.
    - e.g., Hough et al. (1990) military sample.
  - Even more variability at the individual level.

Do People Fake?

- Griffith, Chmielowski, & Yoshita (2007):
  - Administered the test as part of a job application (high motivation to fake).
  - Administered again after participants had been hired, promised confidentiality (low motivation to fake).
  - At the 2nd administration, asked them to complete the test again and fake "in a way that would best guarantee that you get the job."
  - Within-participants design.
- Results:

Condition	*n*	Min	Max	Mean	SD	Alpha
Honest	60	76.17	199.00	164.92	18.35	0.83
Applicant	60	94.00	205.00	176.13	16.56	0.86
Faked	60	86.00	210.00	191.79	27.23	0.93

Implications

- It seems that at least some people do fake… but they're subtle about it.
- Quite a bit of variability – not everyone faked.
- Faking **did** change outcomes.
  - Of the top 10 applicants, only 4 would have been in the top 10 based on their honest score.
  - Depending on the selection ratio, between 31% and 66% of participants would not have been hired based on their honest scores.
  - Another study (Zickar et al., 1996) – even with only a few fakers, the fakers end up at the top of the distribution.
- Suggests that faking **is** a legitimate concern in high-stakes situations.

What Can We Do About It?

- Griffith et al. had very powerful, within-persons data.
  - Showed exactly who faked and how much.
  - Utterly impractical outside a research context!
- Drop people whose scores are "just too good"?
  - There's an obvious problem with this one…
- Develop a measure to identify fakers?
  - "Social desirability"
  - "Unlikely virtues"
  - "Lie" scales

Social Desirability Scales

- Items that (theoretically) no one can honestly "strongly agree" with. E.g.:
  - "My table manners at home are as good as when I eat out in a restaurant."
  - "I don't find it particularly difficult to get along with loud-mouthed, obnoxious people."
    - Both from the Crown-Marlowe social desirability scale.
- Idea: high scores on these measures indicate a tendency to exaggerate or deliberately manipulate your responses to give a more positive impression of yourself.
  - Which means you are also more likely to exaggerate or manipulate responses on other scales.

Correcting for Social Desirability

- Three options:
  - exclude people whose social desirability scores are too high.
  - use partial correlations to control for social desirability.
  - model social desirability directly in your regression or SEM model.
- However… all of these assume that social desirability is unrelated to your criterion (so you are controlling for irrelevant variance).
  - But SD scores are *positively* related to quite a few important criteria…
  - And controlling for SD often **reduces** criterion relationships! 
    - Piedmont et al.
  - So controlling for SD is removing ***relevant*** variance!

*Over*correcting for Social Desirability

- Schmitt & Oswald argued that correcting for faking can actually *reduce* criterion performance in the selected group.
  - If your faking measure is positively correlated with the criterion.
  - Or if your faking measure is positively correlated with the true predictor.
- They found that, overall, correcting for faking didn't affect the mean criterion performance of the selected group.
  - In other words, you didn't get better results by correcting for faking/SD.

More Issues with SD Scales

- They are themselves quite remarkably **easy to fake**.
  - Viswesvaran & Ones (1999)
- They may be **unrelated** to other operationalizations of faking behavior.
  - Peterson et al. (2011) found that SD scores:
    - Were uncorrelated with score change between an applicant and an honest condition.
    - Did pretty poorly in identifying "fakers".

What to do?

- Don't use fakable measures alone.
  - Use as a screen-out, not a screen-in technique.
    - (Trust low scores, but not high scores).
  - Get additional data.
    - Use in combination with other, less fakable predictors.
    - Verify with other-reports (not a perfect comparison, but useful and often good predictors in their own right; Piedmont et al.).

"Invalid" Responses

- What about responses that just don't make sense?
  - E.g., an uncooperative test-taker.
- Most likely won't affect the overall predictive validity coefficient or other overall relationships…
  - But does have an impact on decisions about the individual test-taker!
- IRT offers some methods for detecting statistically unlikely response patterns.
  - E.g., getting hard items right but missing easy ones.
- Piedmont et al. recommend using multiple methods (and sources) for convergent evidence about the accuracy of an individual assessment.

Bottom Line

- Ultimately, purpose becomes very important.
  - If you are making conclusions about the test or construct as a whole, faking/SD probably isn't hurting your conclusions.
  - If you are making decisions about individuals, faking **does** change outcomes for individual people.
- Social desirability measures lack construct validity!
- Finding alternatives to self-reports is a more effective way to avoid the faking problem.

Questions?

For next time: Precision of individual test scores

Practical considerations in scale development – shortening scales., replication.

Read: DeVellis p. 101-102