

A Top Ten List of Measurement-related Errors

Alan D. Mead
Illinois Institute of Technology

1: Ignoring error

- You...
 - Tested two candidates for a job, and one scored 77 and one scored 74, which would you pick?
 - Found validities for tests A and B of $r_{AC}=.05$ and $r_{BC}=.25$ with $N=30$, which would you use?
 - Recruited people for a developmental intervention if they were in the bottom 10% of a pre-test. Will your intervention successfully raise a post-test score?

2: Being too precise

- Maximizing internal consistency maximizes reliability if (and only if) errors are uncorrelated
- In many situations, you would be better off with more, shorter (less precise) measures
- Validity is probably more important than reliability, but 98% of psychometric theory is about maximizing reliability



3: Creating narrow measures

- Our tools (e.g., internal consistency reliability, factor analysis) encourage us to create unidimensional scales
- Excessive internal consistency can rob your measure of content validity
- Many criteria are multidimensional

4: Assuming R^2 goes to 1.0

- Our best predictors have validities near .50, that's only *accounting for 25%* of the variability of the criterion
 - Should we be sad?
- What's the maximum of R^2 ? What's the most variance that we can account for in the criterion?
- 1.0/100% is only the *theoretical* maximum
 - How predictable is behavior? I'd guess “not very”



5: Capitalizing on chance

- Taking a non-cross-validated statistic as Gospel
- Examples include...
 - R^2
 - Reliability of a test after selecting items
 - Validity of an empirically-keyed test
- Worse problems when...
 - You select a few elements from a large set
 - Statistics are imprecise (e.g., small N)
 - Elements vary greatly in “base rate” of “goodness”

6: Misusing NHST

- Using NHST when it's not needed
 - World is round, $p < .05$
 - Simulation studies
- Having too little power
 - What is the power to detect a true correlation of .35 with $N=100$ (and otherwise reasonably favorable conditions)? [See Schmidt, Hunter & Urry, 1976]
- “Proving” H_0
 - DIF, comparability

7: Relying on back-translation

- Carefully scrutinizing a back translation, to check that translation:
 - The flesh is willing, but the spirit is weak
 - The vodka is good, but the meat is rotten
- Can you just “carefully scrutinize” a test, rather than pilot testing?
 - No, people are not very good at this
- Also, some concepts back-translate well, but are meaningless (e.g., ice hockey)



8: Ignoring model-data fit

- There is a “scientific beauty” to statistical models that is hard to explain
- But, things turn ugly if the model fails to closely describe real-world data
 - Assumptions of the model (and robustness of the violations)
 - Identifiability of the model (can we fit our model?)
 - Global fit (e.g., AIC, GFI, Chi-Square)
 - Local fit (individual elements of the model)

9: Drinking the empiricism Kool-Aid

- Theory is more important than data
 - Data are often flawed
 - Data is always sampled with error (often substantial)
- Theory plus appropriate data (and analysis) is the strongest
 - But it's a lot less common than you might suppose
 - Most of our individual studies are either (1) very narrow or (2) flawed

10: Validating without power

- When feasible, test validation is expensive and difficult
- How big a sample do you need to validate?
 - To achieve 90% power when true validity is .35 and criterion reliability is .60, you need at least $N=276$
 - A sample of $N=105$ would only provide 50% power!
 - Multiply by 2-10 if there is moderate to severe range restriction or much less reliable criterion
- And if you do not reject H_0 , then you know essentially nothing!

Bonus: Assuming linearity

- Relationships are, often, well-approximated by linear models
- We also know that cognitive ability tests generally have linear relationships with the criterion
- However, it does not follow that all predictors will always have linear relations with criteria
 - Consider 16PF Factor G, Rule-Consciousness

Bonus: Assuming linearity

Injured good Samaritan ticketed for jaywalking

Thu Feb 26, 7:34 am ET

 Buzz Up |  Send ▼ |  Share ▼ |  Print

DENVER – A good Samaritan who helped push three people out of the path of a pickup truck before being struck and injured has gotten a strange reward for his good deed: A jaywalking ticket.

Family members said 58-year-old bus driver Jim Moffett and another man were helping two elderly women cross a busy Denver street in a snowstorm when he was hit Friday night.

Moffett suffered bleeding in the brain, broken bones, a dislocated shoulder and a possible ruptured spleen. He was in serious but stable condition Wednesday.

The Colorado State Patrol issued the citation. Trooper Ryan Sullivan said that despite Moffett's intentions, jaywalking contributed to the accident.

Moffett had been driving his bus when the two women got off. In the interest of safety, he got out and, together with another passenger, helped the ladies cross.

Moffett's stepson, Ken McDonald, said the driver of the pickup plowed into his stepfather, but not before Moffett pushed the two women out of the way.

When he awoke in intensive care, he learned of the ticket. "His reaction was dazed and confused. I was a little angry," said McDonald.

The other man also was cited for jaywalking, while the pickup driver was cited with careless driving that led to injury. Sullivan said the two elderly women haven't been cited but the investigation is ongoing.

Bonus: Assuming linearity

When he awoke in intensive care, he learned of the ticket. "His reaction was dazed and confused. I was a little angry," said McDonald.

Another man also was cited for jaywalking, while the pickup driver was cited with careless driving that caused no injury. Sullivan said the two elderly women haven't been cited but the investigation is ongoing.



Bonus: Assuming linearity

Validity of Rule-Consciousness
(simulated data)

