



THE FUNDAMENTAL ROLE OF CONSTRUCT VALIDATION IN ORIGINAL AND REPLICATED RESEARCH

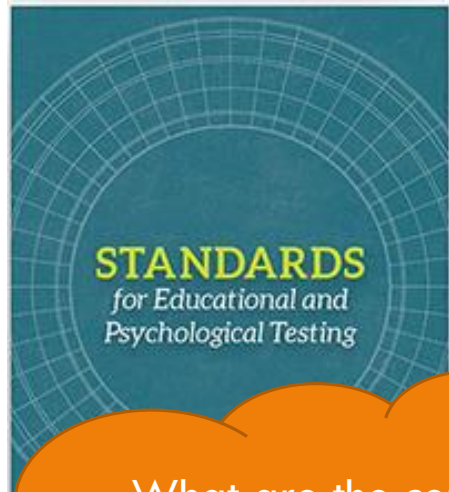
Jessica Kay Flake
McGill University
APS 2018



THE PLAN

Broad look at measurement practices in psychology and how those practices connect to current debates and discussions about replication in psychology.

BACKGROUND: CONSTRUCT VALIDATION



- Theoretical Foundation
 - What is the construct?
 - Does it exist?
 - Item content selection
 - Response processes

What are the common practices for validation in substantive research, when the focus isn't development?

Substantive

Structural

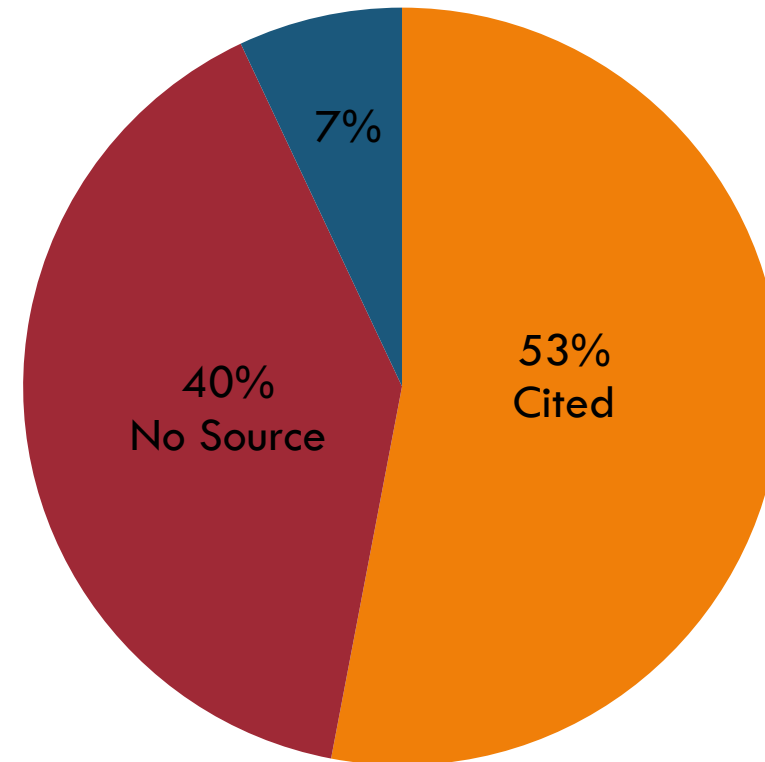
- Psychometrics
 - Item analysis
 - Measurement models
 - Reliability
 - Measurement invariance

- Nomological Net
 - Group differences
 - Predictive validity
 - Convergent and discriminant validity
 - Criterion validity

External

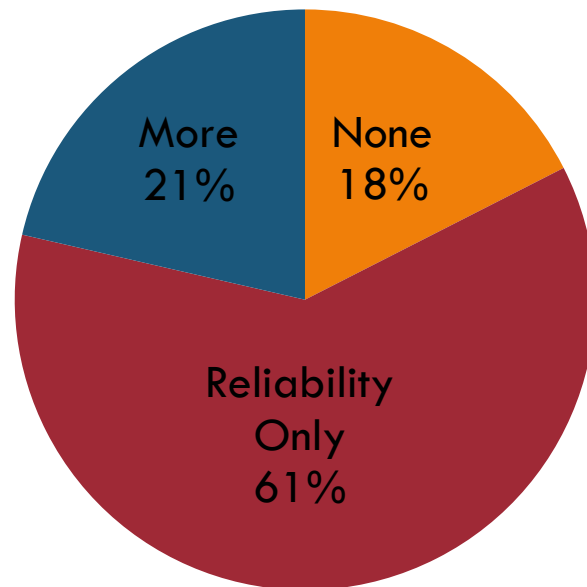
STATE OF CURRENT PRACTICE IN JPSP

Coded 35 articles (33%)
700 instances of measures
87% were item-based scales
30% of those scales were 1-item

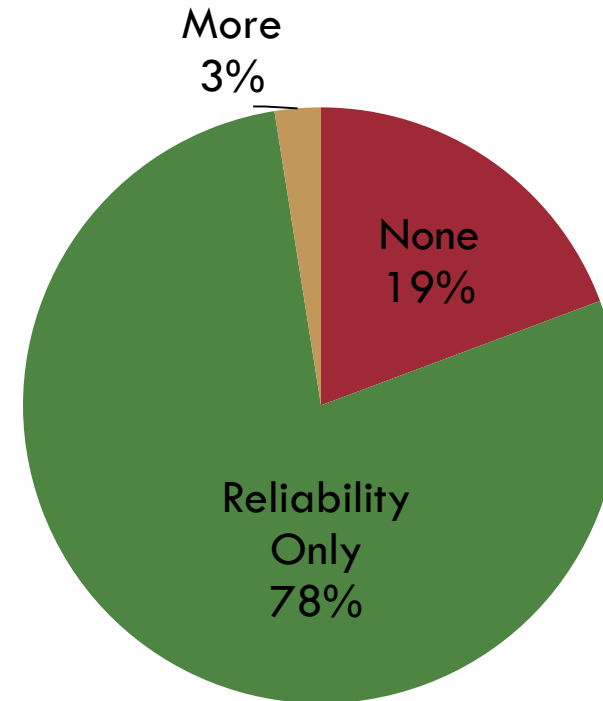


HOW MUCH CURRENT VALIDITY EVIDENCE?

Validity Evidence for Previously Developed Scales

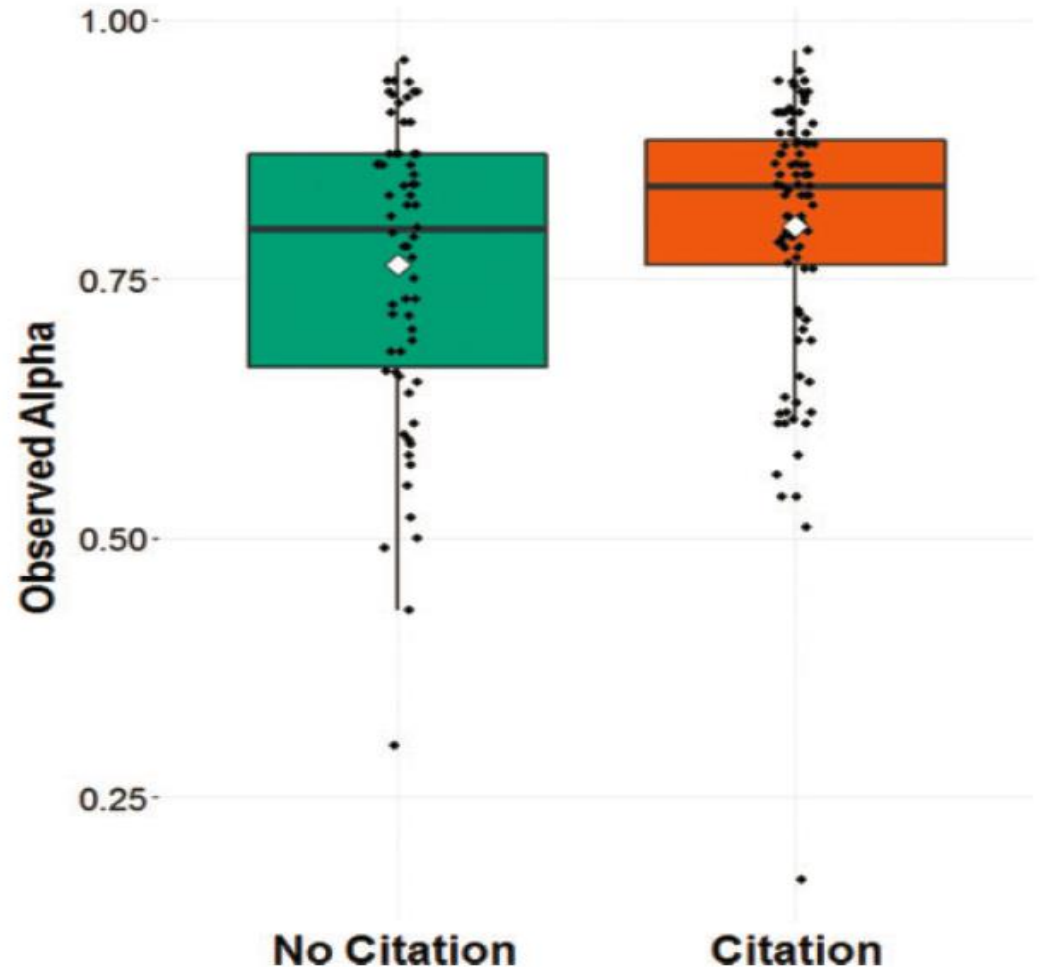


Validity Evidence for New Scales



MEASUREMENT SCHEME

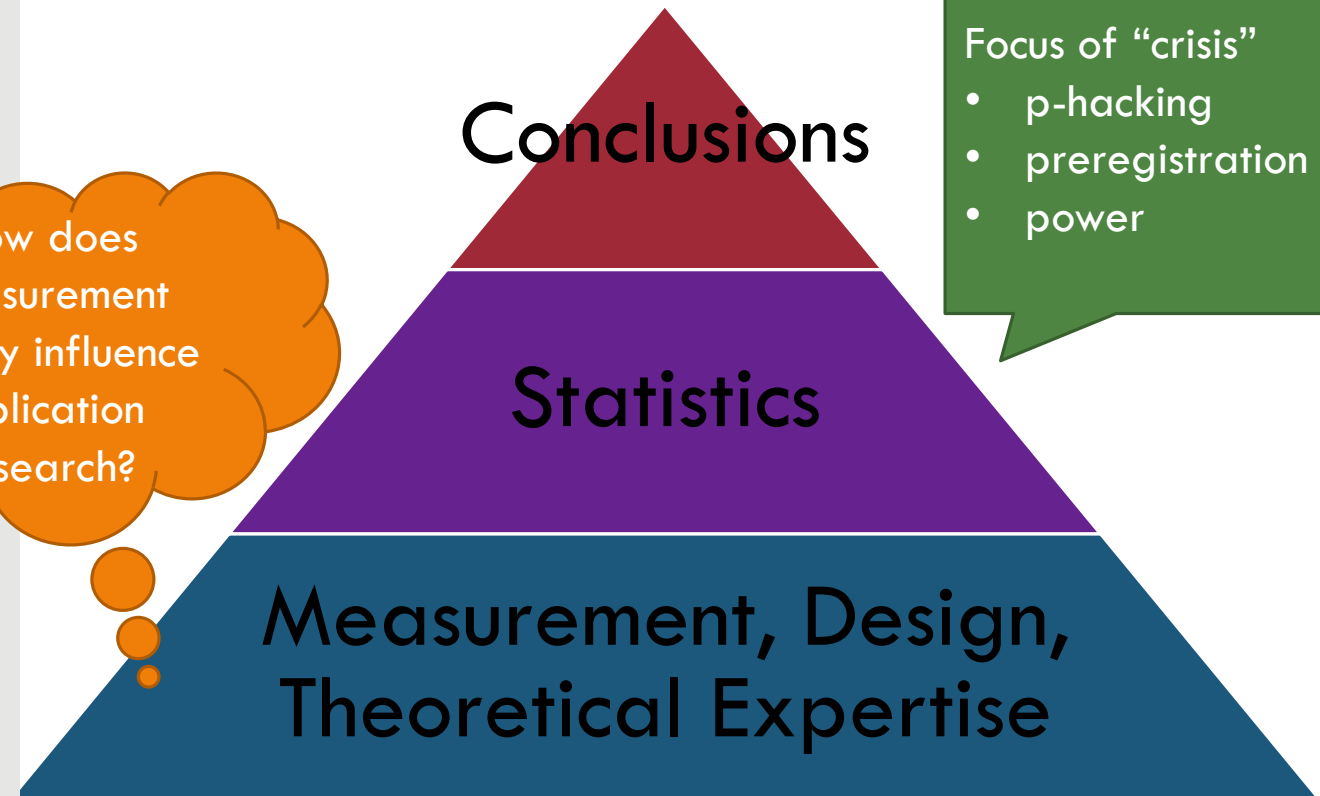
1. On-the-fly measurement is a norm
2. Alpha as the sole source of “validity” evidence is common, and often questionable
3. Cavalier item and scale changes/removal is common within studies



CONNECTING TO THE “CRISIS”



How does measurement quality influence replication research?



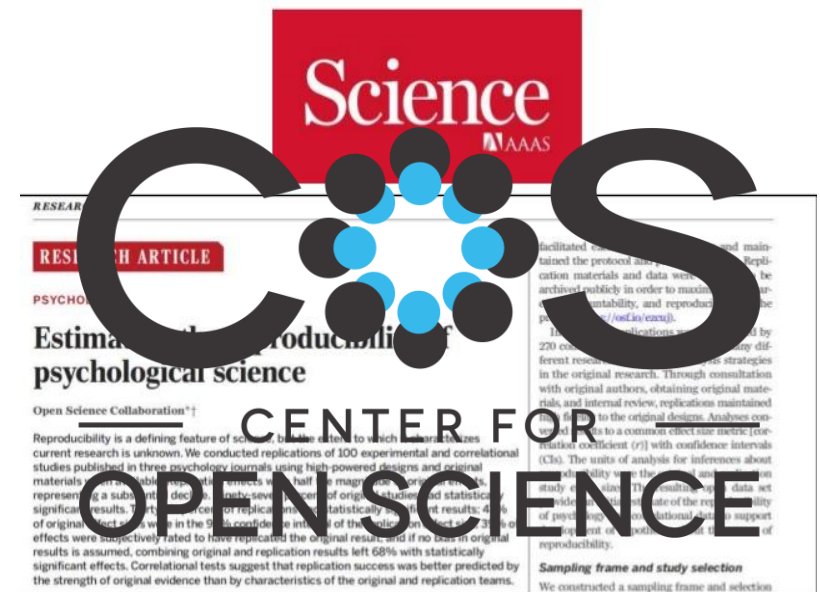
THE REPRODUCIBILITY PROJECT: PSYCHOLOGY (RPP)

100 studies taken from papers published in 2008 from Psychological Science, Journal of Experimental Psychology: General, and Journal of Personality and Social Psychology

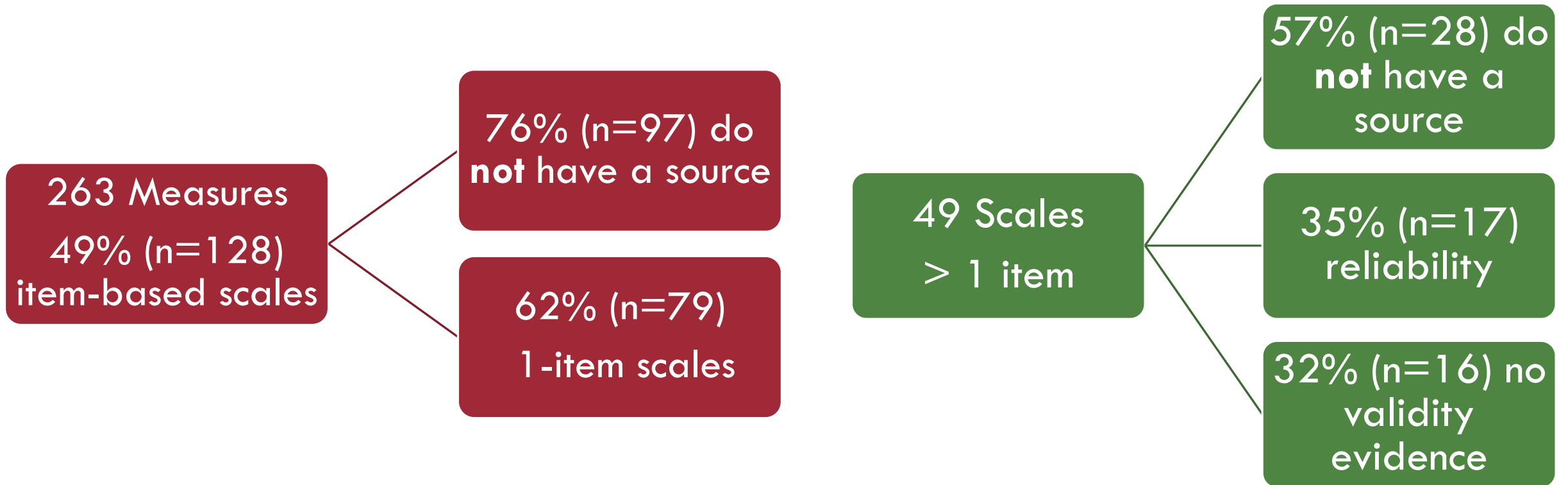
How much validity evidence in the original studies?

How much validity evidence or any clear threats to validity in the replicated studies?

On-going 82/100 studies in the analyzed dataset thus far



MEASURES AND EVIDENCE IN ORIGINALS



MEASUREMENT CHALLENGES COMPLICATE REPLICATIONS

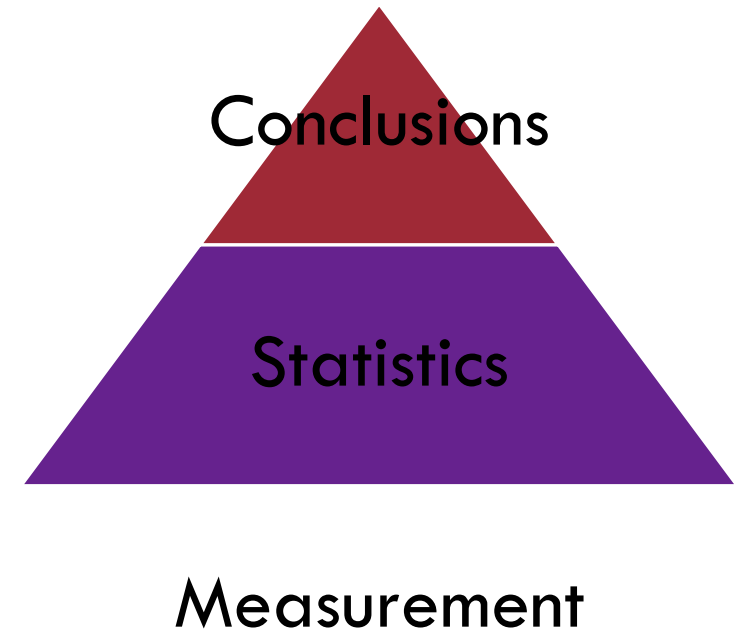
1 / 5 replications (that include surveys) encounter a measurement challenge

Scales in original study were “on-the-fly”

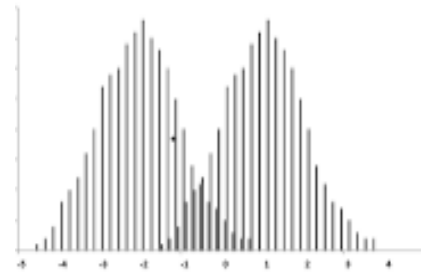
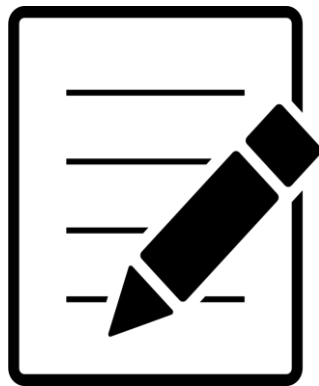
- Unexpectedly low reliability
- Adding and removing items

Even with previously ‘developed’ scales

- Breaking scales apart or aggregating differently
- Lack of translated versions
- Lack of measurement invariance



MEASUREMENT CHALLENGE EXAMPLE

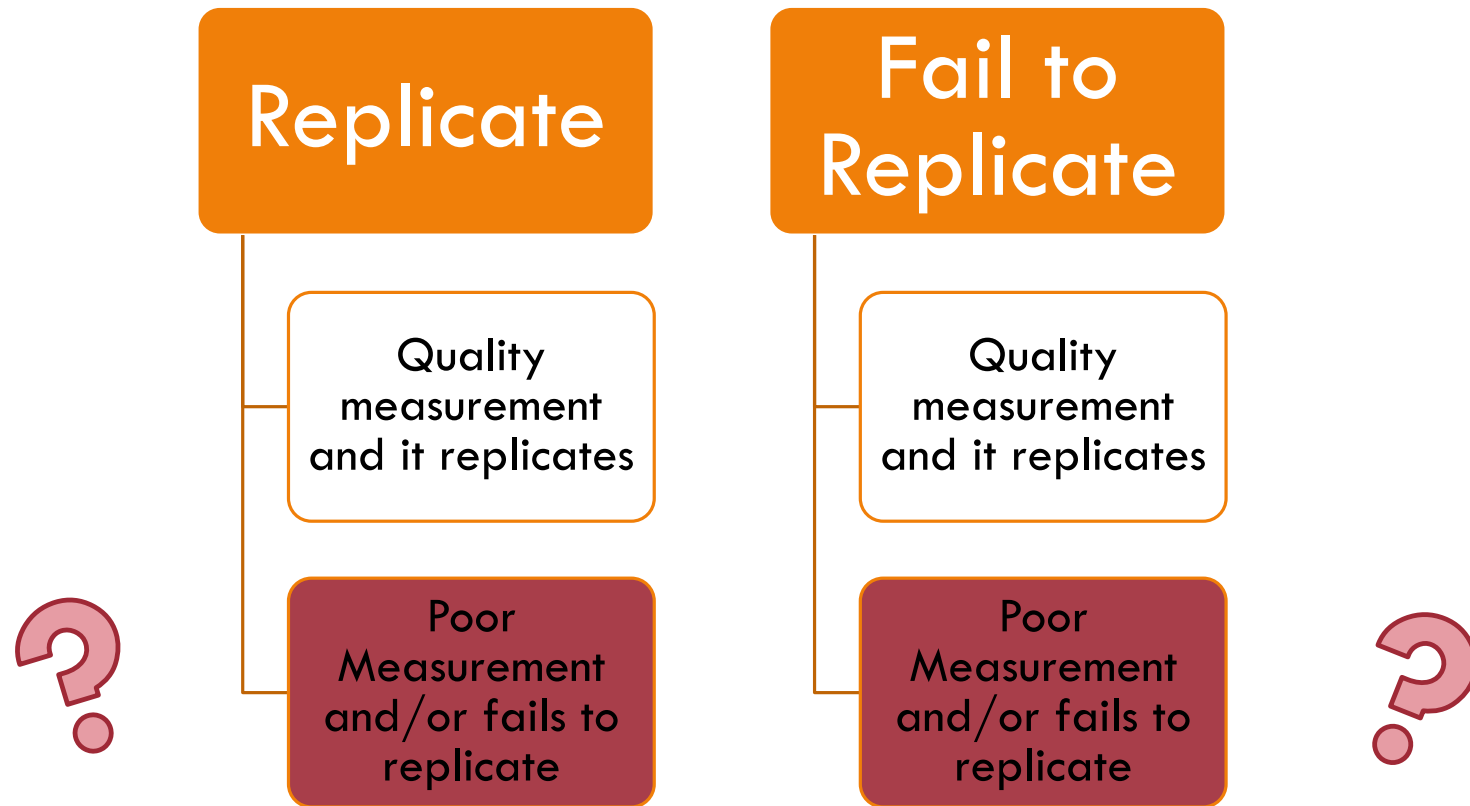


There is no effect



The scale changed

INTERPRETING THE RESULTS OF REPLICATION STUDIES WITH POOR MEASUREMENT IS BEFUDDLING



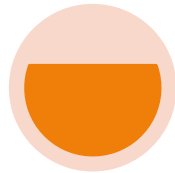
WHAT TO DO? EVALUATE THE VALIDITY EVIDENCE YOU HAVE AND PLAN ACCORDINGLY



None

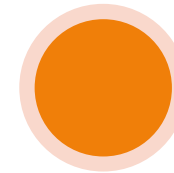
Plan a validation,
not a replication

Face
validity?



Reliability

More is needed
 α has assumptions
(e.g.,
unidimensionality),
consider if it is
appropriate.



More

Hypothesize ways
the evidence may
not apply to your
replication

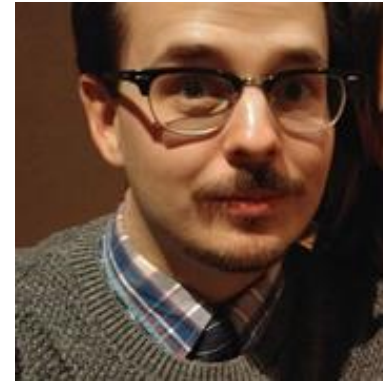
Test for invariance
if suspected
differences

Whatever the result of the replication, consider how construct validity may have contributed and interpret your results in light of the evidence you have

THANK YOU AND ANY QUESTIONS?



Eric Hehman | Ryerson University



Ian Davidson | York University



Octavia Wong | York University



Jolynn Pek | The Ohio State University

Research Assistants: Nyiesha Grant, Andrew Kim,
Lina Kanawati, Bianca Ugucioni