Reliability

Dr. Ian Hussey

Digitalisation of Psychology

What is reliability?

Reliability

Intuitive definition:

“The probability of getting the same score if you measured twice”

Reliability

Intuitive definition:

“The probability of getting the same score if you measured twice”

- Neither Internal Consistency nor Test-Retest Reliability tell you this
- As forms of correlation, they quantify the preservation of rank

Who cares! Just report its “adequate”

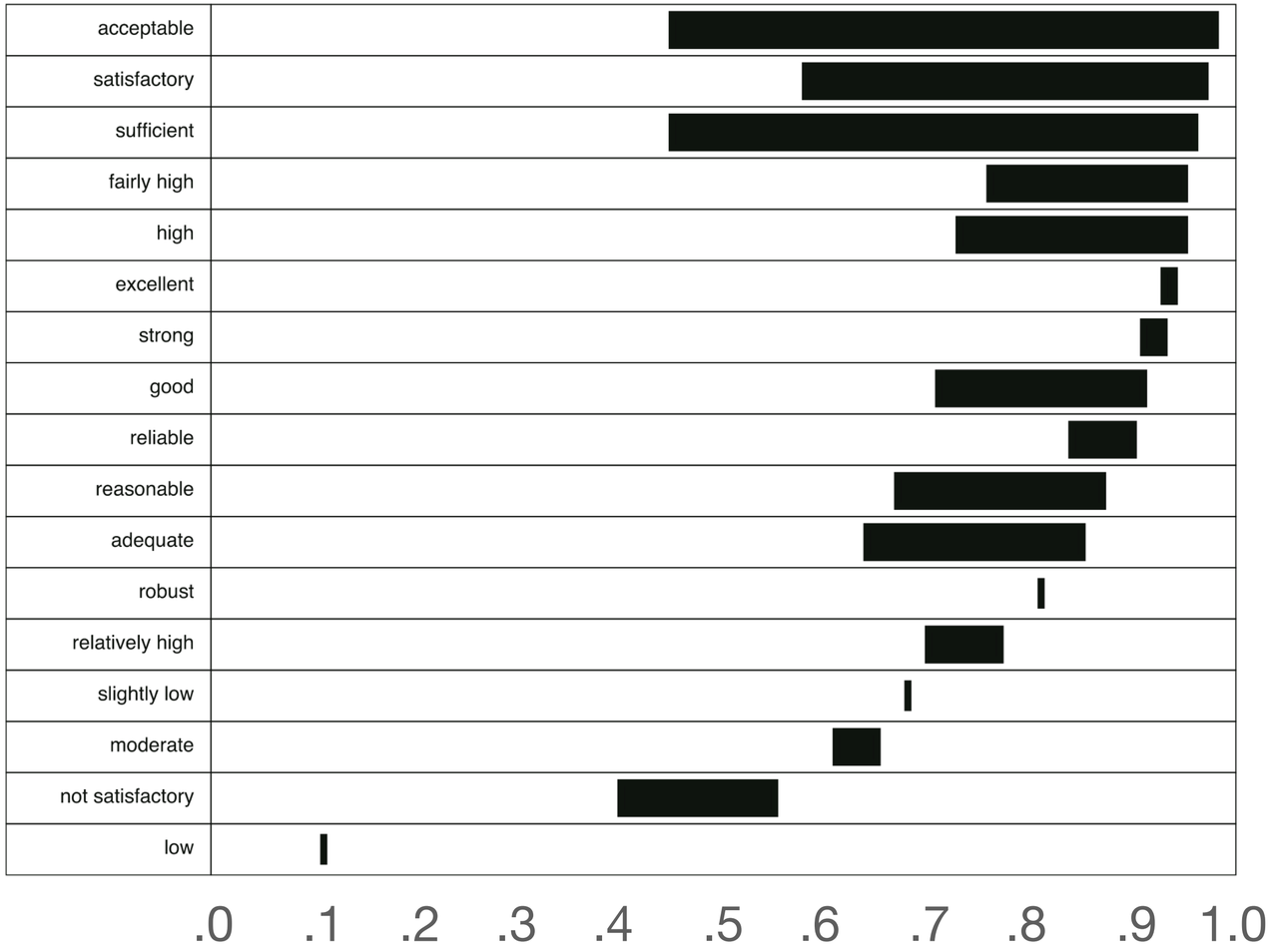
Most papers report nothing about reliability

Those that do, report Cronbach’s alpha

(Flake et al. 2017)

Who cares! Just report its “adequate”

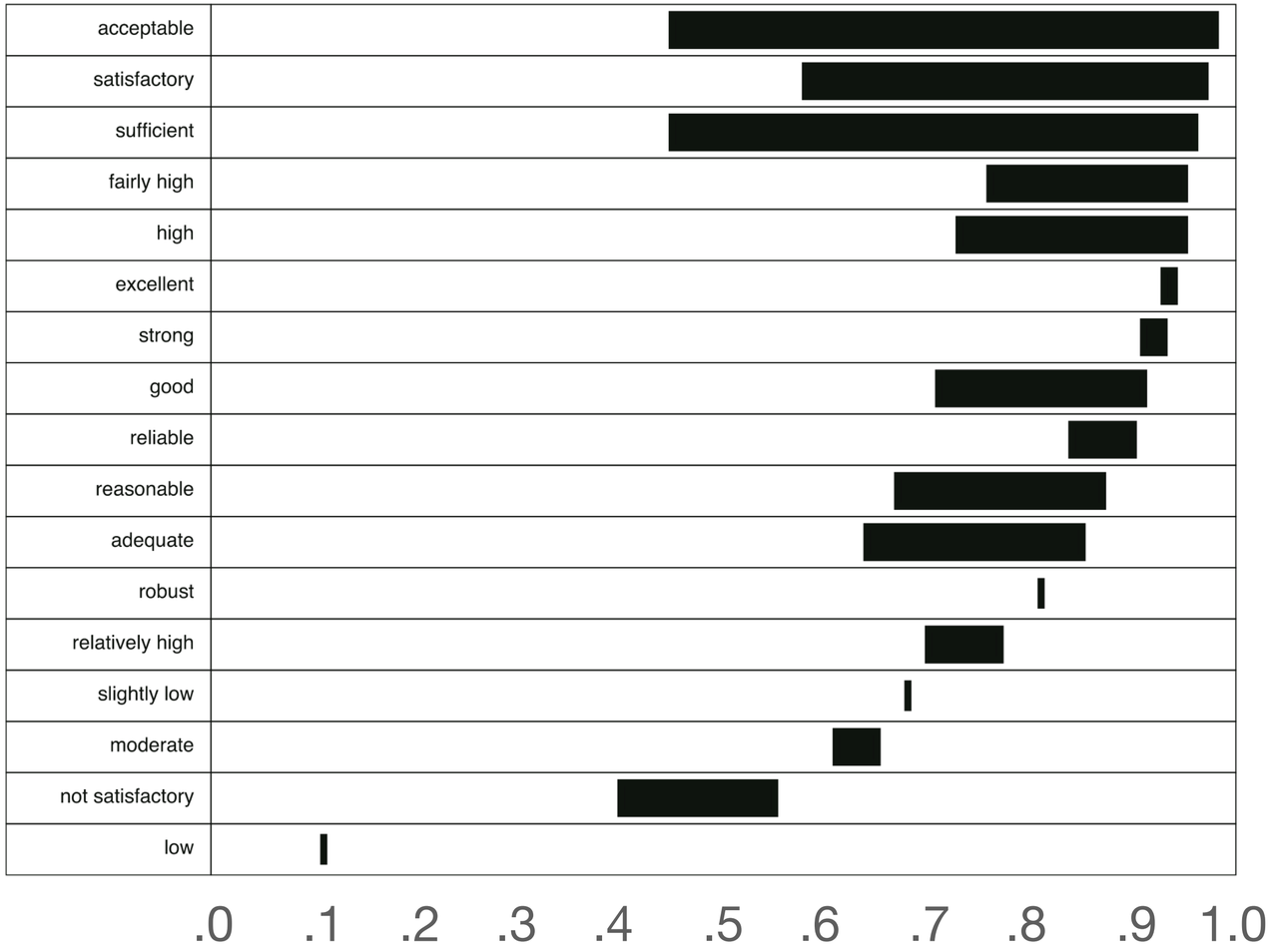
Call your alpha value whatever you want
(Taber, 2018)



Who cares! Just report its “adequate”

Call your alpha value whatever you want
(Taber, 2018)

Including “marginally reliable”



Who cares! Just report its “adequate”

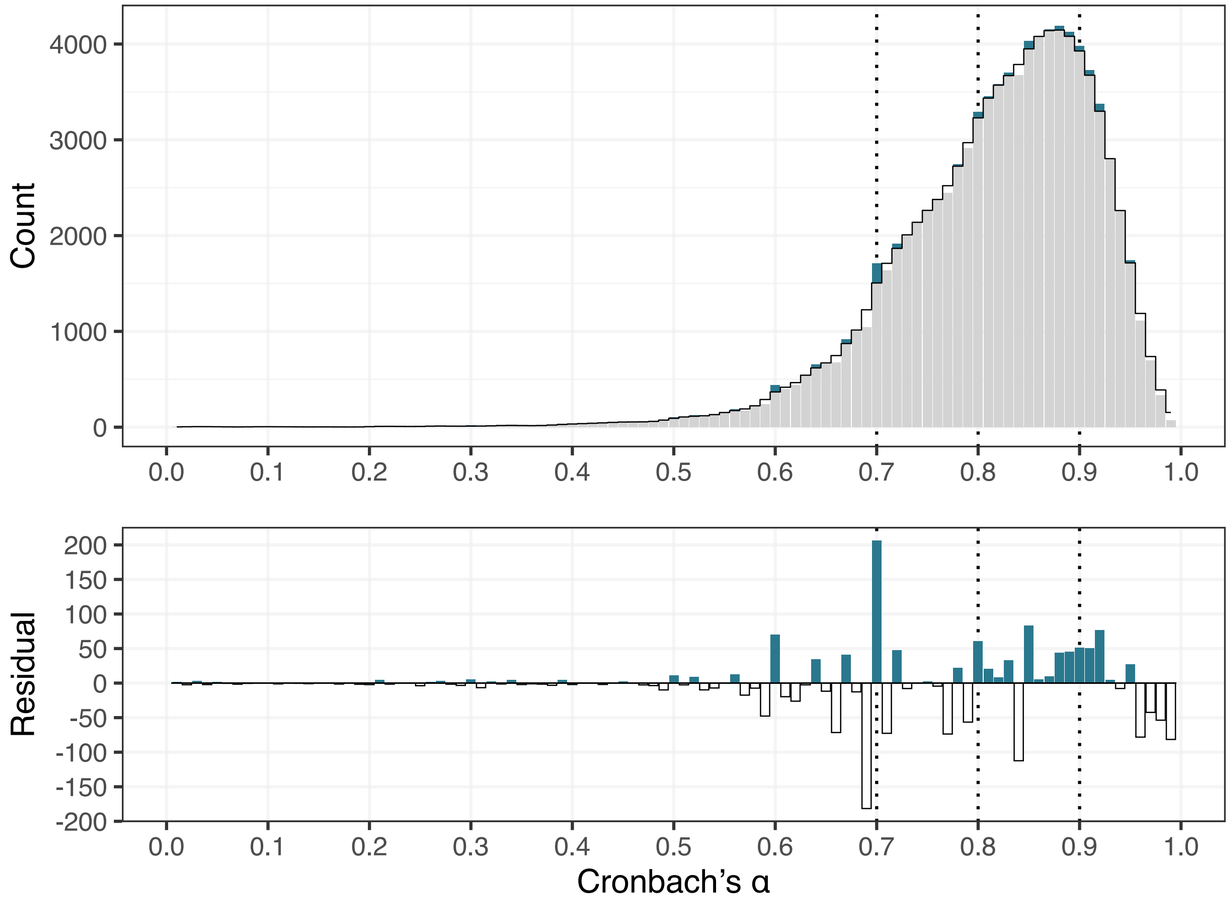
Same for test-retest reliability
(Watson, 2004)

> “researchers almost invariably concluded that their [test-retest] correlations were ‘adequate’ or ‘satisfactory,’ regardless of the size of the coefficient or the length of the retest interval.”

Schmeasurement and hacking

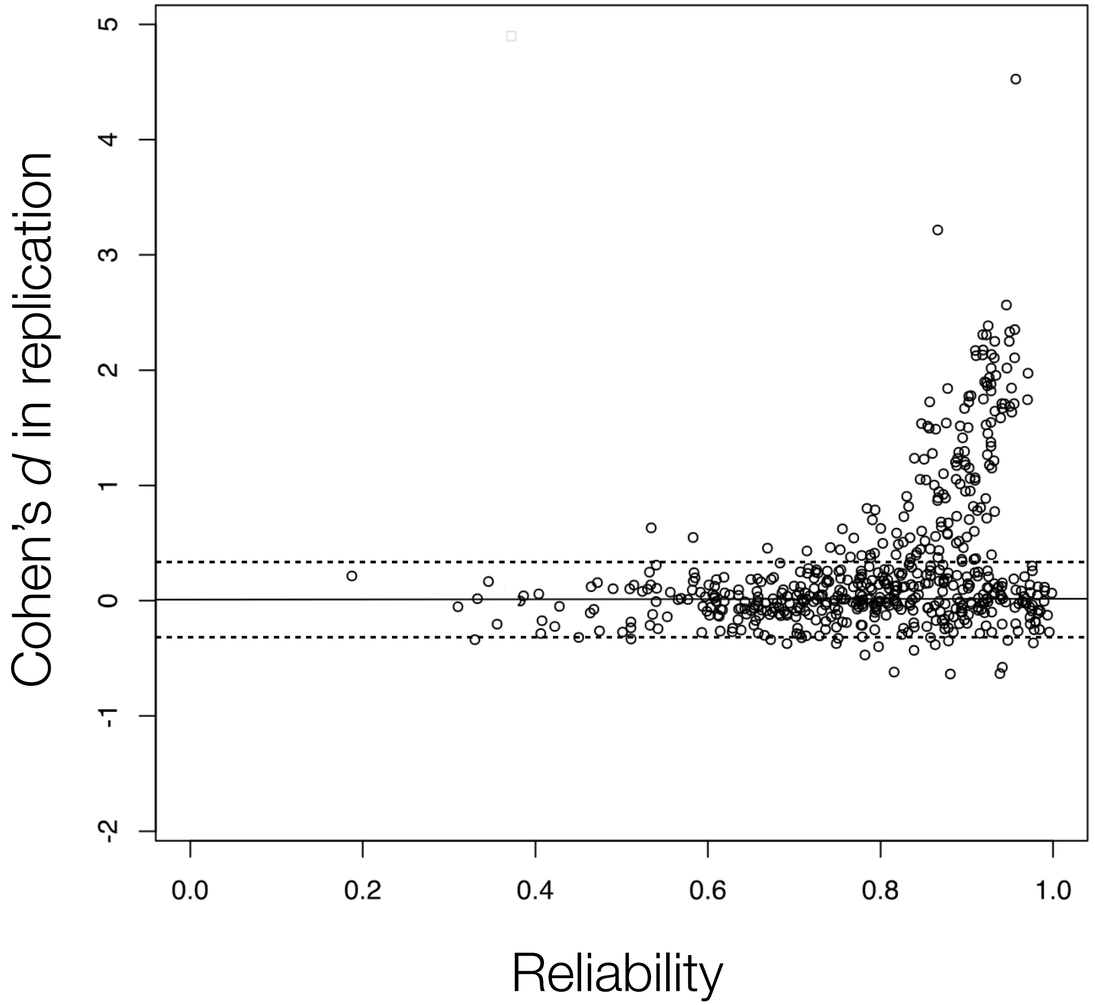
Systematic disinterest in reliability and validity
(Flake and Fried, 2021)

Cronbach's alpha-hacking
(Hussey et al., 2024)



Reliability is necessary but not sufficient for replicability

- When reliability is low, replicated effect size is low or null
- When reliability is high, replicated effect size ranges from null to large
  - Markon (2021)



Reliability is necessary but not sufficient for replicability

- Reliability's influence on distribution [Quarto document]
- Reliability's influence on effect size
  - https://matthewbjane.shinyapps.io/effect_size_artifact_bias/

Credibility checks

- It’s easier to assess reliability than run a replication
- Checks
  - Possibility (disattentuate effect sizes for reliability)
  - Plausibility (disattentuate effect sizes for reliability)
    - https://errors.shinyapps.io/docc/

- Examples
  - Vahey et al. 2015 (r = .45) vs Hussey et al 2020 (reliability = .19)

Questions?