



# Betrouwbaarheid in comparatief beoordelen



San Verhavert



*α*



# Scale Separation Reliability SSR



# Interne Consistentie



Bradley-Terry-Luce model

Scale Separation Reliability  
SSR

Cronbach

Interne Consistentie

Inter-Rater Reliability

IRT

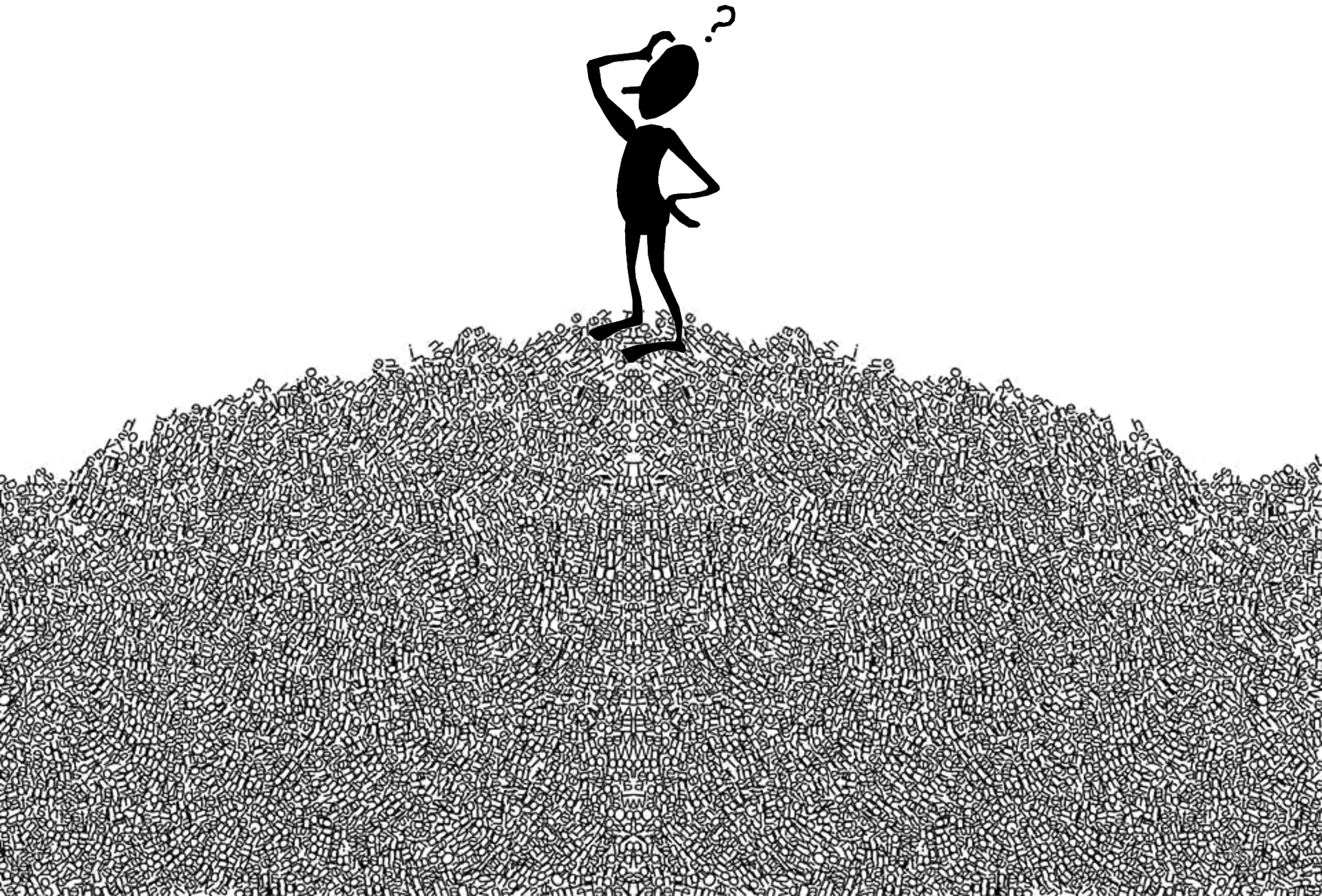
Rasch

$\alpha$

Test-Retest Reliability

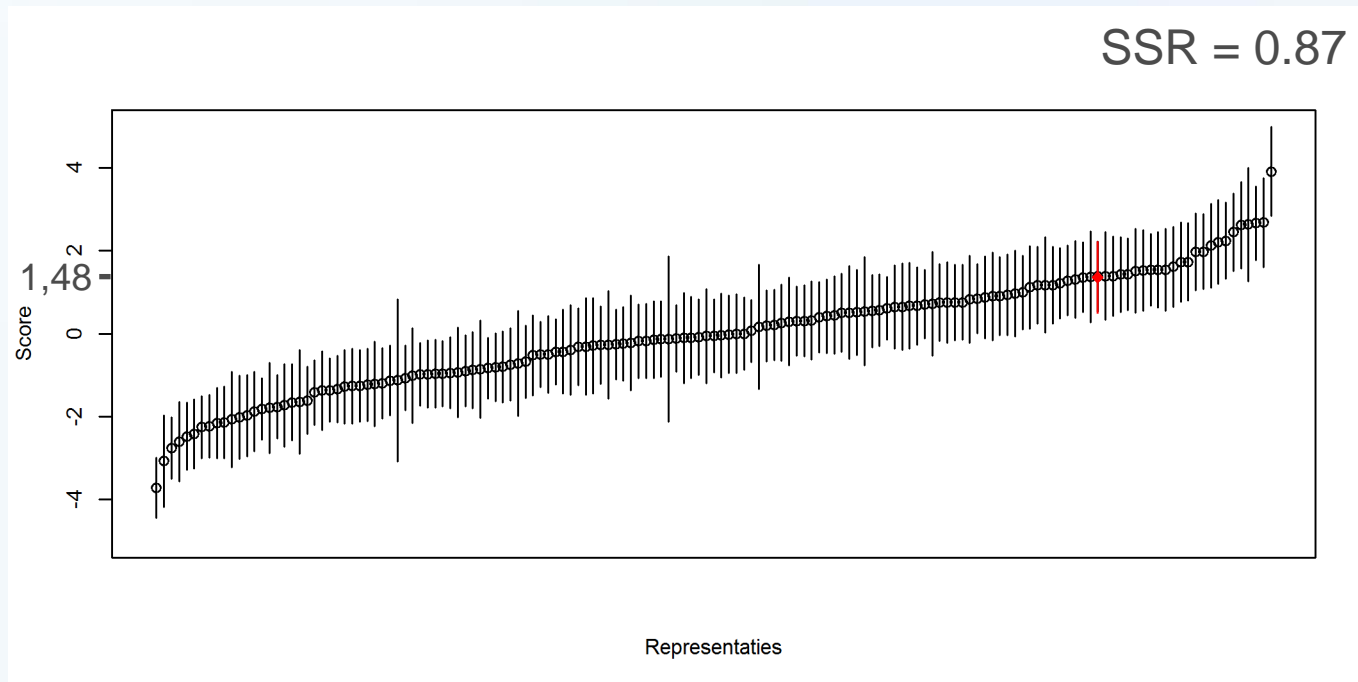
correlatie





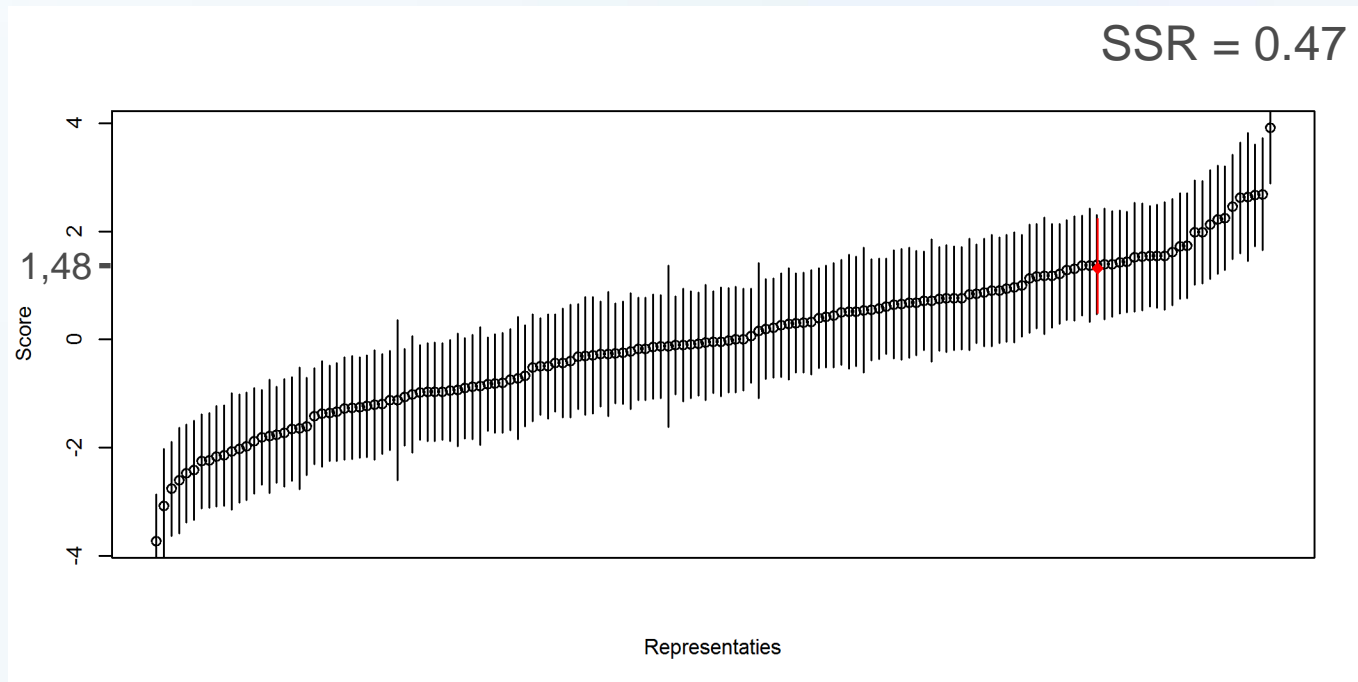
# Betrouwbaarheid? Belangrijk?

- Theoretisch en statistisch



# Betrouwbaarheid? Belangrijk?

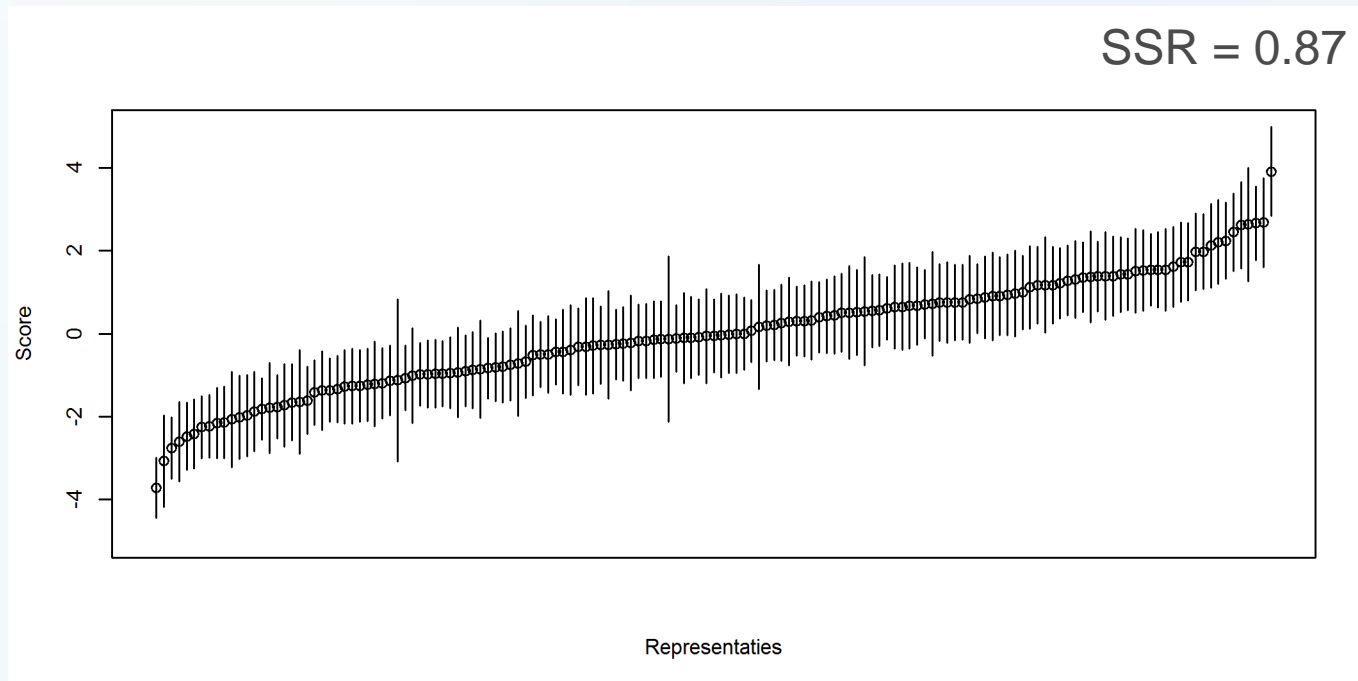
- Theoretisch en statistisch





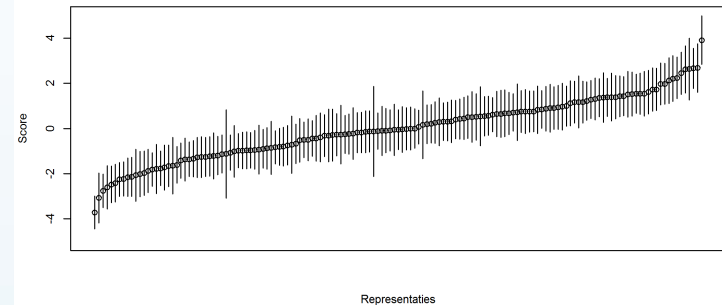
# Betrouwbaarheid?

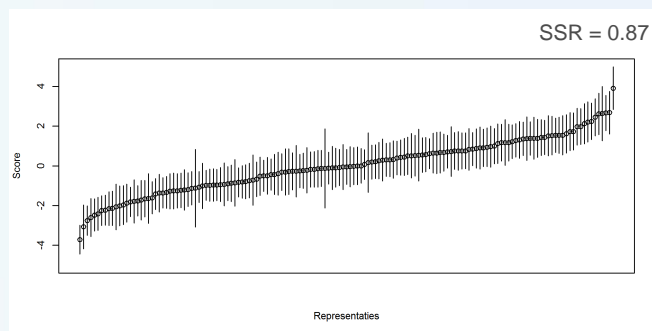
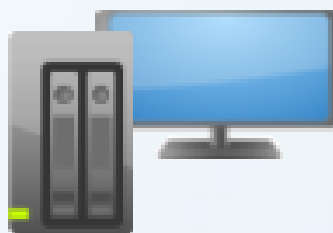
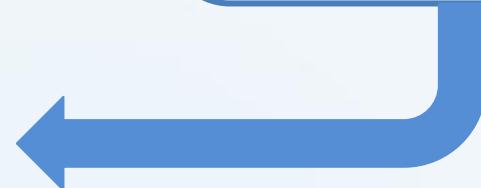
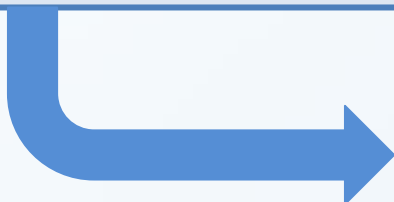
- Theoretisch en statistisch

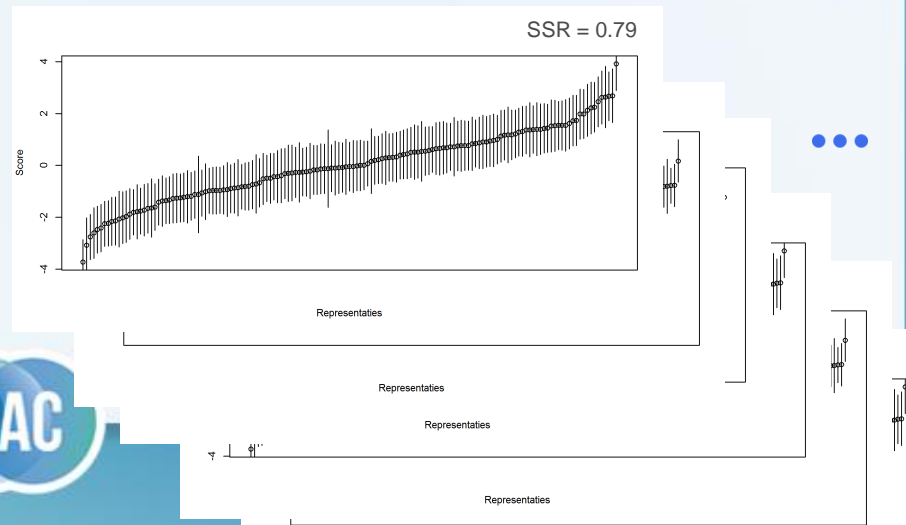
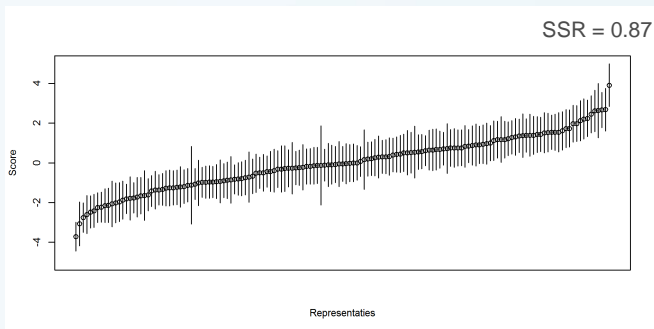
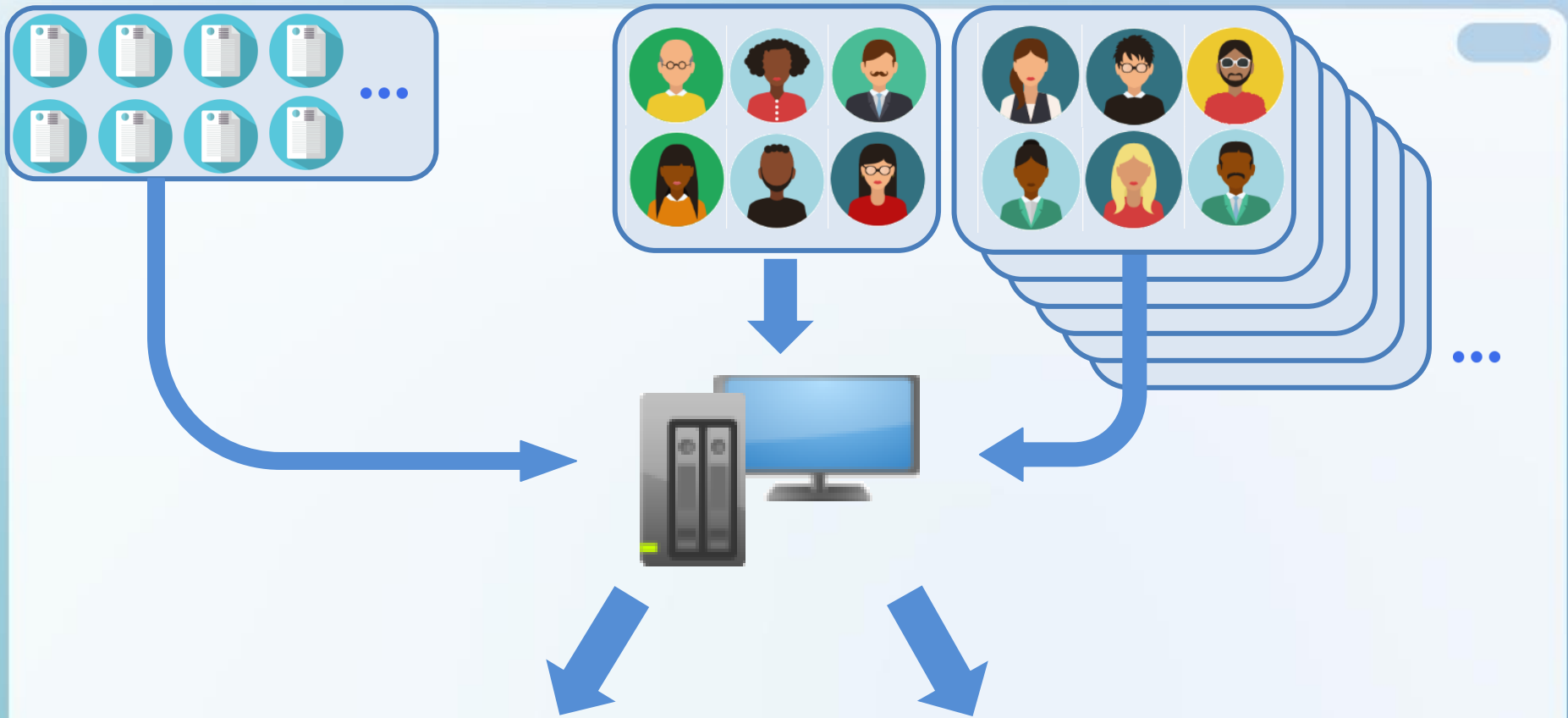


# Betrouwbaarheid?

- Theoretisch en statistisch
- Overeenkomst
  - Tussen groepen beoordelaars
  - Binnen een groep beoordelaars

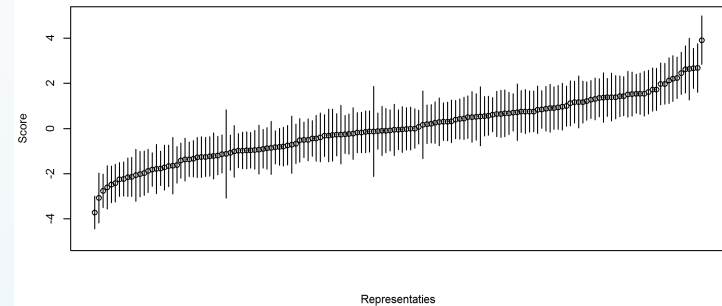


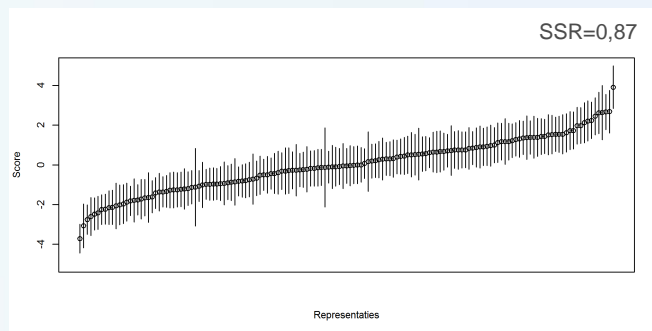
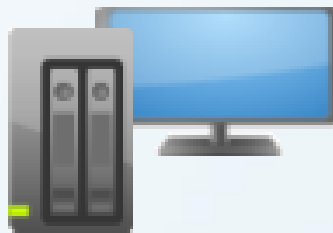
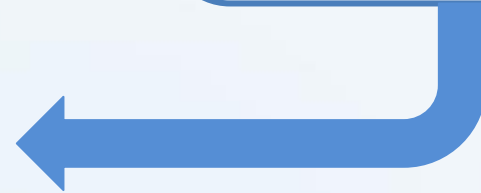
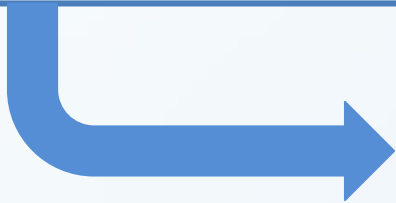


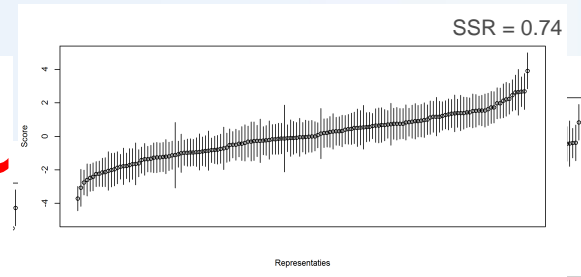
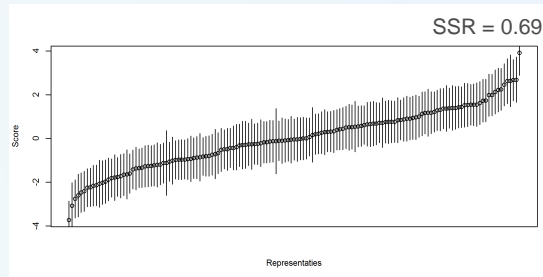
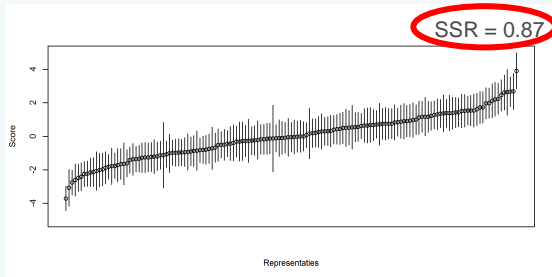
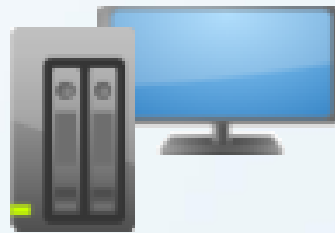


# Betrouwbaarheid?

- Theoretisch en statistisch
- Overeenkomst
  - Tussen groepen beoordelaars
  - Binnen een groep beoordelaars







# Betrouwbaarheid?

- Theoretisch en statistisch
- Overeenkomst
  - Tussen groepen beoordelaars
  - Binnen een groep beoordelaars
- Betrouwbaarheid is belangrijk





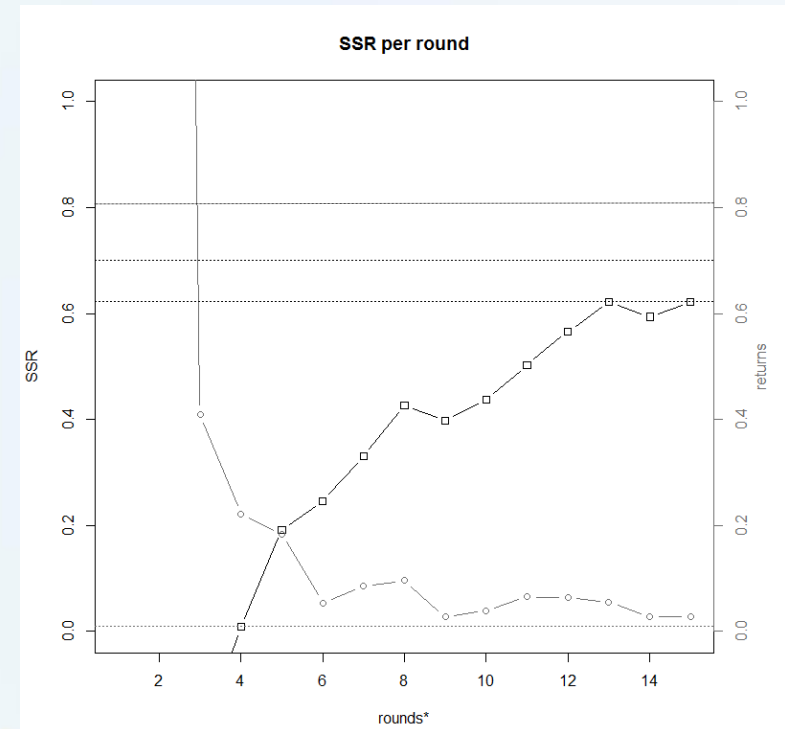
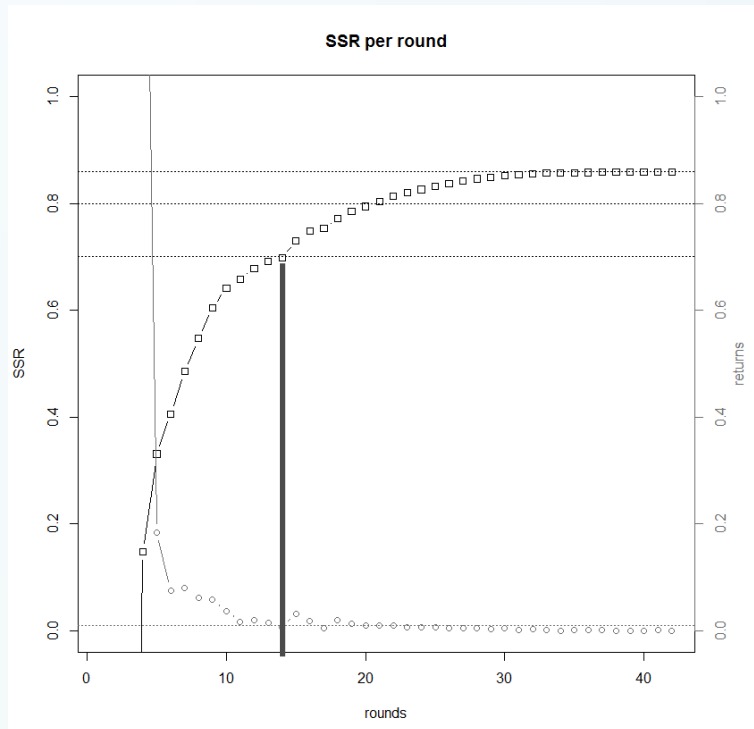
# betrouwbare schaal

- Hoeveel beoordelaars?
- Hoeveel werk?
- Hoeveel representaties?
- Voorwaarden beoordeelaars?



# Invloed op betrouwbaarheid

- Aantal vergelijkingen per representatie



(Verhavert, Bouwer, De Maeyer, Donche, in prep)

# Invloed op betrouwbaarheid

- Aantal vergelijkingen per representatie
- Aantal vergelijkingen in totaal
- Aantal beoordelaars
- Aantal vergelijkingen per beoordelaar
- Expertise van de beoordelaar
- Soort feedback
- Soort representaties
- ...



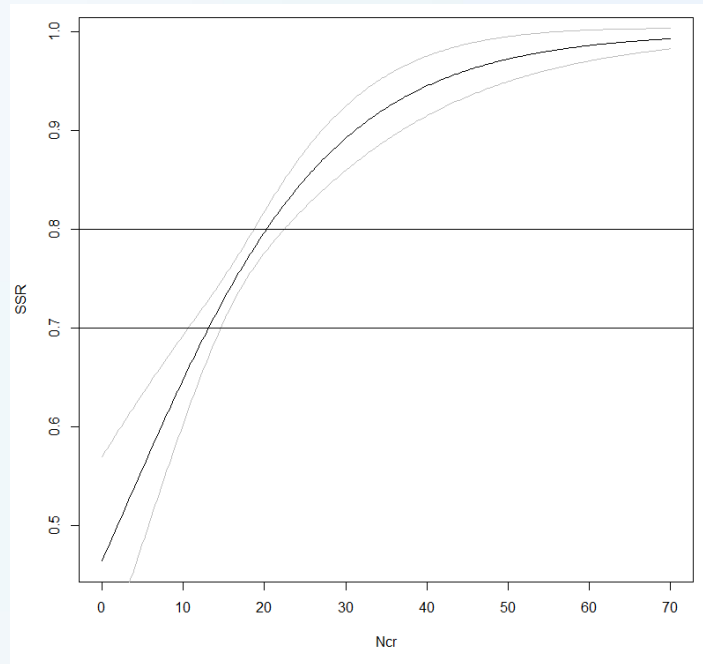
# Invloed op betrouwbaarheid: resultaten

- Aantal vergelijkingen per representatie
- ~~Aantal vergelijkingen in totaal~~
- ~~Aantal beoordelaars~~
- Aantal vergelijkingen per beoordelaar
- ~~Expertise van de beoordelaar~~
- ~~Soort feedback~~
- ~~Soort representaties~~
- ~~...~~



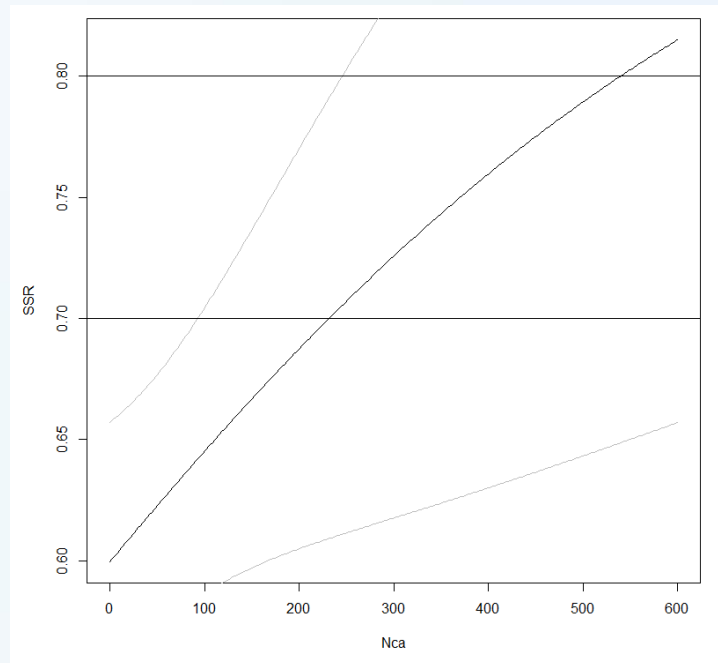
# Aantal vergelijkingen

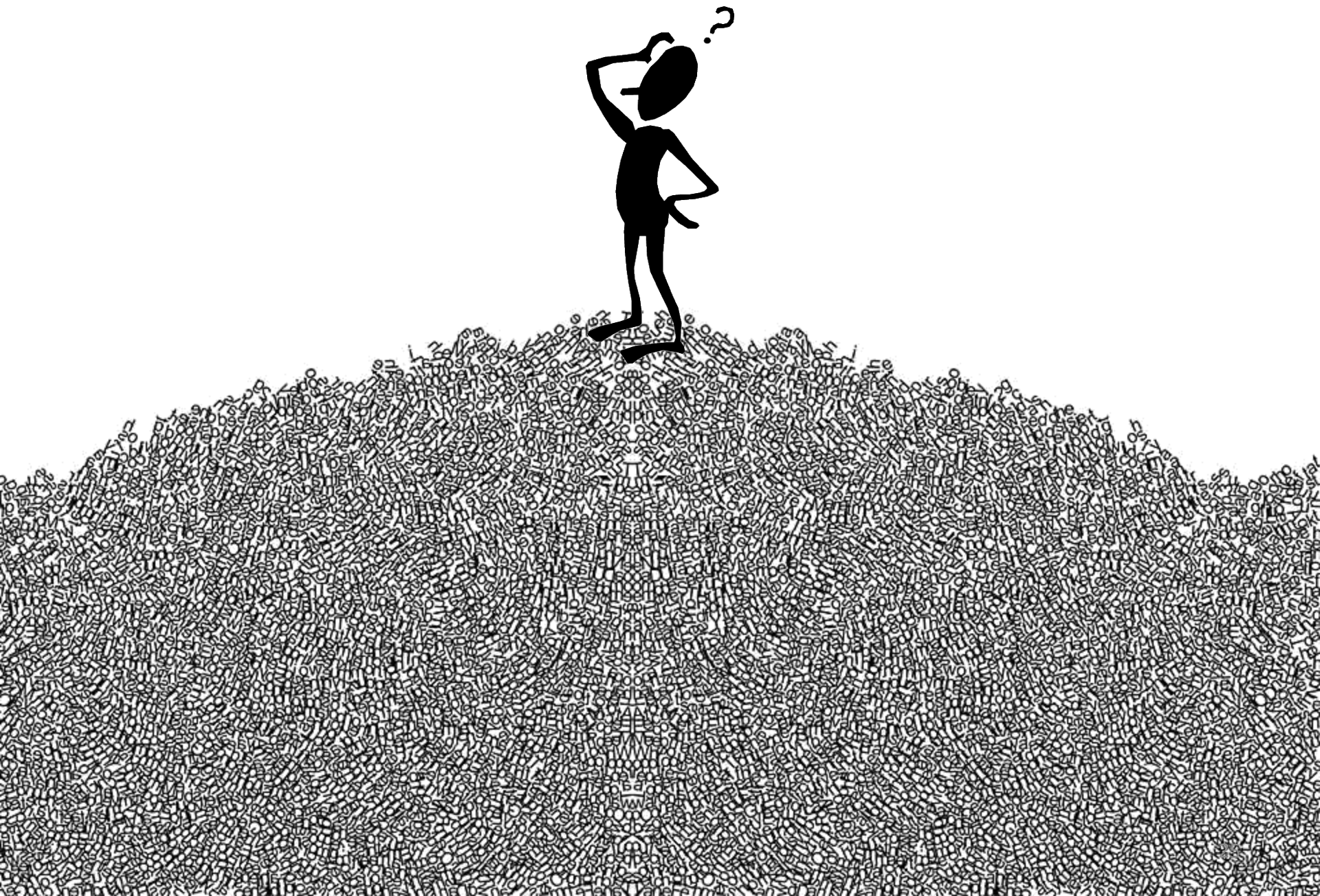
- .70: 8-12 vergelijkingen per representatie
- .80: 16-20 vergelijkingen per representatie



# Aantal vergelijkingen beoordelaar

- Stijging van 0.01 in betrouwbaarheid  
=> 20 vergelijkingen per beoordelaar extra









Unless you have confidence in the reliability of the ruler, if you use a ruler to measure a table you may also be using the table to measure the ruler.  
– Nassim Nicholas Taleb

“Tenzij je vertrouwen hebt in de betrouwbaarheid van de liniaal kan je even goed de liniaal gebruikt om een tafel te meten als de tafel gebruiken om de liniaal te meten. [vrij vertaald]

– Nassim Nicholas Taleb  
(Libanees-Amerikaanse auteur en statisticus)

# Vragen?

