

# **Scale length does matter: Recommendations for Measurement Invariance testing with Categorical Factor Analysis and Item Response Theory Approaches**

E. Damiano D'Urso

Dr. Kim De Roover

Prof. Dr. Jeroen K. Vermunt

Dr. Jesper Tijmstra





# Paragone fra gruppi



To what degree do you like apples?

not at all  
☐

a little  
☐

rather  
☐

very much  
☐

Do you believe death penalty is ever justified?

①  
Strongly  
Agree

②  
Agree

③  
Neither

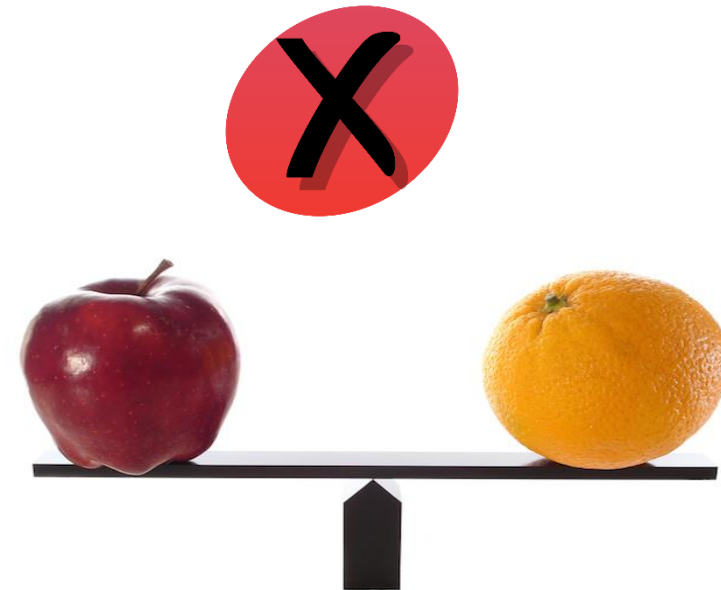
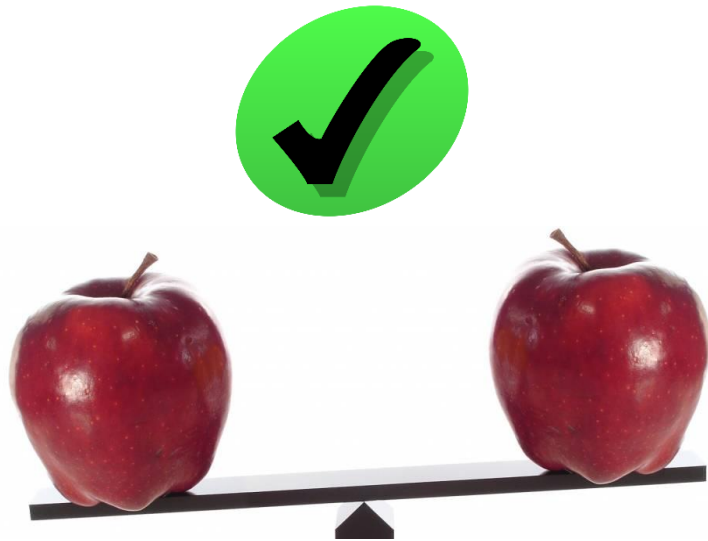
④  
Disagree

⑤  
Strongly  
Disagree



# Invarianza di Misura

- Statistical property of a measurement that indicates that the same construct is measured equivalently across groups (Vandenberg & Lance, 2000)





# Due approcci

**MG-CCFA**



**MG-IRT**

MG- Normal Ogive Graded  
Response Model



# Somiglianze



- ☐ Modelli di misura fundamentalmente equivalenti (Chang et al., 2017)
- ☐ Test loadings/indice di discriminatività
- ☐ Test thresholds (soglie)

# Differenze



- ☐ Tipologia del test:
  - Tutti gli items contemporaneamente (MG-CCFA);
  - Un item alla volta (MG-IRT).
- ☐ Strategie/criteri:
  - MG-CCFA
    - Indici di Fit assoluto:  $\chi^2$ ;
    - Indici di misfit: Root mean square error approximation (RMSEA);
    - Indici di fit comparativo: comparative fit index (CFI);
  - MG-IRT:
    - Likelihood-Ratio Test (LRT);
    - Logistic Regression (LoR).



# Quale scegliere

- A quale livello conviene testare l'invarianza di misura?
- Quali strategie/criteri sono maggiormente prestanti nell'individuare l'invarianza di misura?





# Per rispondere....





# Studi simulativi

- **Studio simulativo 1:** Costrutto misurato equivalentemente (invarianza di misura).
- **Studio simulativo 2:** Costrutto misurato non-equivalentemente (non-invarianza di misura)

Misure

Studio 1: False Positive Rate

Studio 2: True positive Rate, False positive Rate





# RISULTATI & RACCOMANDAZIONI

Lunghezza della scala: 5, 25





# Limiti e direzioni future

- Scale unidimensionali
- Solo due gruppi
- Lunghezza della scala variata solamente a due livelli (i.e., 5 e 25 items)



# Thank you for your attention!

- E. Damiano D'Urso
- E.D.Durso@tilburguniversity.edu

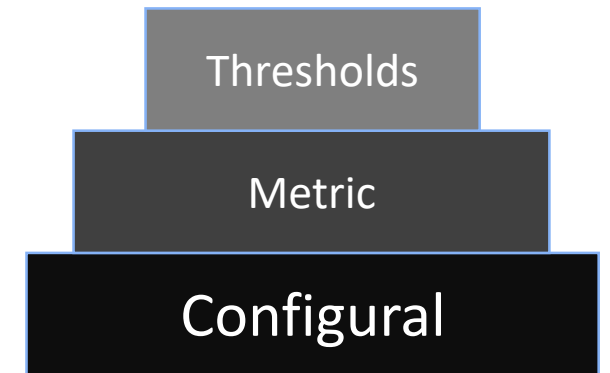
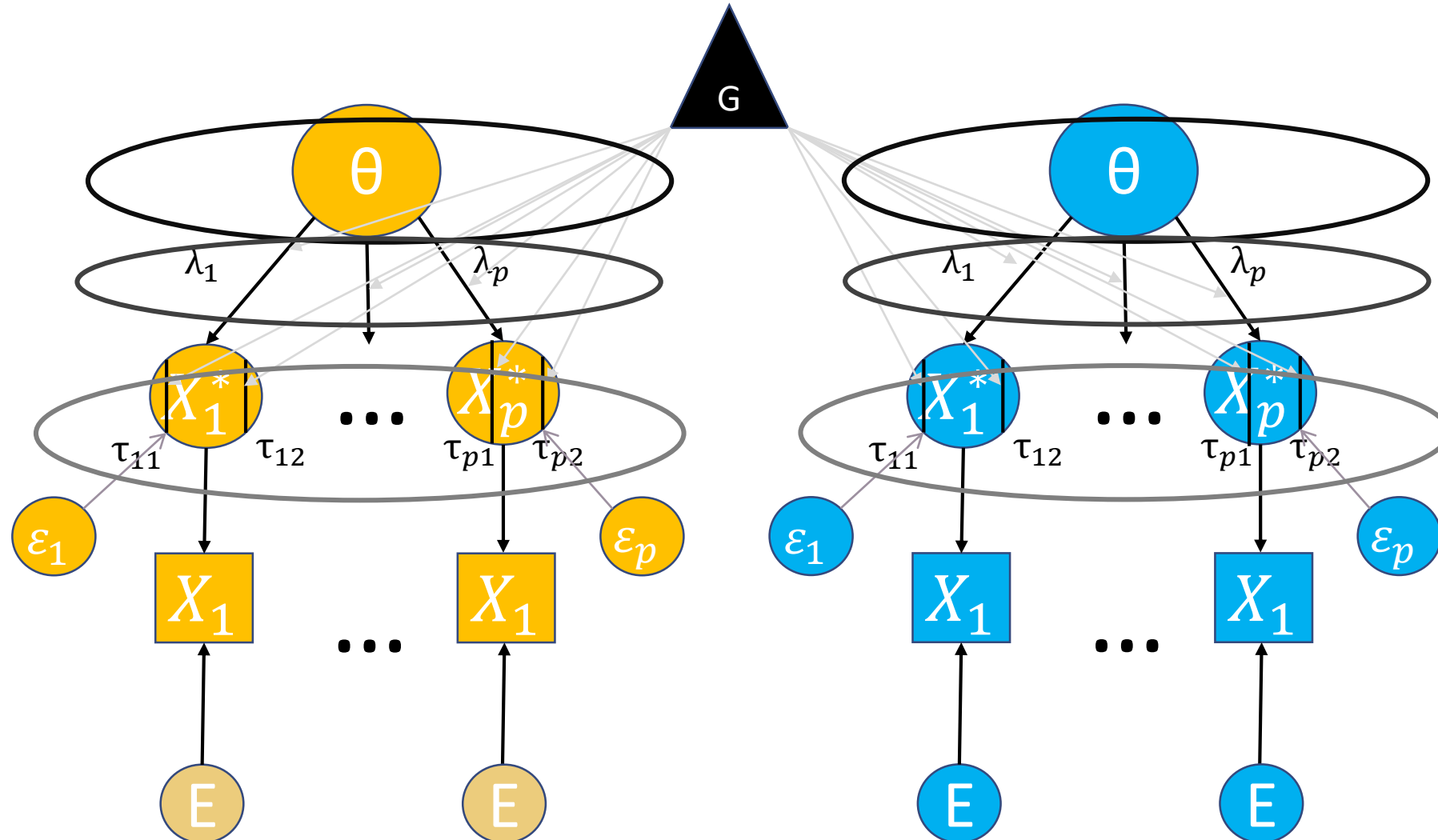


# References

- Chang, Y.-W., Hsu, N.-J., & Tsai, R.-C. (2017). Unifying differential item functioning in factor analysis for categorical data under a discretization of a normal variant. *Psychometrika*, 82 (2), 382–406.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3 (1), 4–70.

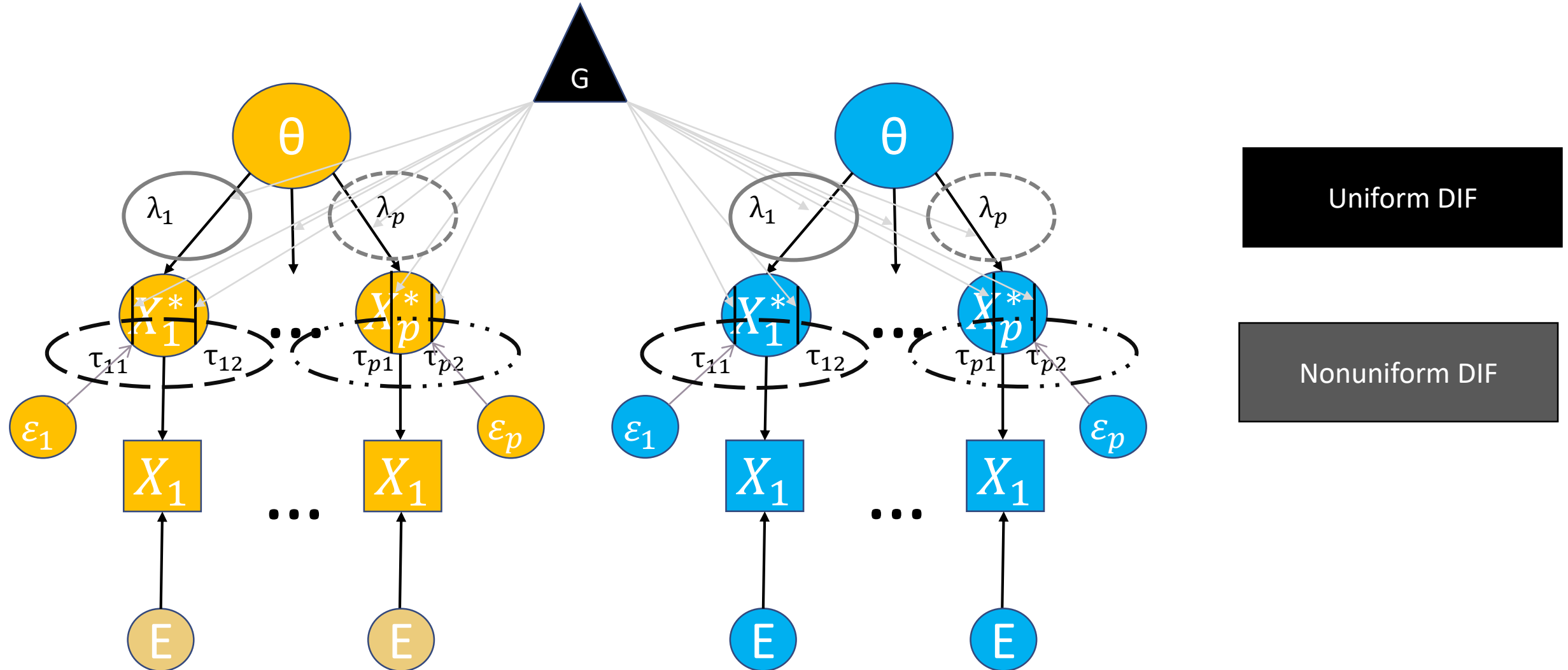


# MI in MG-CCFA $\approx$ MG-IRT





# MI in MG-IRT (GRM) $\approx$ MG-CCFA





# Studio simulativo 1

- Obiettivo dello studio: verificare la performance di MG-CCFA e MG-IRT nel caso in cui ci sia MI
- Fattori:
  - # items: 5, 25;
  - # categorie: 3, 5;
  - dimensione campione per gruppo: 250, 500.



# Studio simulativo 2

- Obiettivo dello studio: verificare la performance di MG-CCFA e MG-IRT nel caso in cui **NON** ci sia MI
- Fattori:
  - items: 5, 25;
  - categorie: 3, 5;
  - dimensione campione per gruppo : 250, 500;
  - % items con loadings non-invarianti: 20%, 40% aligned, 40% misaligned;
  - % items con soglie non-invarianti: 20%, 40% aligned, 40% misaligned;
  - Non-invarianza: piccola, grande.





# Risultati principali (Simulazione 2)

- La prestazione dipende dalla lunghezza della scala:
  - Per scale corte (i.e., 5 items): MG-CCFA (usando sia  $\chi^2$  e uno fra CFI e RMSEA);
  - Per scale lunghe (i.e., 25 items): MG-IRT con Likelihood-Ratio Test (LRT);
- La potenza statistica, nel caso di MG-CCFA, dipende da:
  - Lunghezza della scala: nettamente peggiore con scale lunghe;
  - Tipologia di test: nettamente peggiore quando il test é condotto a livello dell'item;