The Impact of Rapid Guessing on Model fit and Factor-analytic Reliability

Alfonso J. Martinez (Presenter), University of Iowa

Joseph A. Rios, University of Minnesota

Overview

- RG has been found to **distort psychometric inferences** if not properly accounted for.

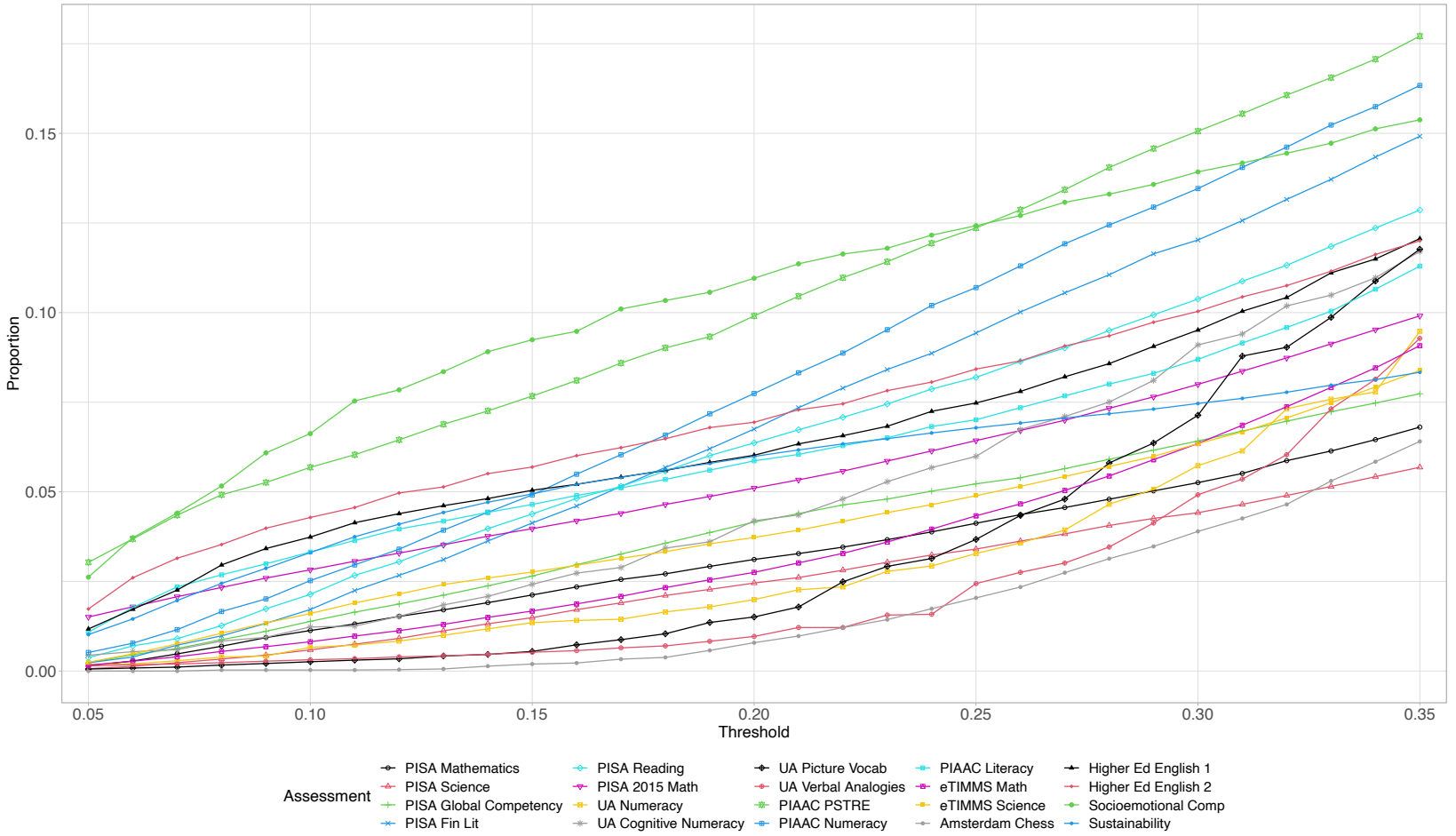
- **The effects of RG on model fit and reliability** are not well understood.

- Model fit and reliability are useful **for identifying low-performing items** and establishing evidence of **convergent and construct validity**

Research Objectives

- The present study **explores how model fit and reliability are influenced by the presence of RG** via an applied analysis
  - Model fit: SEM indices – **SRMR, RMSEA, CFI, TLI**
  - Reliability: **coefficient alpha, coefficient omega**
- **Compare three different RG corrective procedures**
  - Naïve ("do-nothing")
  - Penalized (treat RG responses as incorrect)
  - Effort-moderated with imputation (treat RG responses as missing and impute response based on plausible values)
- Identify RG via **sequential thresholds method**
  - Response is classified as RG if response time falls below $\mathbf{100}x\%$ **of item** i**'s mean response time** $(x = 0.05, \ldots, 0.35)$

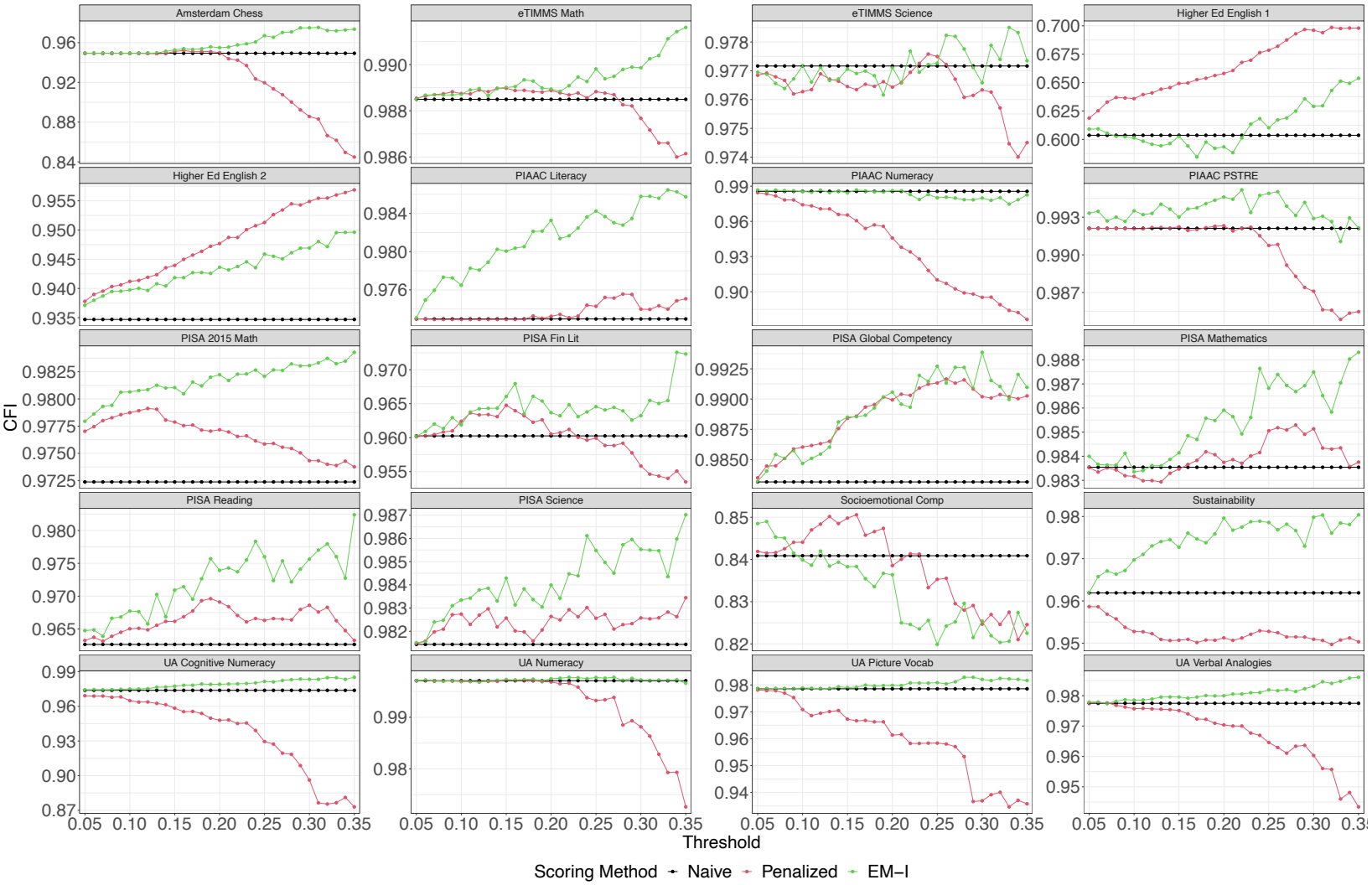
Selected Results (Rates of RG Classification)



Proportion of Examinee-by-Item Interactions Classified as RG

Notes: All results can be found in manuscript (uploaded to NCME server)

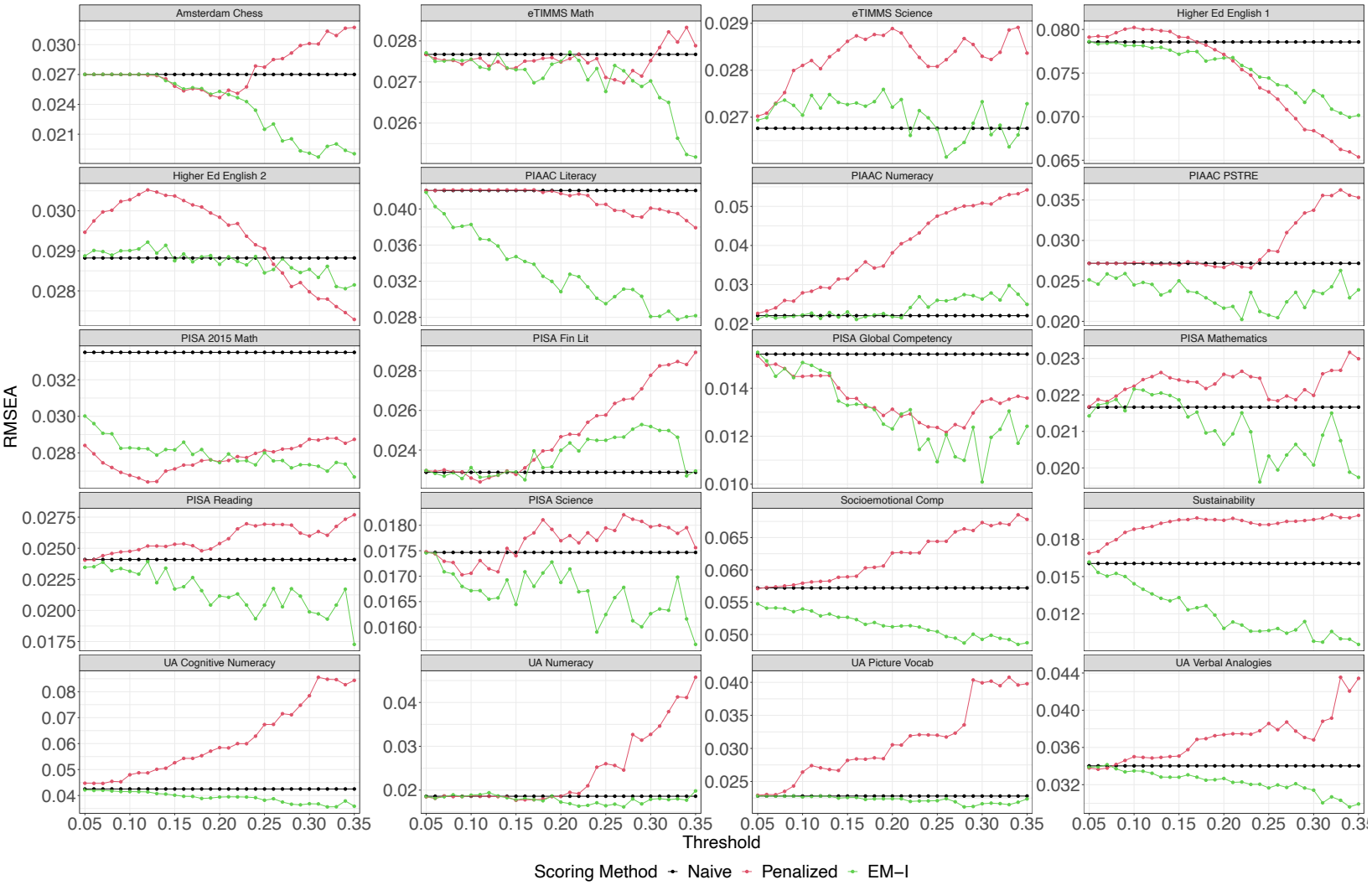
Selected Results (Trends in CFI)



Comparative Fit Index (CFI) as a function of threshold value and scoring procedure

Notes: All results can be found in manuscript (uploaded to NCME server)

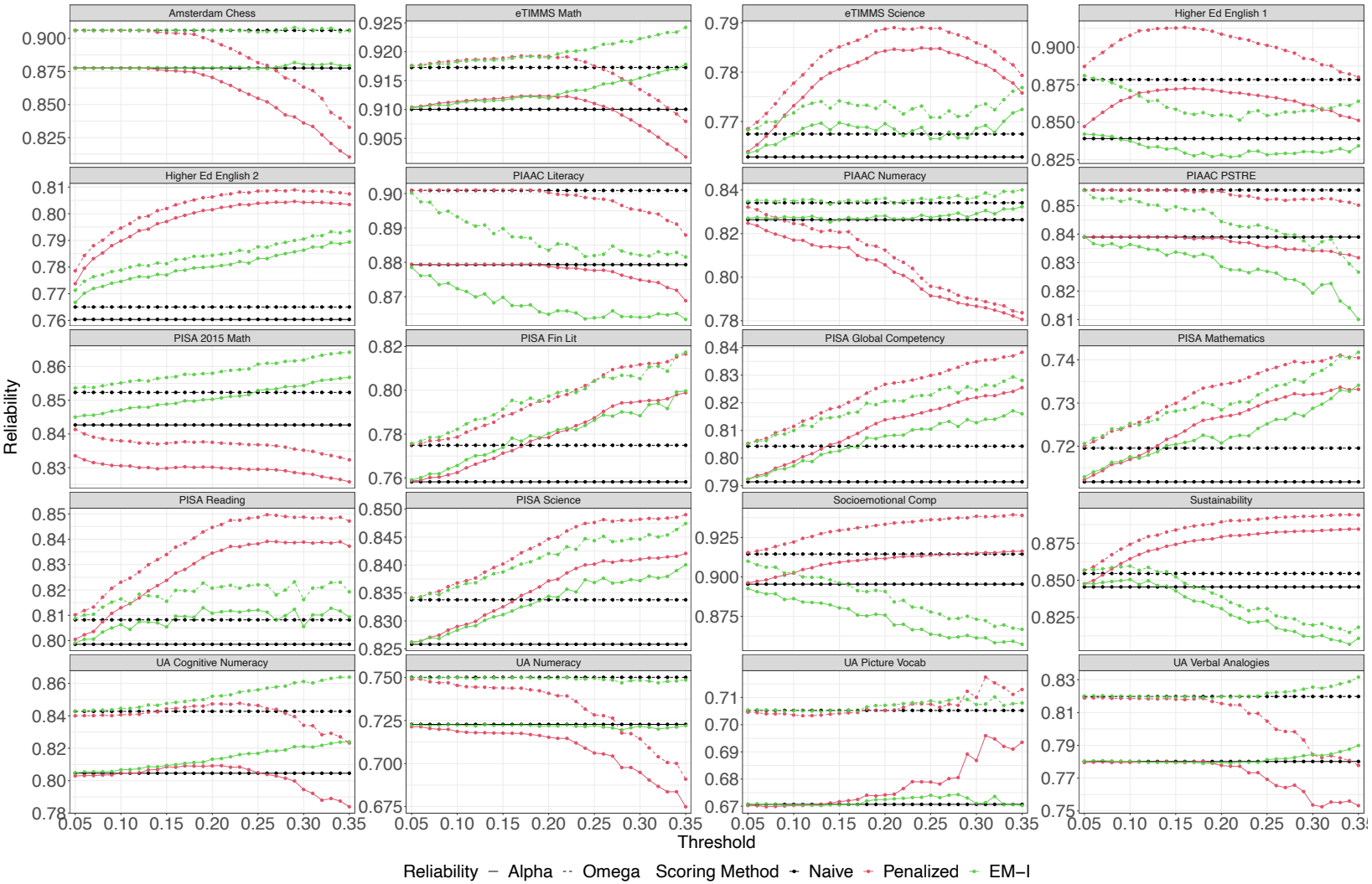
Selected Results (Trends RMSEA)



Root mean square error of approximation (RMSEA) as a function of threshold value and scoring procedure

Notes: All results can be found in manuscript (uploaded to NCME server)

Selected Results



Reliability as a function of threshold value and scoring procedure

Notes: All results can be found in manuscript (uploaded to NCME server)

Main Takeaways

- Best model fit under **EM-I scoring** if the **threshold was liberal** (e.g., $t > 0.25$)

- Reasonable model fit under penalized scoring if the threshold was conservative (e.g., $t < 0.10$)
  - Naïve scoring may be more appropriate although it ignores effects of RG

- Omega reliability was higher for all threshold values for all assessments
  - **No systematic trends** in reliability across corpus
  - Needs to be investigated on **a case-by-case basis**