EAM2025

XI Conference

European Association of Methodology

23RD - 25TH JULY 2025

Spain Tenerife Canary Islands

Ipsativity indices for forced-choice assessments

Rodrigo S. Kreitchmann, Miguel A. Sorrel, Diego F. Graña, and Francisco J. Abad

Universidad de La Laguna
Cabildo de Tenerife
Tagoro
cajasiete
Gobierno de Canarias
Consejería de Universidades, Ciencia e Innovación y Cultura
Agencia Canaria de Investigación, Innovación y Sociedad de la Información
ICI Instituto Canario de Igualdad
tea
hogrefe

- Ipsativity vs. Normativity
- Normativity Indicator
- Simulation with Real Item Pool
- Conclusions
- Future Directions

Ipsativity:

- From the Latin *ipse* (he, himself): The scores can only be interpreted for oneself, not allowing to compare between persons.
- Interdependence between multidimensional scales derived generally from the data format.
- E.g., forced-choice responses.

The forced-choice format:

- Respondents are required to rank statements (items) in a block:

	Most like me
I adapt to setbacks. (ES)	○
I fulfill my commitments. (CO)	○

- Under CTT, scores reflect the number of endorsements of the items in each dimension.
- The total number of endorsements across all dimensions is constant.
- That is, the scores reflect proportions of the endorsement of different dimensions out of a constant total.

Consequences of ipsative scoring:

A respondent has a higher score in trait A than trait B.

~~Respondent 1 has a higher score than Respondent 2.~~

No absolute interpretation of the scores (underidentified origin).

Some properties of ipsative scores (Clemans, 1966; Hicks, 1970):

- Sum/mean of scores is constant across respondents .
  - A higher endorsement of items in one dimension necessarily implies a lower endorsement of those of the other dimensions.
  - Impossibility to rank respondents in multivariate space.
- Negative interdependence between scales:
  - For truly independent traits, observed correlation of $\frac{-1}{D-1}$.
- Sum/mean of validity coefficients of 0.
- Thus, internal consistency and external validity are compromised.

Overcoming ipsativity:

- Building strategies to avoid constant sum/mean of scores:
  - CTT: including reversely coded items, with negative scoring key.
  - IRT: Using items with diverse discrimination parameters.
- Still, it is difficult to determine the amount of ipsativity in scores.

Normativity Score Decomposition

- Conceptually, scores are fully ipsative if their sum (or mean) is constant across all individuals.
- It is our aim to quantify the true variance of within-person average across dimensions ($\bar{\hat{\theta}}_i$).
- I.e., Variation of within-person mean scores that is due to true within-person mean score variation.

Normativity Score Decomposition

$$\hat{\theta}_{i,k} = \bar{\hat{\theta}}_i + \left(\hat{\theta}_{i,k} - \bar{\hat{\theta}}_i\right)$$

Estimated Score

Within-person Average Estimate

Within-person Differential Score Estimate

$$\bar{\hat{\theta}}_i = \bar{\theta}_i + \varepsilon_{\text{mean}}$$

$$\left(\hat{\theta}_{i,k} - \bar{\hat{\theta}}_i\right) = \left(\theta_{i,k} - \bar{\theta}_i\right) + \varepsilon_{\text{diff}_k}$$

Normativity Score Decomposition

- Estimated score variance can be decomposed as:

$$\mathrm{var}\left(\hat{\theta}_{i,k}\right) = \mathrm{var}\left(\bar{\hat{\theta}}_i\right) + \mathrm{var}\left(\hat{\theta}_{i,k} - \bar{\hat{\theta}}_i\right) + 2 \cdot \mathrm{cov}\left(\bar{\hat{\theta}}_i, \hat{\theta}_{i,k} - \bar{\hat{\theta}}_i\right)$$

- True score variance and error variance can be decomposed in similar fashion.

(Theoretical) Normativity Index:

In terms of asymptotic error $\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_i|\boldsymbol{\theta}_i)$, it can be found that:

- The error variance of the within-person average estimate equals the mean of the conditional asymptotic error covariance matrix.

$$\mathrm{var}\left(\bar{\hat{\theta}}_i\right) = \mathrm{mean}[\mathrm{Cov}(\widehat{\boldsymbol{\theta}}_i|\boldsymbol{\theta}_i)]$$

(Theoretical) Normativity Index:

Reliability (normativity) of mean scores:

$$r^2_{\bar{\hat{\theta}}\bar{\theta}} = \frac{\int \mathrm{var}\left(\bar{\hat{\theta}}\right) g(\bar{\hat{\theta}})}{\mathrm{var}(\bar{\theta})},$$

where $\mathrm{var}(\bar{\theta})$ is the mean of the populational trait covariance matrix, $\mathrm{Cov}(\boldsymbol{\theta})$.

(Empirical) Normativity Index:

For empirical scores:

- The posterior standard deviation (PSD) of within-person means can be computed from square root of the average of the posterior covariance matrix.

(Empirical) Normativity Index:

Reliability (normativity) of empirical mean EAP scores:

$$r^2_{\bar{\hat{\theta}}\bar{\theta}} = \frac{\mathrm{var}\left(\bar{\hat{\theta}}\right)}{\mathrm{var}\left(\bar{\hat{\theta}}\right) + \left[N^{-1}\sum_{n=1}^{N}\mathrm{PSD}^2\left(\bar{\hat{\theta}}\right)\right]},$$

where $\mathrm{var}\left(\bar{\hat{\theta}}\right)$ is the variance of the estimated within-person averages.

Simulation Study with a Real Item Pool

Study Design:

- Block pool: 154 equally-keyed blocks of Big Five factors.
- Test lengths: 10, 20, 30, 40, 50, and 60 blocks.
- 1,000 simulees with $\boldsymbol{\theta} \sim MVN(\mathbf{0}, \Sigma_{NEO\text{-}PI})$.
- Criterion 1: simulated with correlation 0.3 with every $\boldsymbol{\theta}$ (multiple $R^2 = 0.36$). Big common variance with all traits.
- Criterion 2: simulated with correlation 0.6 with θ_1 (multiple $R^2 = 0.36$). Common variance with trait 1 only.

Procedure:

- For each test length condition, randomly assemble 100 tests from the pool (balanced content constraint).

Data Analysis:

- Recovery of true normativity (squared correlation between true and estimated within-person mean).

- Effect of normativity on multiple R-squared with criteria:

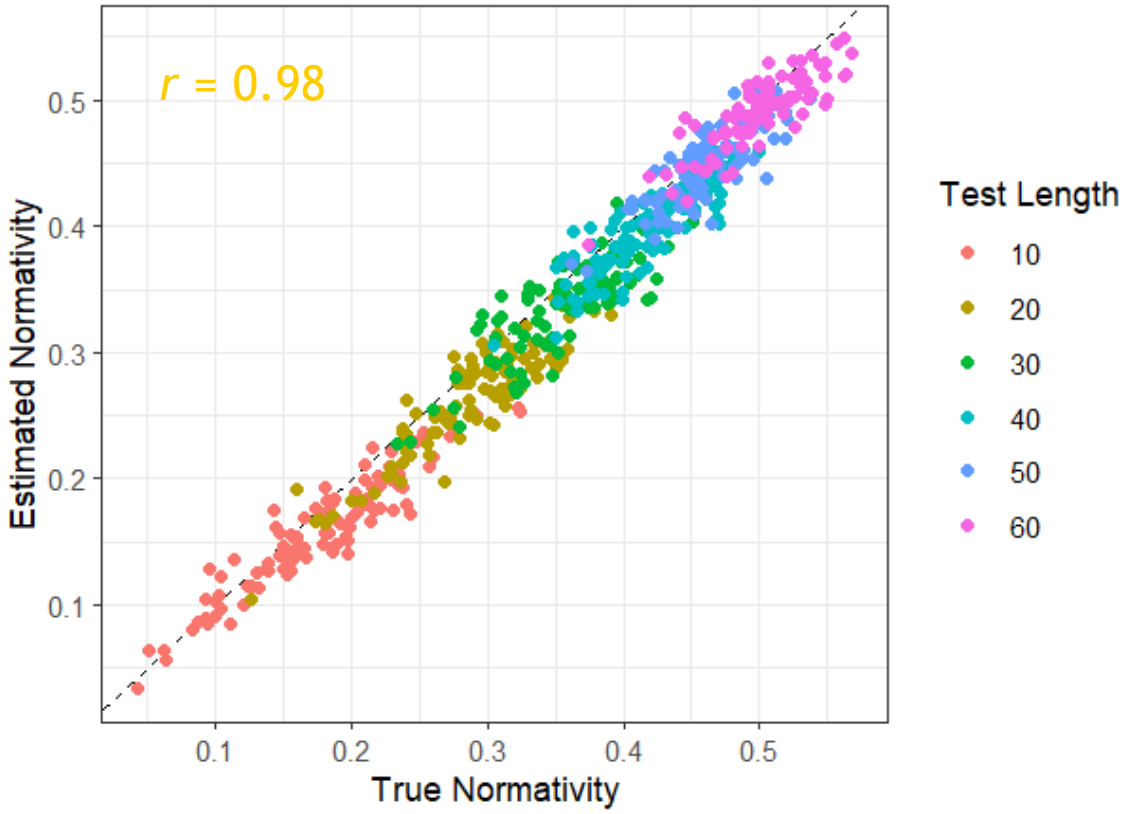
$$Y_1 \sim \hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3 + \hat{\theta}_4 + \hat{\theta}_5$$

$$Y_2 \sim \hat{\theta}_1 + \hat{\theta}_2 + \hat{\theta}_3 + \hat{\theta}_4 + \hat{\theta}_5$$

- Effect of normative on correct selection rate ($\boldsymbol{\theta} > 0$).

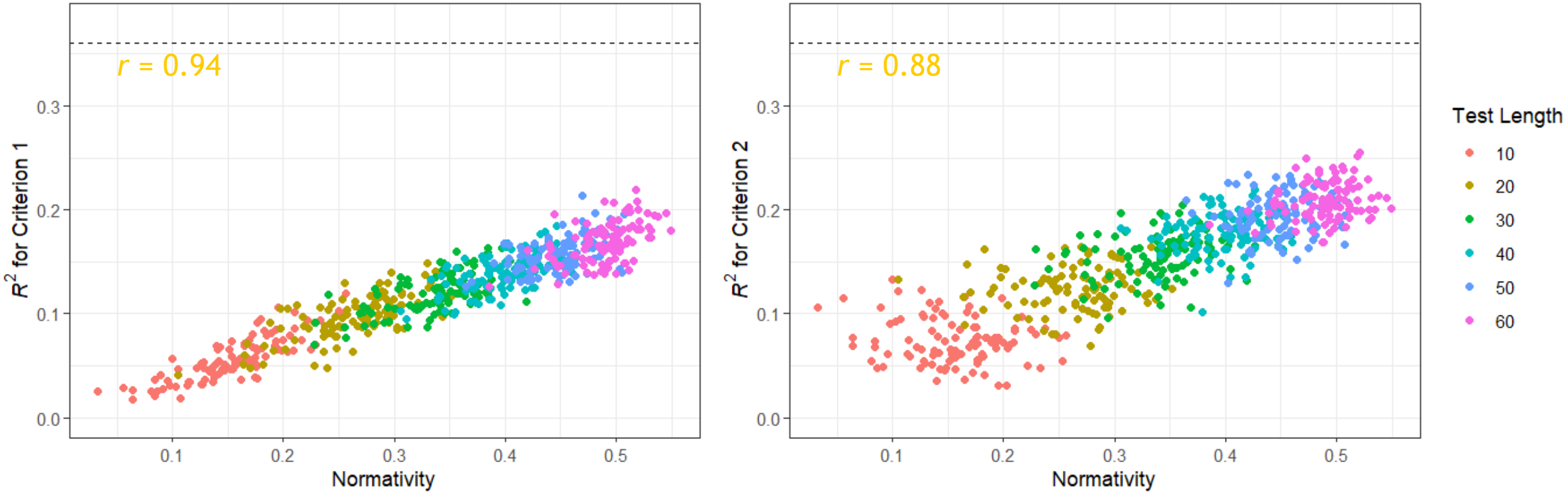
Results:

- Recovery of true normativity:



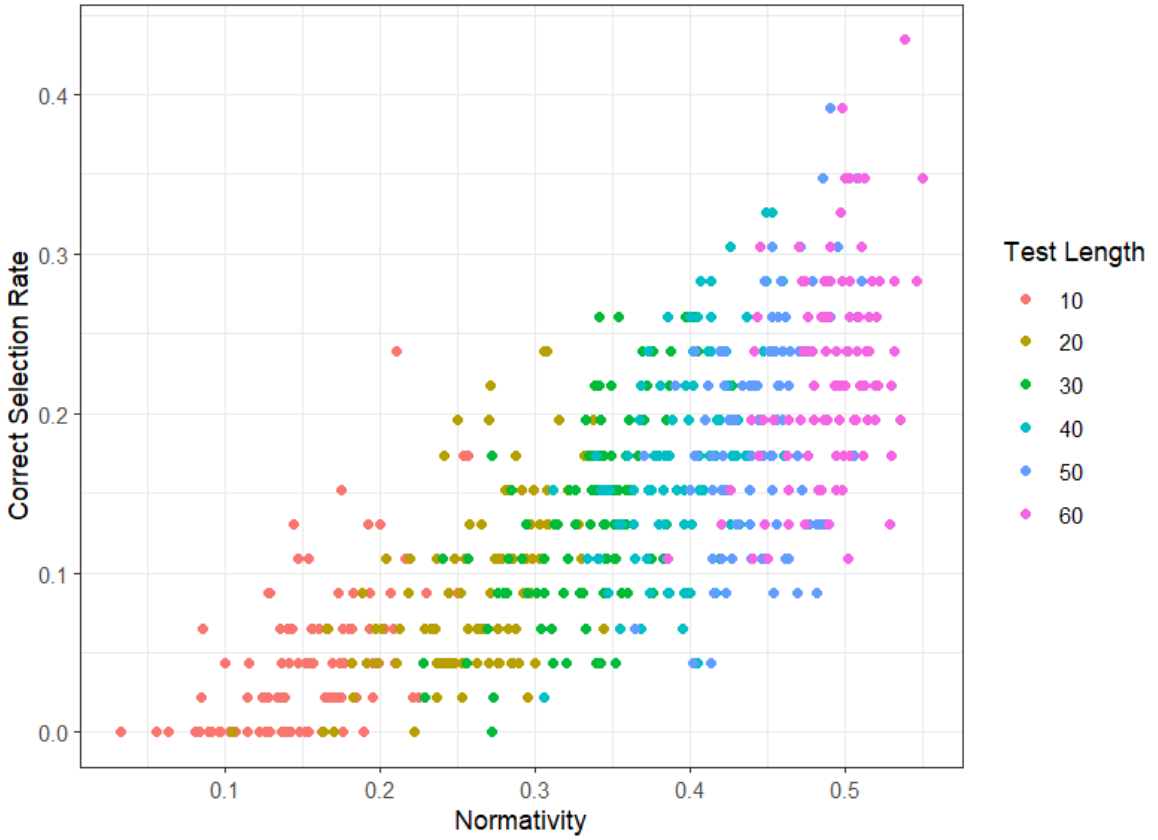
Results:

- Effect of normativity on multiple R-squared with criteria:



Results:

- Effect of normativity on correct selection rate ($\boldsymbol{\theta} > 0$):



Conclusions:

- Calculation of normativity indices offer good recovery of the squared correlation between true and estimated within-person means.
- Normativity has a positive effect on external validity estimates, especially when criteria are related to multiple traits (e.g., life satisfaction).
- Normativity has a positive effect on correct selection rates (r = 0.78), more than reliability alone (r = 0.72).

Thank you!
¡Muchas gracias!

For more information:

Rodrigo S. Kreitchmann
rschames@psi.uned.es

PID2022-137258NB-100 funded by MICIU/AEI /10.13039/501100011033, by ESF+ and by UAM-IIC Chair of Psychometric Models and Applications

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES

Cofinanciado por la Unión Europea

AGENCIA ESTATAL DE INVESTIGACIÓN