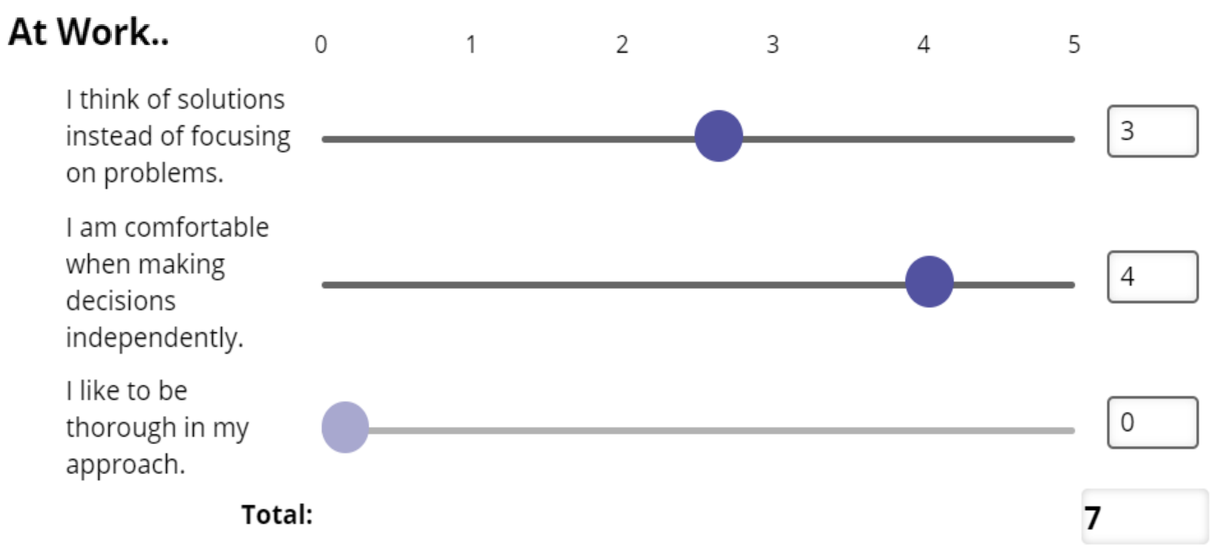
DETECTING CARELESS RESPONDING IN IPSATIVE DATA

Professor Anna Brown

University of Kent, Canterbury, UK

'IPSATIVE' RESPONSE FORMATS



- Impossible to endorse all desirable alternatives
- Facilitate differentiation and "slow" thinking (Kahneman, 2011)
- Popular since proper scaling methods have become available, e.g. Thurstonian models family (Brown & Maydeu-Olivares, 2011; 2013; 2018; Brown; 2016a; Brown; 2016b)
  - Normative trait scores can be obtained from ipsative data

CARELESSNESS IN IPSATIVE ASSESSMENTS

- Like any other questionnaires, ipsative questionnaires can be subject to careless responding when respondents are not sufficiently motivated to give their full attention to the questions.
- However, detecting such responding can be more challenging than when using Likert scales
  - modelling of ipsative responses is inherently multidimensional;
  - method factors need to take to account the comparative nature of ipsative responses

OBJECTIVES

To describe and evaluate two alternative strategies for dealing with careless responses in ipsative data:

(1) identifying (and ultimately removing from the sample) careless responders using 'person fit' indices designed for ipsative formats;

(2) controlling for careless responding using method factors embedded in the Thurstonian IRT model (Brown & Maydeu-Olivares, 2012).

EMPIRICAL STUDY

- Bespoke questionnaire for assessing applicants to public sector jobs in the UK
  - measures 24 non-cognitive skills, covering the Big Five domains
  - consists of 276 multidimensional 'graded response' pairs

I work effectively without getting distracted

Much More

A Little More

Equally

A Little More

Much More

I express myself confidently

- Data can be analysed as ordinal or continuous

- Sample.
  - N=1,388 volunteers who participated in a trial

‘PERSON FIT’ INDICES

- Test takers should express preferences in line with their trait scores; for example, if they are higher on trait A than on trait B, they should prefer items measuring A over items measuring B **consistently** (adjusted for item parameters)
- Comparing observed responses with responses expected under the measurement model
  - We know observed responses to pair of items $\{a,b\}$ for person i

    $\text{Observed}_{\{a.b\}i}$ (for example, =4)
  - We compute observed response according to the Thurstonian model

    $\text{Expected}_{\{a.b\}i} = \text{intercept}_{\{a.b\}} + \text{loading}_{\{a\}} * \text{TraitA}_i - \text{loading}_{\{b\}} * \text{TraitB}_i$
- For each test taker, ‘fit’ between their observed responses and their expected responses are measured by summarising either:
  - Discrepancies
  - Concordance

PERSON FIT AS A MEASURE OF DISCREPANCIES

- Ferrando (2010) proposed a simple person-fit statistic for linear factor models (also known as "congeneric"), "***lco***"
  - Summary of squared differences of observed and expected responses

$$lco_i = \sum_{j}^{n} \frac{\left(X_{ij} - \mu_j - \lambda_j \hat{\theta}_i(\mathrm{ML})\right)^2}{\sigma^2{}_{\varepsilon j}}$$

- ***lco*** can be easily extended to Thurstonian factor model
  - The numerator is simply (observed-expected)^2
  - The denominator (error variance) is constant for all pairs, so can be omitted
  - I suggest computing the mean across 276 pairs rather than the sum

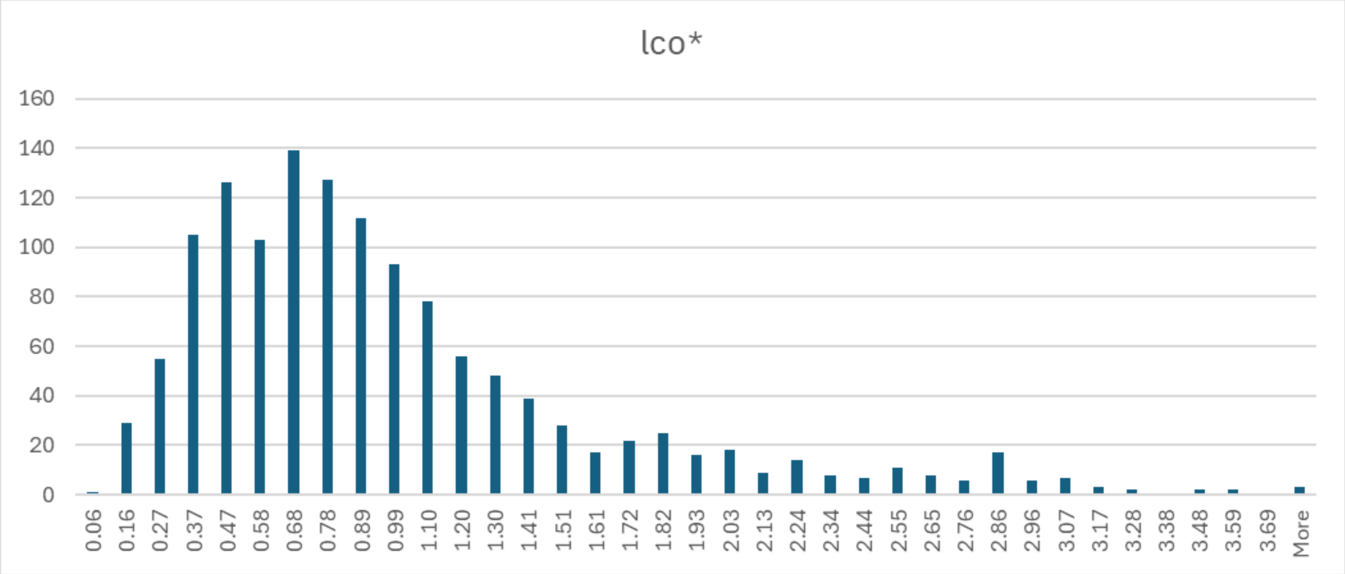
$$lco_i{}^* = \mathrm{MEAN}(\mathrm{Observed}_{\{a.b\}i} \text{ - } \mathrm{Expected}_{\{a.b\}i})^2$$

DISTRIBUTION OF *LCO**

- For the trial sample (<u>before</u> any data cleaning!)
  - We have a long positive tail of outliers (those with large misfit to model)

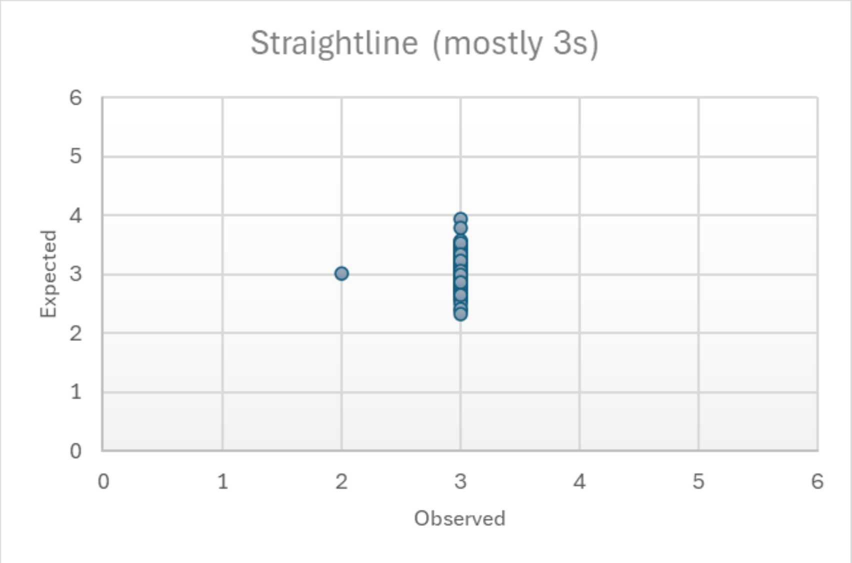
  - Median = 0.770

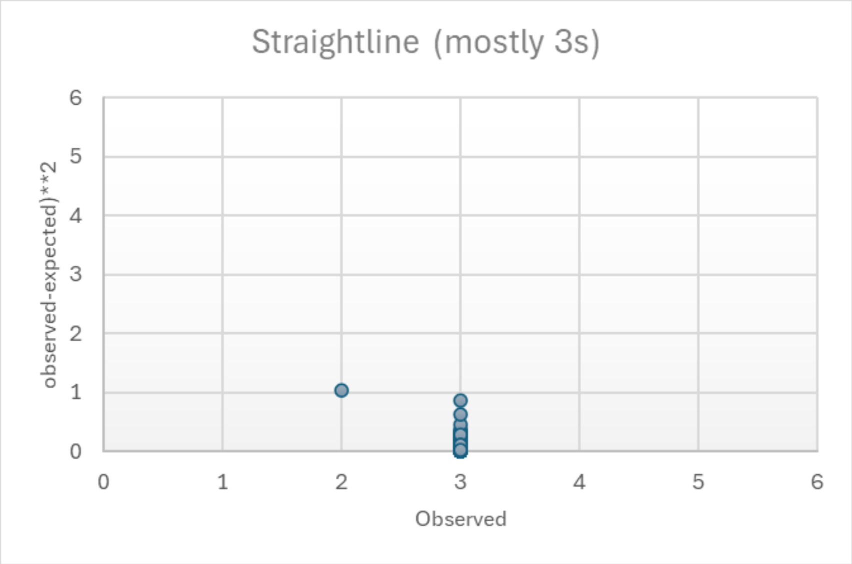
  - 5th percentile = 0.247
  - 10th percentile = 0.311
  - 90th percentile = 1.85
  - 95th percentile = 2.46



PROBLEM

- With polytomous items, candidates providing **midpoint** responses will obtain **average** trait scores
  - Their expected responses will also show central tendency, and will be very similar to their observed responses
- Index *lco** will be very small, suggesting perfect person-fit
- This problem is also described in Ferrando (2010); in single scales, any response pattern using only one or two adjacent categories will have this problem
  - In our case, only patterns with predominant response "3" are problematic





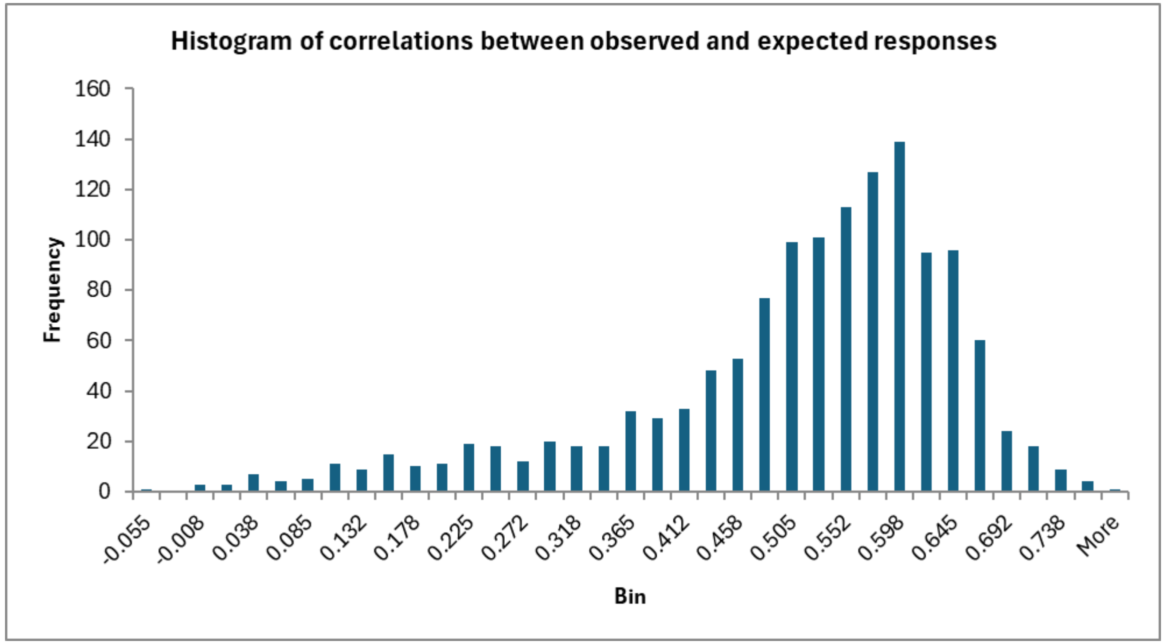
PERSON FIT AS A MEASURE OF CONCORDANCE

- Pearson's **correlation** can be computed to capture concordance between observed and expected responses, for each candidate

  - For genuine responses, the correlation should be positive and high
  - For random responses, the correlation should be near zero
  - For central tendency responses, the correlation should be near zero

DISTRIBUTION OF *COR(OBS, EXP)*

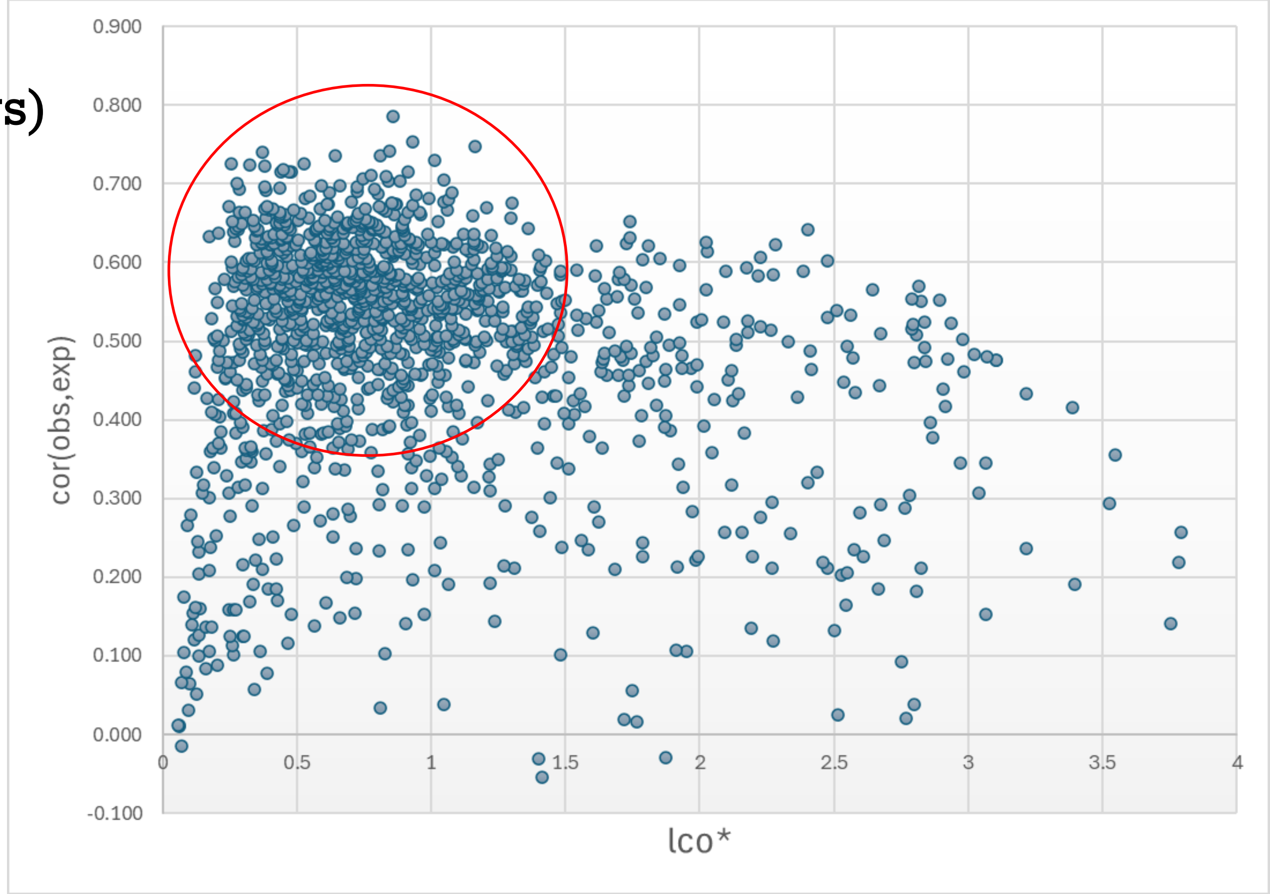
- For our trial sample (before any data cleaning)
  - We have a long negative tail (those with small concordance with the model)
  - Median(cor) = 0.532
  - 5^{th} percentile = 0.170
  - 10^{th} percentile = 0.280
  - 90^{th} percentile = 0.642
  - 95^{th} percentile = 0.664



DO THE PERSON FIT INDICES AGREE?

- Not really (only for careful responders)

- They complement each other for detecting careless responders, who have either:
  - high *lco**
  - or low *cor(obs,exp)*
  - or both



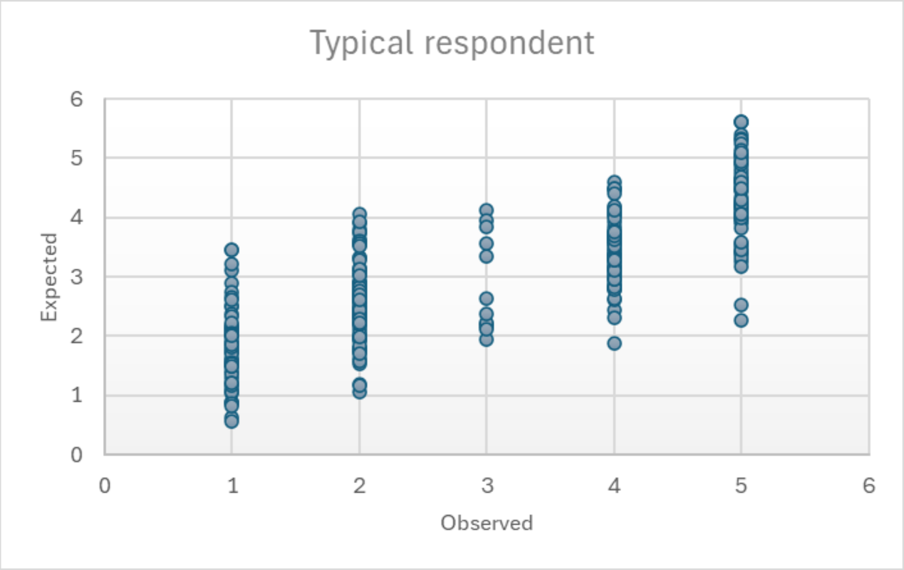
EXAMPLE: TYPICAL *LCO** AND *COR*

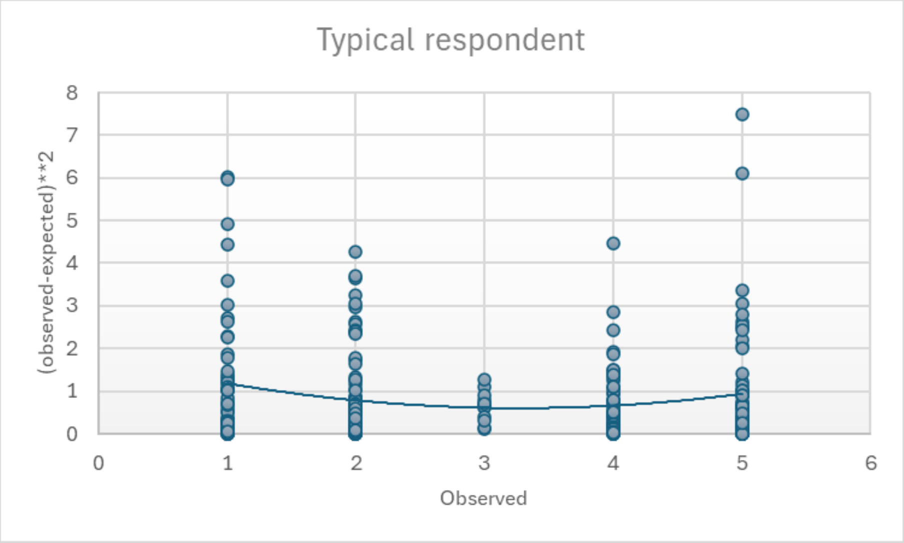
MEAN_RESPONSE 3.01

SD_RESPONSE 1.49

lco* 0.858

cor(obs, exp) 0.785

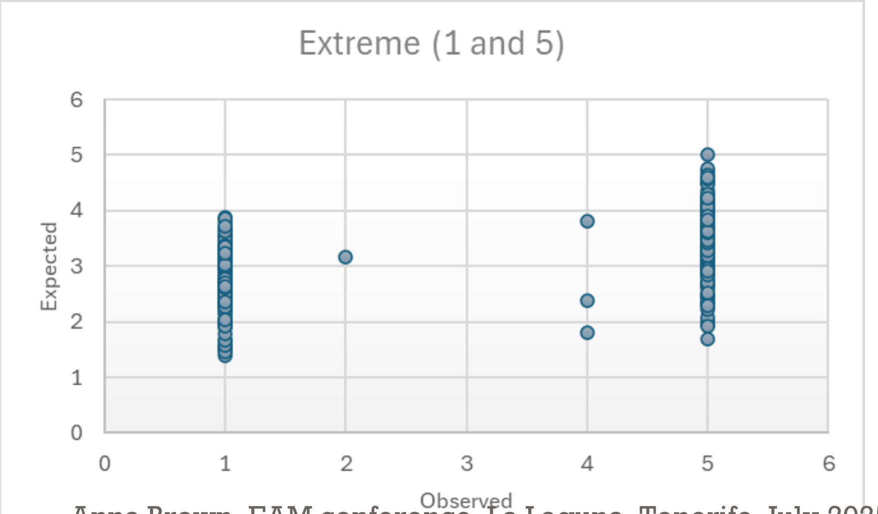


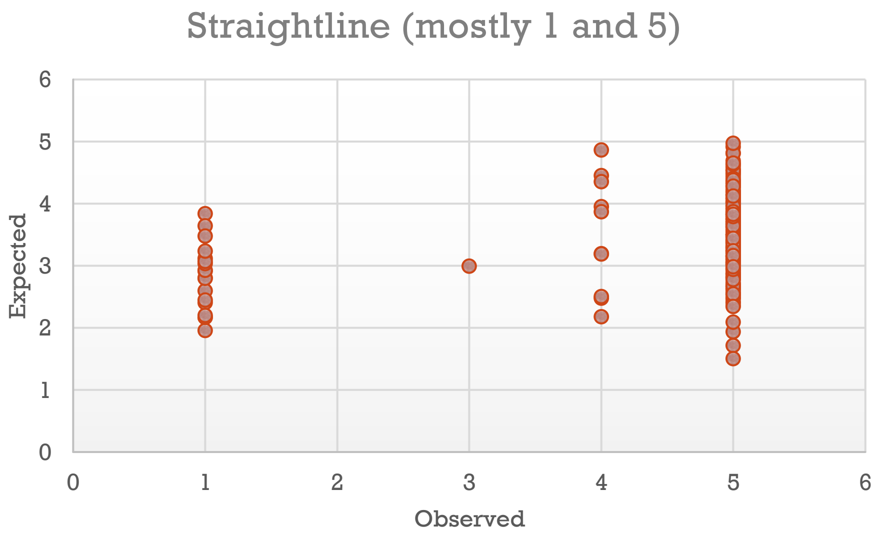


EXAMPLE: OUTLIERS ACCORDING TO *LCO**

MEAN_RESPONSE	3.17
SD_RESPONSE	1.99
lco*	3.105
cor(obs, exp)	0.475

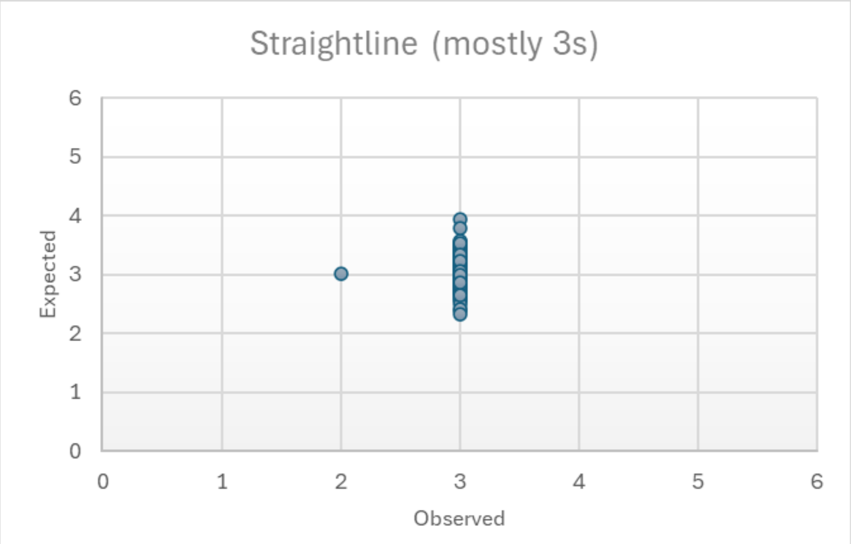
MEAN_RESPONSE	4.7
SD_RESPONSE	1.00
lco*	2.687
cor(obs, exp)	0.247



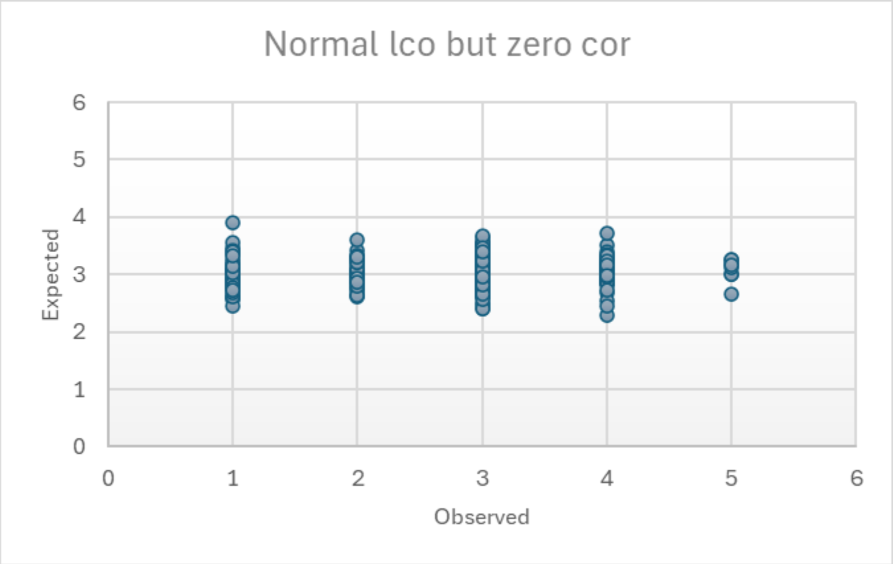


EXAMPLE: OUTLIERS ACCORDING TO *COR(OBS,EXP)*

MEAN_RESPONSE	3.00
SD_RESPONSE	0.06
lco*	0.058
cor(obs, exp)	0.011



MEAN_RESPONSE	2.85
SD_RESPONSE	0.99
lco*	1.040
cor(obs, exp)	0.038



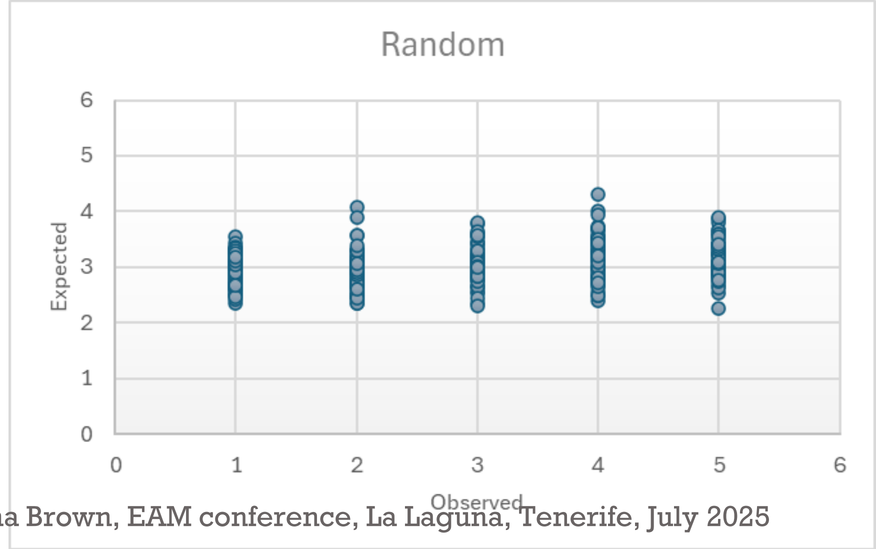
EXAMPLE: RANDOM RESPONSES

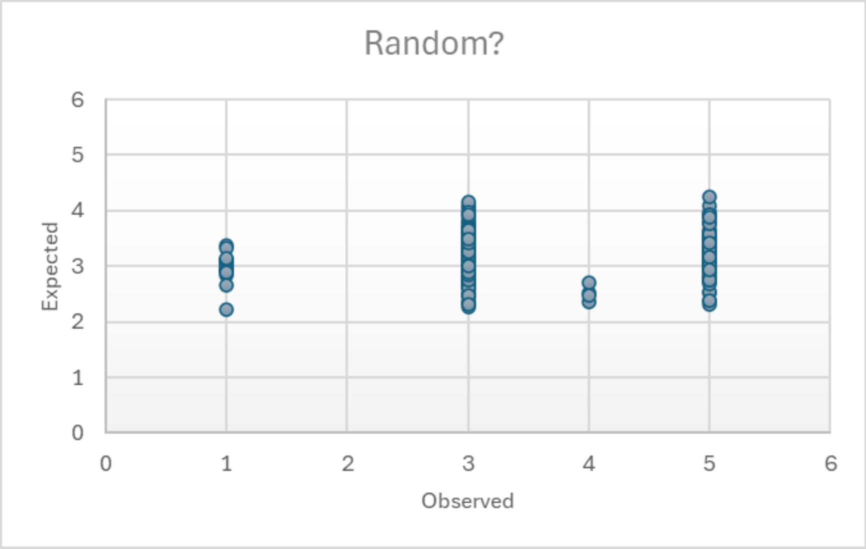
- Computer generated

MEAN_RESPONSE	2.98
SD_RESPONSE	1.419
lco*	1.897
cor(obs, exp)	0.242

- Respondent generated

MEAN_RESPONSE	3.51
SD_RESPONSE	1.11
lco*	1.221
cor(obs, exp)	0.311





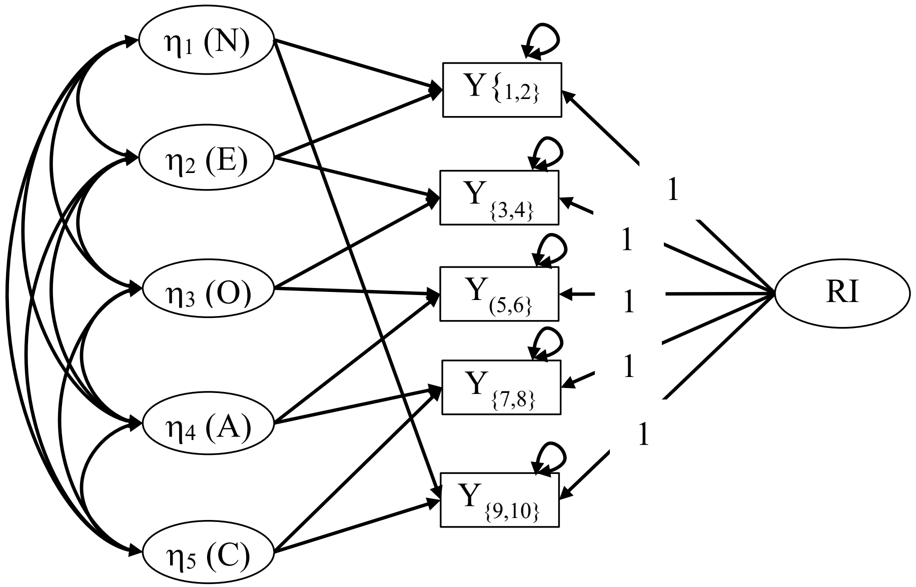
'METHOD' FACTOR

- A 'random intercept' can be added to the Thurstonian factor model to control carelessness expressed as overusing one response option

$$\text{Expected}_{\{a.b\}i} =$$

$$= \text{intercept}_{\{a.b\}} + \text{rand.intercept}_i +$$

$$+ \text{loading}_{\{a\}} * \text{TraitA}_i - \text{loading}_{\{b\}} * \text{TraitB}_i$$



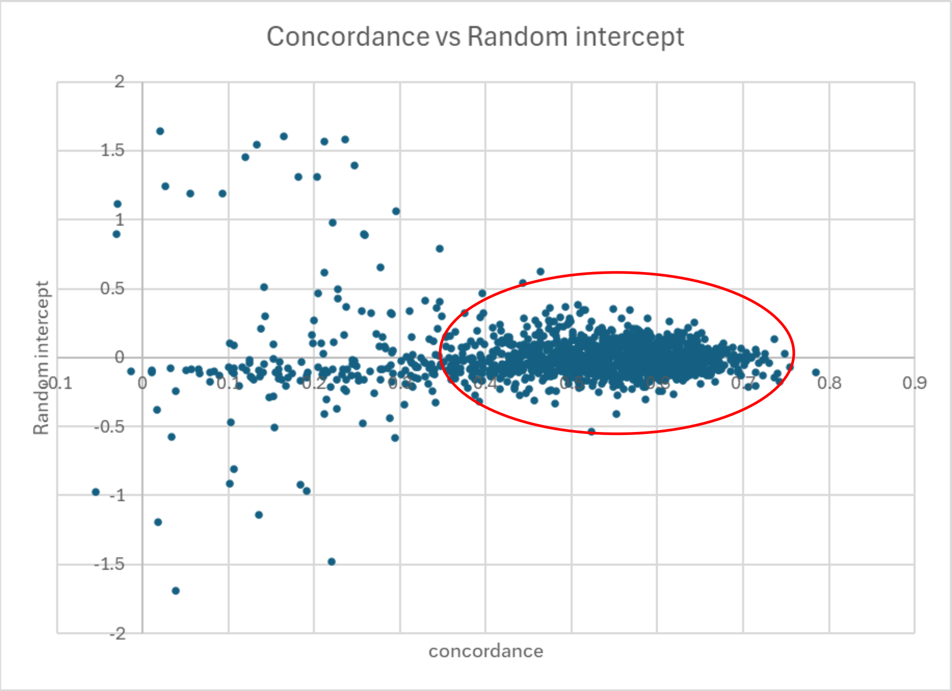
RANDOM INTERCEPT MODEL RESULTS

- The RI factor had variance 0.054 ($p < .001$)
  - 5.4% of the substantive traits' variances
  - explained approx. 4% variance of observed responses
- Goodness of fit
  - baseline Thurstonian model (N=1,388): SRMR = .068
  - Thurstonian model with RI (N=1,388): SRMR = .055
  - baseline Thurstonian model without careless responders detected with *lco** and *cor* (N=1,245): SRMR = .063

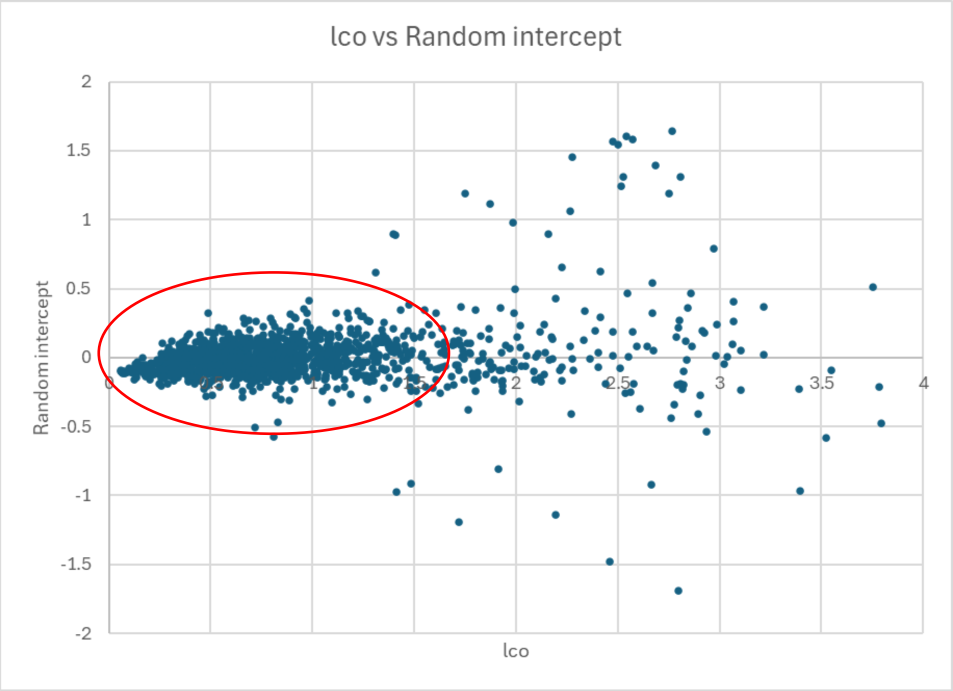
DO PERSON FIT INDICES AND RI AGREE?

Not really (only for careful responders)

corr(cor, RI) = -.052



corr(lco, RI) = .190



CONCLUSIONS

- No single index is 100% effective at detecting all types of careless responses
- At the individual level, the measures of discrepancy, concordance and random intercept agreed only for careful responders
  - "*All happy families are alike; each unhappy family is unhappy in its own way*" (L. Tolstoy)
- Combination of *lco** and *cor(obs,exp)* work for detecting
  - Random responding
  - Straight-lining (any category)
  - Over-using a category or several categories
- Method (RI) factor works for detecting (and controlling for)
  - Over-using one category

RECOMMENDATIONS

- Determine cut-offs for *lco** and *cor* indices empirically on your data
- Implement these cut-offs for flagging careless responders
- But consider also implementing simple prevention measures during the test administration, for example:
  - Allow only a certain proportion of responses in certain category, for example, no more than 20% of "equally true" (the middle category)
  - Warn the test taker that they should not select too many responses in the middle category, and when they exceed the limit, warn them that their profile will be void

THANK YOU!
ANY COMMENTS OR QUESTIONS?

a.a.brown@kent.ac.uk

http://annabrown.name