Reduce randomly guessing effects in a university generic skills test

Luc T Le, Australian Council for Educational Research (ACER)

Available at: https://works.bepress.com/luc_le/34/
Reduce randomly guessing effects in a university generic skills test

Luc Le

ITC 2018
Montreal, CA, July 2018

Australian Council for Educational Research
Presentation Outline

• Background on STAT
• Rational & Research Question
• Data
• Analysis
• Results
• Summary
What is STAT?

- Special Tertiary Admissions Test
- Contains 70 multiple choice items to be completed in 2 hours
  - 35 Verbal Reasoning (32 core + 3 trial)
  - 35 Quantitative Reasoning (32 core + 3 trial)
- Rasch model (Rasch, 1966/1980) has been used to calibrate items and measure candidate ability
Data Tailoring

- **Idea**: responses from lower ability candidates to a hard item could be considered as randomly guessing
- Remove such responses from item analysis
- **Example**: with a MC item with 4 options, when the model probability to answer correctly to an item <0.25, the responses could be considered as randomly guessing and being removed from item analysis

**Reference source**:
- Andrich D., Marais I., Humphry S. (2012)
- RUMM 2030 (Andrich et al, 2012);
- Winsteps (Linacre, 2012)
Research Question

Can Tailoring data be effective for item difficulty estimation in STAT and in other similar generic skills tests?
Data

- STAT 2016-2017
- 6462 Australian candidates
- About 50% females and 50% males
- Age: median = 22
Analysis

- **Tailoring data**: remove all responses s.t. the corresponding probability for such person answers correctly to the item is less than 25%.
- **Calibrating items** (Rasch model) using ACER ConQuest
- **Comparing** Rasch item difficulty estimates from Tailoring data with those from all response data.
- **Simulation**: 3 sets of responses to 32 items with 100 replications: with no randomly guessing; with randomly guessing responses, and with Tailoring data
Results
Mean ability by suspected randomly guessing – Quantitative reasoning
Mean ability by suspected randomly guessing–Verbal reasoning
Item difficulty estimate: Tailoring data vs All - Quantitative reasoning

![Graph showing item difficulty estimates](image)
Item difficulty estimate: All vs. Tailoring data - Verbal reasoning
Simulations and analysis

• Generating 32 dichotomously items with difficulty from [-1, 1]

• In each replication, a sample of 5000 ability is randomly drawn from a standard normal distribution N(0,1)

• Set 1 (general): generating responses to the 32 items base on the Rasch model

• Set 2 (Randomly guessing): With Set 1 data, randomly assign to 0 or 1 (prob.=3/4, 1/4) for any responses s.t. $\theta_n - \delta_i \leq -1.10$ logits

• Set 3 (Tailoring Randomly guessing): removing all guessing responses from Set 2

• 100 replications
Recovery of item difficulty estimates
Bias of item difficulty estimate

![Graph showing bias of item difficulty estimate]
Summary

• There is a small change in item difficulty estimates when Tailoring response data in STAT

• Tailoring data could help to remove randomly guessing responses and improve the Rasch item difficulty estimation

• However, Tailoring such responses that are not random guessing would lead to be more biased in estimation

• More constrained conditions for data Tailoring
Thank You