Does Our Instrument Include the Right Response Options?
Empirical Evidence to Evaluate Floor and Ceiling Effects for Multi-Item Ordinal Instruments

MSR23

Nicolai D. Ayasse, Fraser D. Bocell, Cheryl D. Coon
Clinical Outcome Assessment Program, Critical Path Institute

CLINICAL OUTCOME ASSESSMENT PROGRAM

CRITICAL PATH INSTITUTE 20 YEARS OF IMPACT

Introduction

- **Floor or ceiling (collectively "scale attenuation") effects** in an ordinal item or instrument are marked by a *large percentage of participants* (a) *endorsing the lowest or highest response option, or* (b) *assigned the lowest or highest possible score value*.
- They are problematic when observed due to items or response options not adequately capturing the range of the latent variable (θ) intended to be measured.
- Although evaluation for scale attenuation effects is standard practice when psychometrically evaluating an instrument, there is ***no standard reference*** agreed upon in the field to define them.

Objective:
- To generate empirical evidence to inform the evaluation of scale attenuation effects, via simulation study.

Conclusions

- It is important to account for *key characteristics of the instrument and items*, i.e., the ***number of response categories*** and ***number of items*** in the instrument, when examining the sample proportion assigned the lowest or highest possible score value or endorsing the lowest or highest response option.

Score-Level Proposed Reference Range:

$$p_{AVZ} = \frac{1}{i(j-1)+1} \textbf{ to } \frac{1}{i(j-1)+1} \times 1.25$$

where p_{AVZ} is the sample proportion assigned the lowest (*A*) or highest (*Z*) possible score value, *i* is the number of items in the instrument, *j* is the number of response categories

- *An alternative metric* (possibly more sensitive) is to consider the sample proportion assigned the highest or lowest ~10% of possible score values.

Item-Level Proposed Reference Range:

$$p_{avz} = \frac{1}{j} \textbf{ to } \frac{1}{j-2}$$

where p_{avz} is the sample proportion endorsing the lowest (*a*) or highest (*z*) response option, and *j* is the number of response categories

Recommendations:

- Instruments meeting these criteria should be examined carefully, although *meeting the criteria alone should not be taken as definitive evidence of problematic attenuation effects*.
  - *Note*: This simulation assumes θ is normally distributed
- Instruments are designed with item locations spread along the target population's θ range, so observing possible attenuation effects in ~1-2 items without considering broader context is not necessarily informative.

Early in development: Examine individual items for possible attenuation effects.
- If observed: Do you have the right response options?

Later in development:

1. Examine the *instrument scores* for possible score-level attenuation effects.
   - If observed: Do you have the right items in the right θ range to adequately represent your target population on the concept of interest?
2. Examine the *individual items* for possible item-level attenuation effects.
   - If observed: Are there specific items that might be particularly problematic?

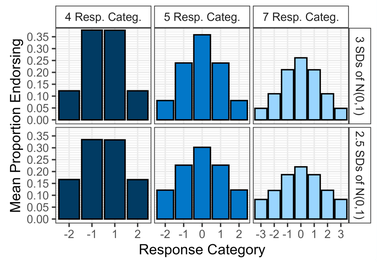
Methods

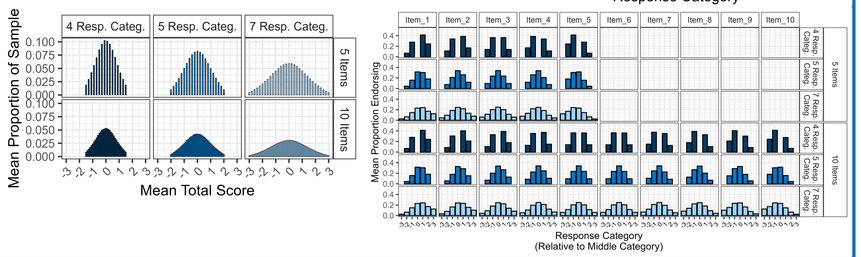
Simulation Study Design

Replications: 500 — Number of (ordinal) items in instrument: 5 or 10
Sample size: N=500 — Number of response categories for ordinal items: 4, 5, or 7
Simulate normal latent variable (θ) with population mean μ=0 and standard deviation σ=1

Control Conditions

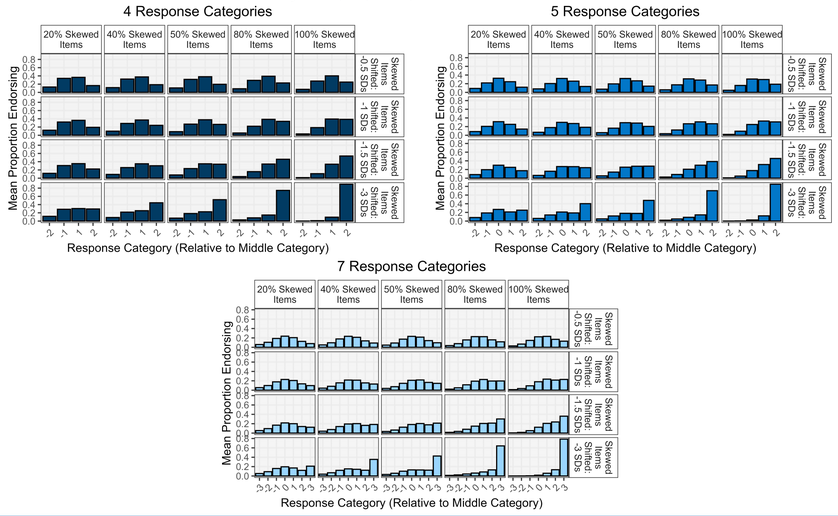
- Item response thresholds are centered at 0 across the instrument
- Two sets of mean thresholds chosen based on evenly dividing 2.5 and 3 standard deviations of the normal distribution (N[0,1]):
  - 4 responses: -1.5, 0, 1.5 *and* -1.25, 0, 1.25
  - 5 responses: -1.8, -0.6, 0.6, 1.8 *and* -1.5, -0.5, 0.5, 1.5
  - 7 responses: -2.14,-1.29,-0.43, 0.43, 1.29, 2.14 *and* -1.79,-1.07,-0.36, 0.36, 1.07, 1.79





Experimental Conditions

- Shift item intercepts for a given percentage of items ("skewed items")
- Shift center of item intercepts by -0.5, -1, -1.5, -3 σ



Analysis

- The bias of the resulting instrument total scores was examined (total score = mean of items)

$$Bias = Score_{Observed} - \theta$$

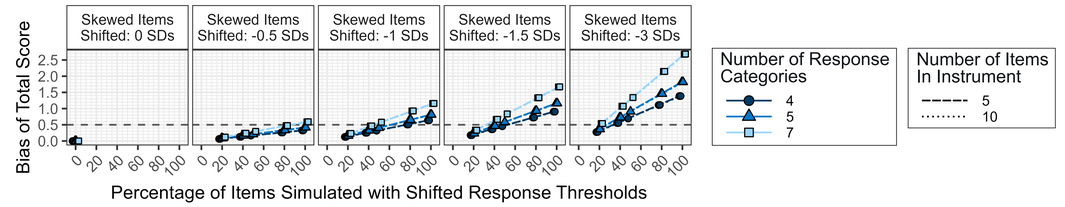
- Bias threshold set at ½ σ:
  - Reference: Observed sample proportion endorsing the highest response category, or assigned the highest possible score value (*alternatively*: highest ~10% of possible score values, here ranging 9.5-12.5%), corresponding to when the total score bias exceeded ½ σ
- For 4, 5, or 7 responses: we incrementally tested different sample proportions and examined their corresponding classification accuracy for identifying when total score bias exceeded ½ σ
  - Proposed reference range errs on the side of specificity (as opposed to sensitivity)

Acknowledgements

Critical Path Institute is supported by the Food and Drug Administration (FDA) of the Department of Health and Human Services (HHS) and is 56% funded by the FDA/HHS, totaling $23,740,424, and 44% funded by non-government source(s), totaling $18,881,611. The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement by, FDA/HHS or the U.S. Government.

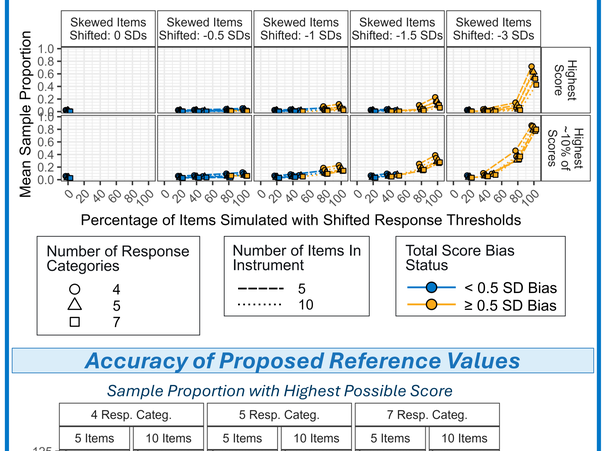
Results

Total Score Bias



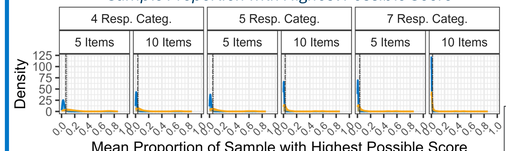
Score-Level Results

Proportion Assigned Highest Possible Score Value(s)

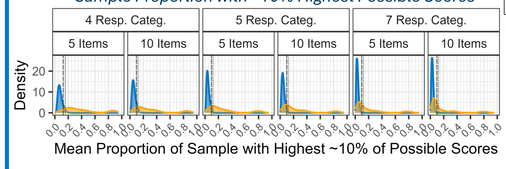


Accuracy of Proposed Reference Values

Sample Proportion with Highest Possible Score

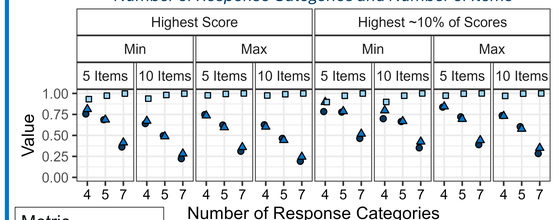


Sample Proportion with ~10% Highest Possible Scores

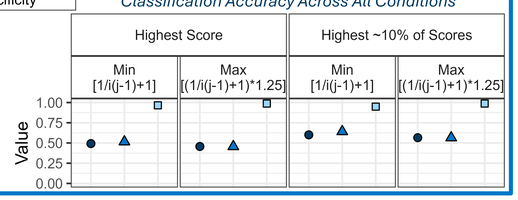


(Vertical lines indicate the proposed reference values)

Classification Accuracy by Number of Response Categories and Number of Items

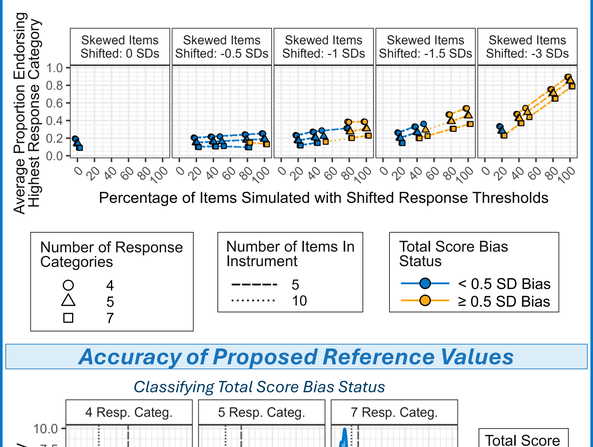


Classification Accuracy Across All Conditions



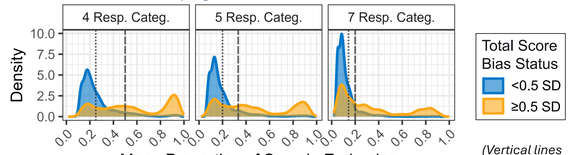
Item-Level Results

Proportion Endorsing Highest Response Option



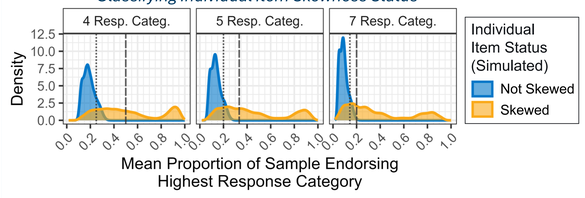
Accuracy of Proposed Reference Values

Classifying Total Score Bias Status

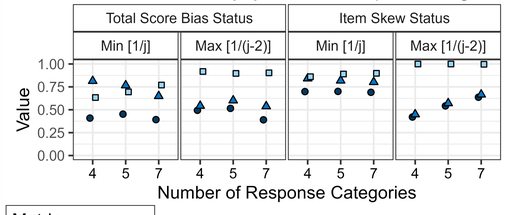


(Vertical lines indicate the proposed reference values)

Classifying Individual Item Skewness Status



Classification Accuracy by Number of Response Categories



Classification Accuracy Across All Conditions