INTRODUCTION

Across the globe, testing agencies have a common goal of creating items that maintain measurement invariance and are thus free of group characteristic biases (Khorramdel et al., 2020). Even though test agencies’ efforts to consult with local talents to create cultural and gender relevant items in LMICs, there is evidence that some of these items are biased and may be unfair to some students based on their economic status, gender, and location. We examine the cultural and gender relevant items included on the cognitive assessments for the Young Lives school effectiveness survey data from Ethiopia (2016/2017) (Boyden et al., 2016; Rossiter et al., 2017).

AIM

The purpose of this paper is to examine how gender and location (rural vs urban) may impact item quality for assessments administered in low-to-middle income countries (LMICs). This study examines:

1. The item fit statistics and item targeting of the YL cognitive assessments for Ethiopia under the Rasch model.
2. Differential item functioning (DIF) by gender and location for the cognitive assessment data.

METHODS

- Item responses to the cognitive assessments were analyzed using the Rasch model. This is consistent with the hypothesized unidimensional nature of this data (Rossiter et al., 2017; Linacre, 2022).
- Item fit statistics were examined to see how well the data fit the Rasch model, and Wright maps were utilized to clarify item targeting.
- We conducted differential item functioning (DIF) analysis to explore the possibility of item-level performance differences (Walker, 2011) based on gender and location.

RESULTS

Test-Item Targeting

The person-distributions for both the Math & English assessments showed the ability of students on the same measurement scale as the item difficulty; this pattern was acceptable but suggests that more difficult items should be added to the assessment for highly proficient students at the upper part of the scale.

Wright maps showing the person distribution with item labels

Differential Item Functioning

For both English and Mathematics items, we used the Rasch-Welsh Test and a significance level of 0.05 based on the absolute value of the DIF contrast (Linacre, 2022).

- Overall, there was no support for any individual items that favored boys or girls.
- However, for location (urban versus rural), 15% of the mathematics items showed moderate to high DIF (DIF ≥ 0.64 logits) and slight to moderate DIF (DIF ≥ 0.43 logits), (Zwick, Thayer & Lewis, 1999).
- The English test had 17.5% of items showing moderate to high DIF. For example, the first test item which asked students to identify a school bag, had a DIF contrast of .82. This item was harder for students in rural schools.

Examples of some English Items that exhibited location DIF

CONCLUSIONS

Based on the literature from LMIC, most large-scale test developers seem to focus on creating unbiased questions to avoid gender DIF than they do for location. Thus, it is no surprise that this same trend is seen within these YL data where the gender-based differences were negligible in terms of DIF but potentially biased items are indicated when location is considered.

E-mail Wilberforce.2@osu.edu with suggestions.

BIBLIOGRAPHY


ACKNOWLEDGEMENT

The data used in this poster come from Young Lives, a 20-year study of childhood poverty and transitions to adulthood in Ethiopia, India, Peru, and Vietnam (www.younglives.org.uk). Young Lives is funded by UK aid from the Foreign, Commonwealth & Development Office and a number of further funders. The views expressed here are those of the authors and are not necessarily those of Young Lives, the University of Oxford, FCDO or other funders.