

Differential Item Functioning: An Analysis of An Academic Resilience Scale

Valerie Ofori Aboah, Latif Kadir, Ann O'Connell (Ed.D.)

INTRODUCTION

- Differential Item Functioning (DIF) is a common measurement issue that plagues measurement scales, rendering results from such scales biased as it yields an unfair advantage to a particular group (Zumbo, 1999).
- Wang et al. (1994) define Academic Resilience as “the heightened likelihood of success in school and other life accomplishments, despite environmental adversities brought about by early traits, conditions, and experiences”.
- Academic resilience is positively related to self-efficacy (Martin & Marsh, 2006; Carlson, 2001) and high levels of on-task behavior (Waxman et al., 2012).
- This study analyzed the 5-item Academic Resilience Scale from the 2018 US PISA dataset using DIF and Differential Test Functioning (DTF).

RESEARCH QUESTION

Does the 5-item Academic Resilience Scale used in PISA 2018 possess DIF and DTF in terms of gender?

In other words, can the scale be used to compare academic resilience between genders?

SAMPLE & MEASURING SCALE

- Data Source: US 2018 PISA dataset
- N = 4838 (15-year-olds)
- 50.1% Male

The 5-point Academic Resilience Item

Numbers	Items
1.	I usually manage one way or another.
2.	I feel proud that I have accomplished things.
3.	I feel that I can handle many things at a time.
4.	My belief in myself gets me through hard times.
5.	When I'm in a difficult situation, I can usually find my way out of it.

The Academic Resilience item used a 4-point Likert scale ranging from 1 (Strongly Disagree) to 4 (Strongly Agree).

METHOD

- The Item Response Theory (IRT) likelihood ratio test was used to conduct the DIF analysis
- P-values were adjusted using the Benjamini–Hochberg's procedure
- The Graded Response Model was employed for both the reference (females) and focal (males) groups (Desjardins & Bulut, 2018)
- Effect sizes were computed according to the recommendations by Meade (2010)

RESULTS

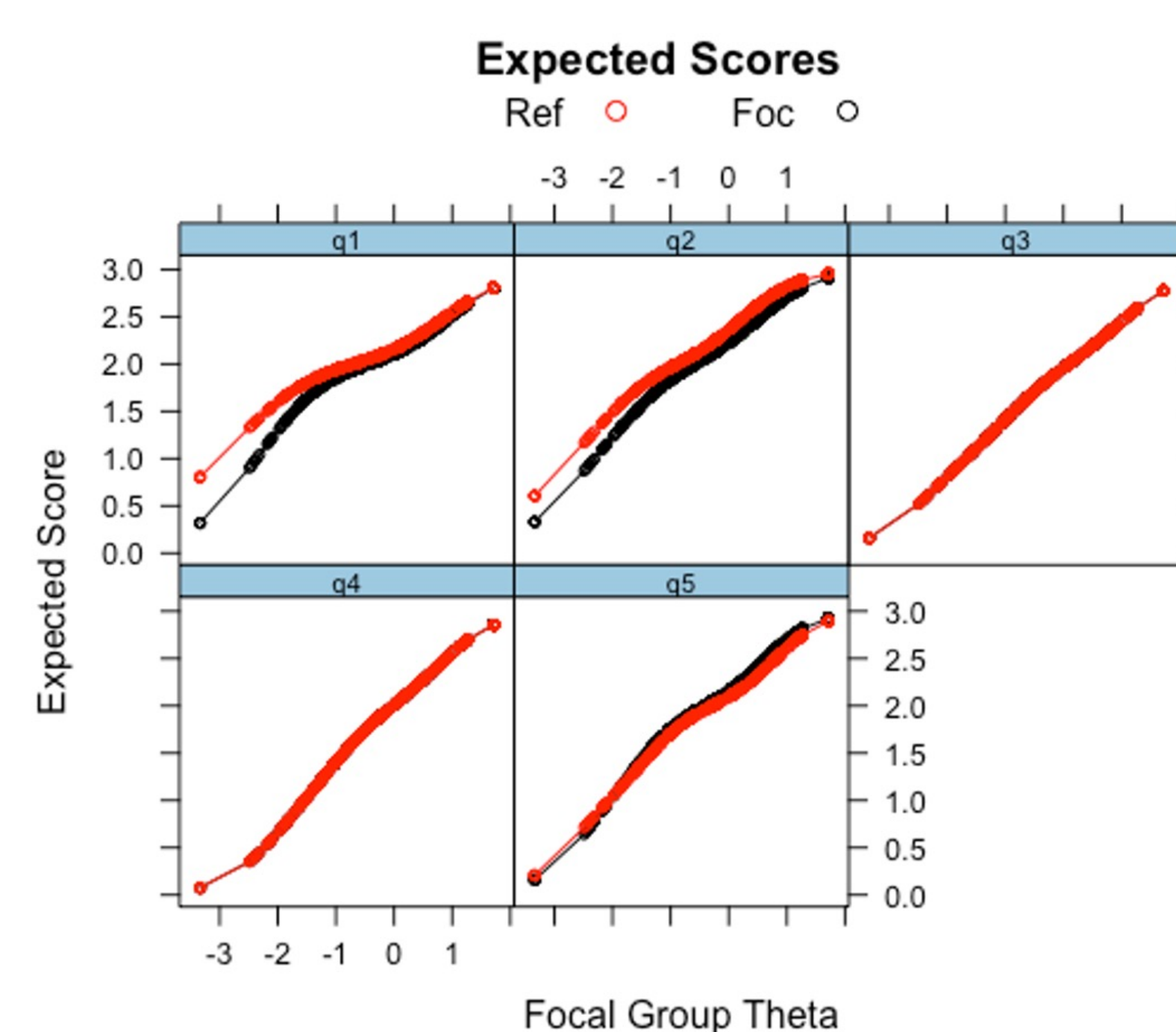
GRM Model Fit Indexes (One Factor Model)

Group	N	RMSEA	RMSEA 90% CI	SRMSR	CFI
Females	2268	.08	[.065, .096]	.04	.98
Males	2280	.09	[.078, .110]	.05	.98

Two-Stage Likelihood Ratio Test

Items	All-others-as anchors model		Anchor-item model (2 nd round)	
	G ²	BH _p	G ²	BH _p
1. I usually manage one way or another.	24.91	<.001	29.23	<.001
2. I feel proud that I have accomplished things.	72.60	<.001	60.61	<.001
3. I feel that I can handle many things at a time.	7.84	.098	-	-
4. My belief in myself gets me through hard times.	15.95	.004	-	-
5. When I'm in a difficult situation, I can usually find my way out of it.	59.05	<.001	23.00	<.001

Item-level Expected Scores



- The scale is unidimensional
- Items 1, 2, and 5 showed DIF
- However, the effect sizes of these items were small
- The cumulative impact of the DIF, i.e., DTF, was very small (<.2)
- The scale-level effect size for gender as measured by the ETSSD was less than .2, indicating a small effect size according to Cohen (1988).

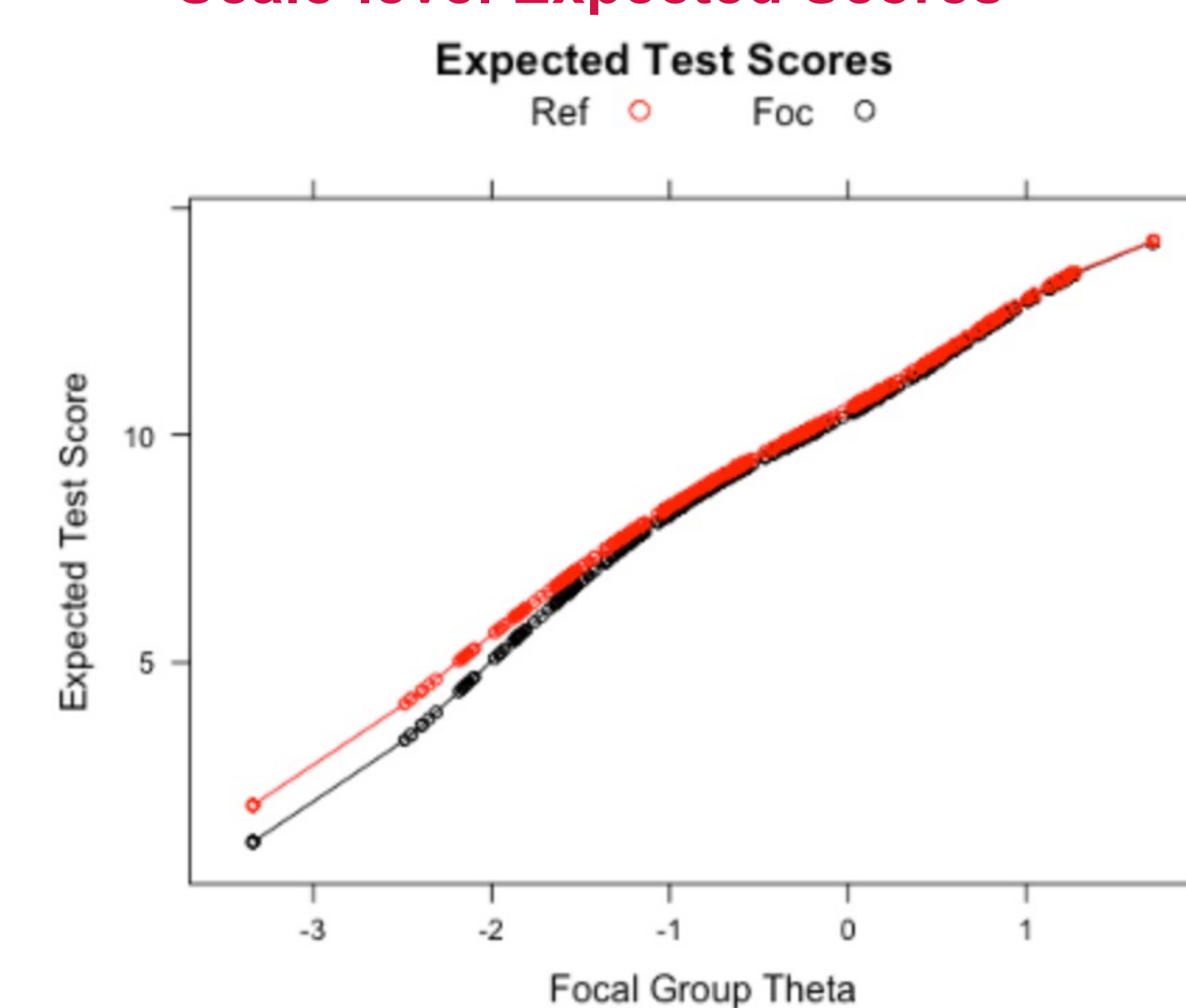
Item-level effect sizes

Items	SIDS	UIDS	ESSD
1. I usually manage one way or another.	.064	.064	.209
2. I feel proud that I have accomplished things.	.128	.128	.342
3. I feel that I can handle many things at a time.	-	-	-
4. My belief in myself gets me through hard times.	-	-	-
5. When I'm in a difficult situation, I can usually find my way out of it.	-.058	-.059	-.138

Scale-level effect sizes

	STDS	ETSSD
Gender	.134	.066

Scale-level Expected Scores



GRM: Graded Response Model
RMSEA: Root Mean Square Error of Approximation
CFI: Comparative Fix Index
SIDS: Signed Item Differences
UIDS: Unsigned Item Differences
ESSD: Expected Score Standardized Differences
STDS: Signed Test Difference in the Sample
ETSSD: Expected Test Score Standardized Difference

LIMITATIONS

- The analysis does not incorporate the complex nature of the sampling weights in the PISA dataset
- The use of only one method in detecting DIF/DTF is not highly recommended by scholars within the field.

CONCLUSIONS

Overall, the results show that the scale seems to operate similarly across gender on both the item and scale levels. Researchers can utilize the PISA Academic Resilience Scale for gender comparisons and can be confident that any disparities identified are genuine rather than mere artifacts of Differential Item Functioning (DIF).

Consequently, conclusions drawn from such gender comparisons can be considered robust and valid. Moreover, since IRT analysis results are independent of the sample, these findings apply to other populations as well, enhancing the generalizability of the study's conclusions.

E-mail oforiaboah.1@osu.edu / kadir.11@osu.edu with suggestions.

BIBLIOGRAPHY

- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57(1), 289-300.
- Carlson, D. J. (2001). Development and validation of a college resilience questionnaire (Publication No. 3016308) [Doctoral dissertation, University of Nebraska]. ProQuest Dissertations and Theses Global.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.
- Desjardins, C. D., & Bulut, O. (2018). *Handbook of Educational Measurement and Psychometrics using R*. CRC Press.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.
- Martin, A. J., & Marsh, H. W. (2006). Academic resilience and its psychological and educational correlates: A construct validity approach. *Psychology in the Schools*, 46(1), 53-83.
- Meade, A.W. (2010). A taxonomy of effect size measures for differential function of items and scale. *Journal of Applied Psychology*, 95(4), 728-743.
- PISA. (2018). PISA 2018 data. <https://www.oecd.org/pisa/data/2018database/>
- Wang, Haertal, G., & Walgberg, H. (1994). Educational resilience in inner-city. In M. Wang & G. E (Eds.), *Educational resilience in inner-city America: Challenges and prospects* (pp. 45 – 72). Hillsdale: Lawrence Erlbaum Associates.
- Waxman, H. C., Rivera, H., & Powers, R. (2012). English learners' educational resilience and classroom learning environment. *Educational Research Quarterly*, 35(4), 53-72.
- Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF)*. Ottawa: National Defense Headquarters, 160.