

Sex Bias in the ADOS: A Multi-Module Comparison of DIF Among 19K Participants from NDAR, SSC, SPARK, and AIC



Austin Wyman¹, Yuen Yvonne Yu², Chenhao Li¹,
Zhiyong Zhang¹, PhD, Caitlin C Clements¹, PhD

¹University of Notre Dame, Department of Psychology;

²Rutgers University, Graduate School of Applied and Professional Psychology



1. Background

- The Autism Diagnostic Observation Schedule (ADOS) is a standardized autism assessment, but it may be susceptible to clinician bias.
- Previous studies investigated sex differential item functioning (DIF) on the ADOS in samples ranging from 700 to 6,300 individuals:
 - Harrison et al.(2017) detected no sex DIF in items from modules 1 through 4.
 - Kalb et al. (2022) detected minimal sex DIF in items from modules 1 through 3.
 - Burrows et al. (2025) detected three sex DIF items across modules T, 1, and 2.
 - Yu et al. (2026) detected medium and large sex DIF in three module 3 items.
- This study is the largest examination of sex DIF in the ADOS and the only study across all modules.

3. Methods

- Likelihood Ratio Test was conducted on algorithm items for each module and language/age level separately (7 models).
- 2-factor graded response model analyzed social communication (SC) and restricted/repetitive behaviors and interests (RRBI).

4. Results + Discussion

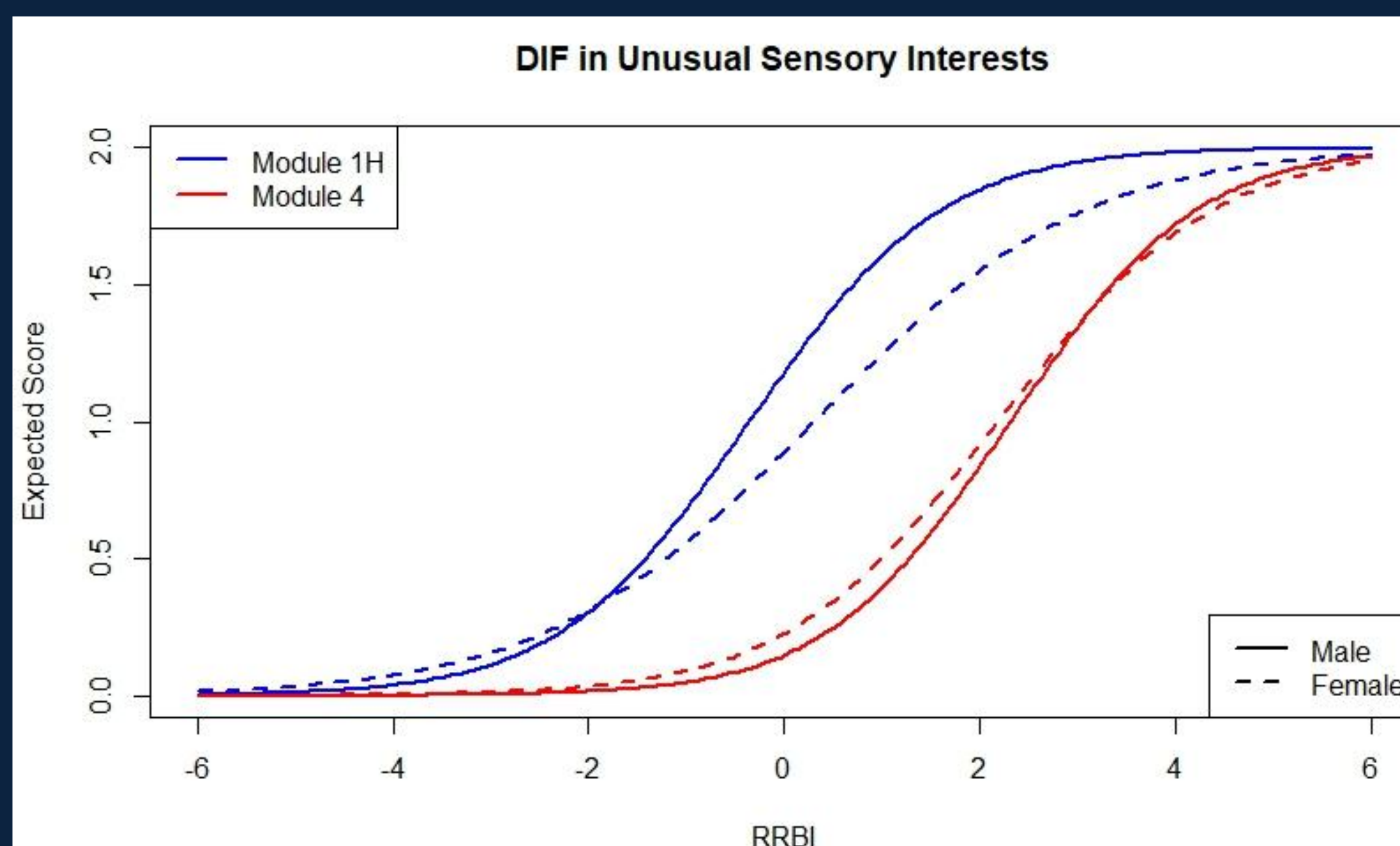
- 11 unique items exhibited significant DIF after correcting for type I error:
 - SC:** Requesting_T, Pointing_{1H}, Showing_{1H}, Facial Expressions to Examiner₄, Emotional Gestures₄, Communication of Own Affect₄, Insight₄, Quality of Social Overtures₄.
 - RRBI:** Unusual Sensory Interests_{1H/4}, Excessive Interests in Unusual Topics_{3/4}, Hand and Finger Mannerisms₄
- Effect sizes indicate that on Mod 4, females tend to be underscored on SC items and overscored on RRBI items, but the opposite pattern was observed for modules for younger ages or less language.
- The observed sex bias in Mod. 4 may reflect a different behavioral presentation in females (e.g., camouflaging), which is more prominent at older ages.

2. Sample

Data	Total	More Info:
SSC	n = 2,851	
SPARK	n = 2,257	
AIC	n = 1,715	
NDAR	n = 16,261	
	N = 23,084	

With valid sex and age data: Male Female

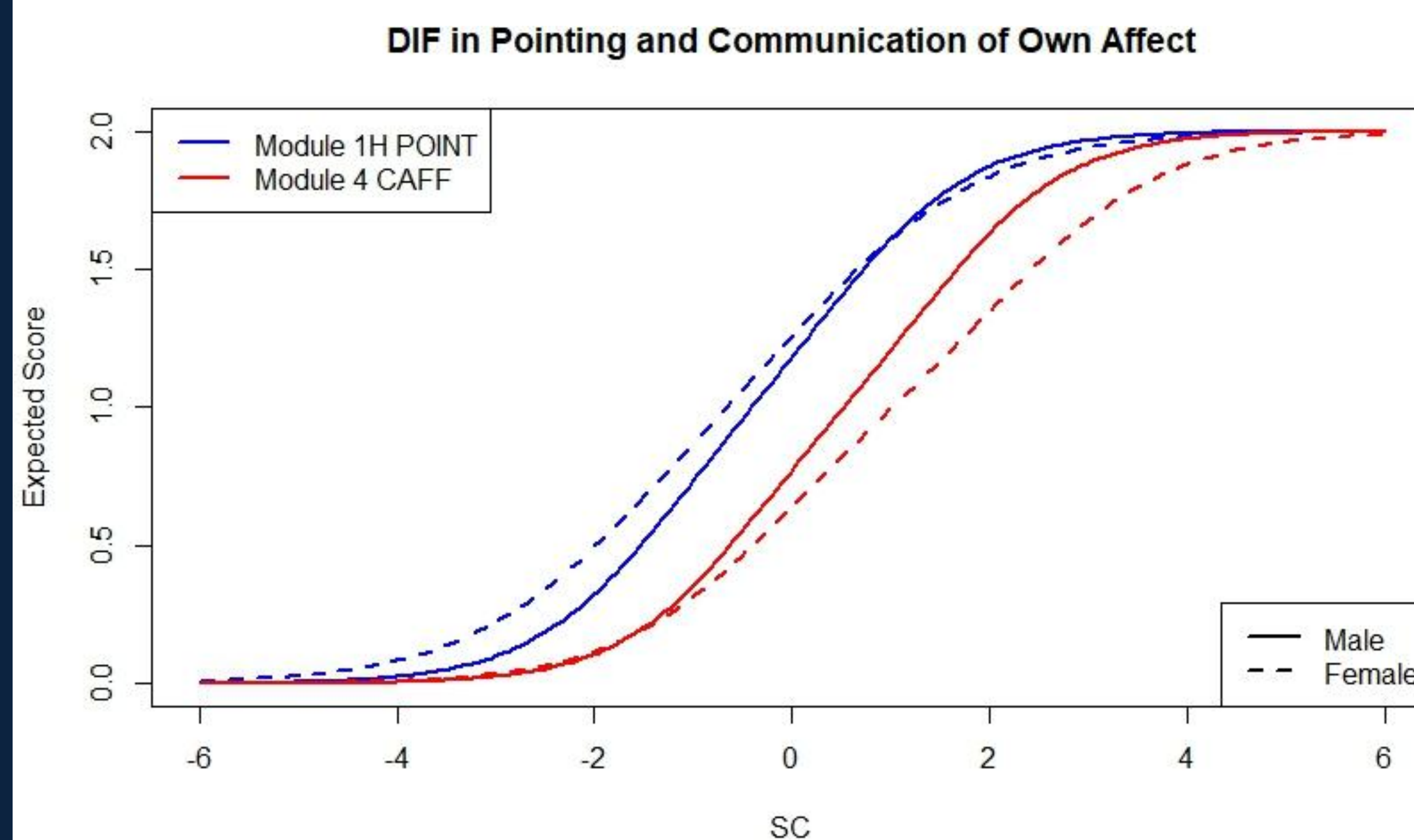
Module T	n = 1,873	1,318	555
Module 1, Lo lang	n = 2,272	1,793	479
Module 1, Hi lang	n = 2,832	2,167	665
Module 2, Lo age	n = 2,055	1,415	640
Module 2, Hi age	n = 1,344	1,067	277
Module 3	n = 6,705	5,294	1,411
Module 4	n = 2,483	1,752	731
	N = 19,564	14,806	4,758



Effect Size

Unusual Sensory Interests (Mod 1, High language):
ESSD = -0.522
(*large*)

Unusual Sensory Interests (Mod 4):
ESSD = 0.482
(*medium*)



Effect Size

Pointing (Mod 1, High language):
ESSD = 0.179
(*small*)

Communication of Own Affect (Mod 4):
ESSD = -0.307
(*small*)

Despite significant sex bias in 11 items across modules, test-level bias appeared modest in most modules.

