Item Response Analysis of Eye-tracking Data Reveals Effectiveness of Streamlined Assays: Results from the Autism Biomarkers Consortium for Clinical Trials (ABC-CT)

A. Naples, F. Shic, G. Helleman, G. Dawson, A. Levin, N. Kleinhans, R. Bernier, S. Jeste, S. Faja, K. Chawarska, C. Sugar, S. Webb, D. Senturk, J. Dziura, J. McPartland

Background

- Currently there are multiple putative eye-tracking biomarkers that quantify attention to social information e.g., faces.
- Traditional analyses aggregate the quantity of time spent looking to social information, e.g., fixation durations.
- Traditional analyses have revealed large effect sizes (d = .78) between autistic and neurotypical children such that autistic children look less at social information [1].
- However, specific stimuli within and across studies are highly variable (e.g., dynamic videos, static photographs). It is unknown how these differences impact the measurement of social attention.
- Here we sought to address the following questions:
  - Do different stimuli measure social attention equivalently?
  - Can different stimuli measure different 'levels' of social attention?
  - If stimuli are *not* equivalent, can we pick and use the 'best' stimuli?

Method

	ASD (n=280)	NT (n=119)
Age (in years)	M=8.55 (1.6)	M=8.51 (1.6)
Sex (male)	215 male	83 male
DAS-II GCA (Full Scale IQ)	96.6 (18.1)	115.1 (12.6)
Verbal Cluster Standard Score	96.4 (20.7)	116.3 (11.2)
Special Nonverbal Composite	97.5 (16.9)	112.2 (14.0)

- Data were collected from 280 autistic (ASD) and 119 neurotypical (NT) children between the ages of 6 and 11 across five sites as part of the Autism Biomarkers Consortium for Clinical Trials (ABC-CT).
- The eye-tracking battery included seventy stimuli presented across two days.
- Stimuli included (Figure 1)
  - Dynamic social videos with speech (DVS)
  - Dynamic social videos without speech (DVN)
  - Static social scenes (SS)
  - Biological and non-biological stimuli (BM)
- Item response theory (IRT) was used to address our objectives. IRT estimates how different stimuli, or items, may have more, or less, precision, depending on the level of the trait they are measuring. In contrast, traditional analyses assume that all items are equally effective measures.
- Each stimuli e.g., image or video, was defined as an item coded as a 0 or 1, with a 1 indicating that a participant had attended to the social aspects of the scene.
- Looking to social information more than 25% of the time was used as a threshold to determine whether someone had sufficiently attended to social information.
- IRT models these stimuli as if they were items on a test and determines
  - How well items measure a central construct (item discrimination)
  - What levels of the construct the items measure, e.g., are they easy or difficult? (item difficulty)

1. Shic, F., Naples, A. J., Barney, E. C., Chang, S. A., Li, B., McAllister, T., ... & McPartland, J. C. (2022). The autism biomarkers consortium for clinical trials: evaluation of a battery of candidate eye-tracking biomarkers for use in autism clinical trials. Molecular Autism, 13(1), 15.

Results: Item Characteristics

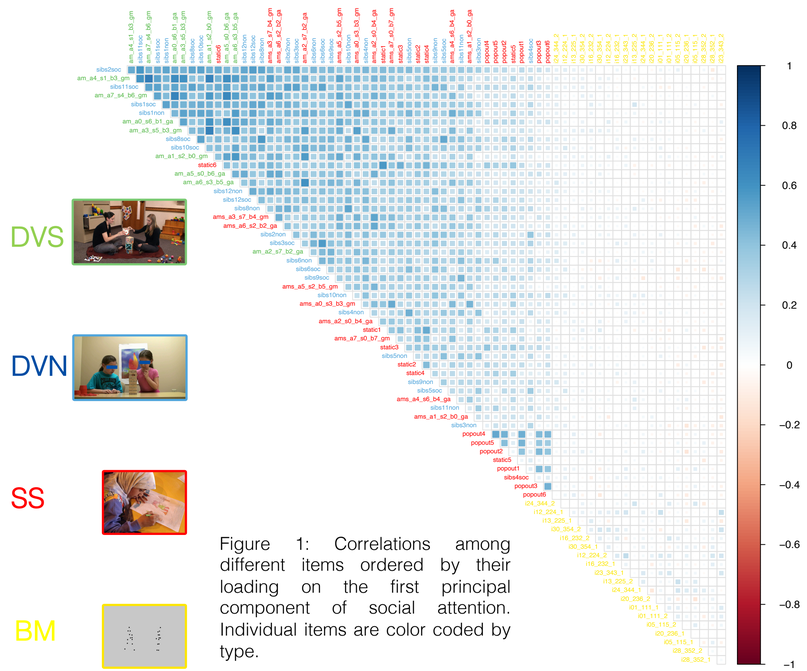

Figure 1: Correlations among different items ordered by their loading on the first principal component of social attention. Individual items are color coded by type.

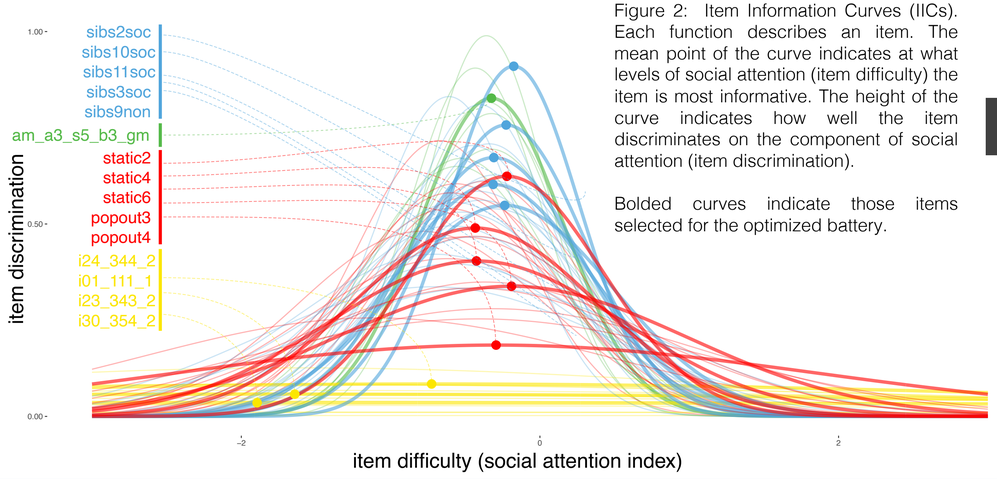

Figure 2: Item Information Curves (IICs). Each function describes an item. The mean point of the curve indicates at what levels of social attention (item difficulty) the item is most informative. The height of the curve indicates how well the item discriminates on the component of social attention (item discrimination).

Bolded curves indicate those items selected for the optimized battery.

Results: optimized battery

- A single component explained 16% of the variance in looking to social information, indicating sufficient unidimensionality.
- IRT analyses reveled that items captured a range of looking variability, some items discriminating among low levels of attention (SS) and others discriminating among higher levels (DSV; Figures 1 & 2)
- Dynamic social videos loaded most strongly on this component while biological motion was virtually uncorrelated with social attention (Figure 1)

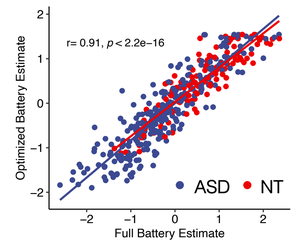

Figure 3: Correlations between all 70 items and an optimized battery of 15 items.

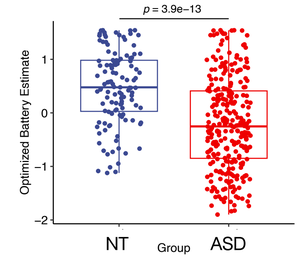

Figure 4: Group comparisons based on the optimized 15 item battery.

- A shortened battery was constructed using the 15 best performing items, which captured 82% of the variance of all 70 items in the sample of 399 children. (Figure 3)
- The IRT estimates discriminated groups with an effect size of d = 1.1, this is greater than the composite estimate (d = .78) and the largest single experiment effect size (d = 1.01) from traditional analyses in [1]. (Figure 4)
- The optimized (15 item) battery discriminated between groups with an effect size of d = 1.06.
- Additionally, the optimized battery correlated with social performance as measured by the ADOS Calibrated Severity Score (CSS) (r = -.2, p <.005).

Conclusions

- We validated that social attention as measured by this battery is unidimensional.
- IRT measures of social attention are either equivalent or outperform traditional analyses in both discriminating groups and in identifying biomarker behavior relationships.
- Streamlining the battery using only high-performing items shows that it is possible to dramatically reduce participant burden – from 70 to 15 items -- in ET batteries without loss of information.
- Ongoing analyses are exploring the effects of variable thresholds and item parameterizations.

McPartland Lab
mcp-lab.org
mcp.lab@yale.edu

Funding for the Autism Biomarkers Consortium for Clinical Trials (ABC-CT) was provided by NIH U19 MH108206 (McPartland).