Comparative Item Response Theory Analysis of the Autism Quotient

Z Winters-Derevjanik and Elia Harper

Dr. Rajesh Kana

CBRA Lab at the University of Alabama

Overview of ICON

Until recently, neuroimaging limitations have rendered researchers unable to sufficiently investigate the connections between the different layers of neurons that make up the neocortex (i.e. inter-layer connectivity).

Our study, ICON, uses a 7-Tesla MRI scanner which allows for previously unattainable details of granular organization in the neocortex. Alongside neuroimaging data, we administered a series of neuropsychological tests including the WASI-II, a multi-modal measure of cognitive ability, and Autism Quotient (AQ), a survey consisting of 50 items designed to assess traits associated with ASD. This poster presentation focuses on the statistical analysis of the AQ-10, a smaller subset of the AQ, using item response theory (IRT).

Abstract

Over the last several decades, the way clinicians and researchers conceptualize autism spectrum disorder (ASD) has shifted drastically. Moving beyond narrow diagnostic criteria has increased the prevalence rate of autism diagnosis.

Disrupted patterns of connectivity across the brain have been implicated in autism, including both local hyperconnectivity and global hypoconnectivity. Technological limitations have prevented deep study of the the connections between the different layers of neurons that make up the neocortex. This study, ICON (Inter-layer Cortical Connectivity in Autism), uses a 7-Tesla MRI scanner which allows for previously unattainable details of granular organization in the neocortex.

Alongside neuroimaging data, we administered a series of neuropsychological tests including the WASI-II, a multi-modal measure of cognitive ability, and Autism Quotient (AQ), a survey consisting of 50 items designed to assess traits associated with ASD. This poster presentation focuses on the statistical analysis of ten questions from the AQ which make up the AQ-10, using item response theory (IRT) methods to analyze responses from a sample of 14 participants. We compared the item parameters within our own sample with a much larger sample of the general population responding to the AQ in order to investigate how the measure may be functioning differently in our sample. Using a Rasch model and differential item functioning (DIF), we review our findings here.

Key Terminology

IRT (Item Response Theory) is a statistical framework for analyzing the relationships between item responses and individual abilities or traits

The Rasch Model is a specific type of IRT model that assumes that the probability of an individual getting an item correct depends only on the individual's ability and the item's difficulty.

Item parameters are characteristics of the items in a test that are estimated by IRT models. In the Rasch Model, there is one item parameter, which is the difficulty of the item. In other IRT models, there may be additional item parameters, such as discrimination or guessing parameters, which capture how well the item discriminates between individuals with different levels of the trait being measured, or how likely it is that an individual will guess the correct answer. Due to the size of our sample, we used the Rasch model.

Differential Item Functioning (DIF) is used to assess how items may function differently between groups of individuals, even when those individuals have the same level of ability on the construct being measured. DIF can arise due to differences in language or culture, or due to biases in the item content. Detecting and correcting for DIF is important to ensure that the test is fair and unbiased for all individuals, regardless of their background or characteristics.

Autism is a neurodevelopmental disability with wide-ranging impacts on everything from executive functions, the immune system, social behavior, physical development, and more. For the purpose of our research, we use the definition of autism spectrum disorder (ASD) provided by the DSM-V. The DSM-V outlines two core symptom domains of ASD:

- Persistent deficits in social communication and social interaction:
  - Deficits in social-emotional reciprocity, such as difficulty initiating or responding to social interactions or conversations.
  - Deficits in nonverbal communication, such as lack of eye contact or facial expressions.
  - Deficits in developing, maintaining, and understanding relationships, such as difficulty making friends or sharing interests.
- Restricted, repetitive patterns of behavior, interests, or activities:
  - Stereotyped or repetitive motor movements, use of objects, or speech.
  - Insistence on sameness, such as inflexible adherence to routines or rituals.
  - Highly restricted, fixated interests that are abnormal in intensity or focus.
  - Hyper- or hypo-reactivity to sensory input or unusual interest in sensory aspects of the environment.

Methods

Datasets

ICON: 14 adult participants tested in Fall of 2022

Male = 6

Autistic = 3

Autism Screening Adult Data Set (Tabtah, 2017): 704 adult participants tested in 2016/2017

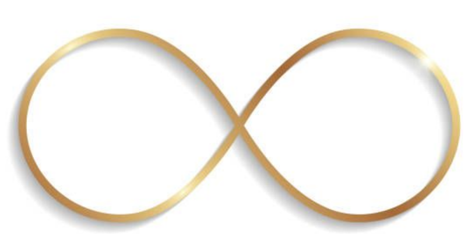
Male = 367

Autistic = 90

Measures

Autism Spectrum Quotient (AQ)/AQ-10

The autism spectrum quotient (AQ) is a 50 item questionnaire that measures autistic traits in adults without intellectual disability. These items are equally sub-divided into 5 phenotypic domains typically associated with autism: social skills, attentional switching, attention to detail, communication, and imagination. (Baron-Cohen et al., 2001)



The golden infinity symbol is used to celebrate autistic individuals, inspired by the symbol for gold on the periodic table, Au.

The AQ-10 is a shortened version of the AQ used to determine whether an individual should be referred for a comprehensive autism assessment. The AQ-10 is comprised of 10 questions from the original AQ, equally balanced across the 5 aforementioned subscales (Allison et al., 2012).

Statistical Methods

Data Preparation: The two datasets were obtained and cleaned for analysis. This involved scoring our dataset dichotomously to match the control dataset, and creating factors to be used in the DIF analysis for the gender and neurotype of participants.

IRT Model Fitting: Item Response Theory (IRT) models were fitted to both datasets. IRT models are commonly used in psychometric analysis to estimate the latent trait of interest (in this case, the level of autistic traits) based on responses to a set of items (in this case, the AQ questionnaire). We used the ltm package for model fitting after testing for unidimensionality.

Parameter Estimation: The parameters estimated from the IRT models were used to evaluate the performance of the AQ items. While the control dataset was large enough to conduct 2 parameter logistic analysis, the small size of the ICON dataset made Rasch analysis the better option.

Differential Item Functioning (DIF) Analysis: DIF analysis was performed to evaluate whether the AQ items were functioning differently between the two datasets (i.e., whether there were different levels of item difficulty or discrimination between the two groups). In this analysis, we used the "difR" package in R to perform DIF analysis using the Mantel-Haenszel method. DIF was evaluated by comparing the item difficulty and discrimination parameters between the two datasets. The two independent variables we examined for differential item functioning were gender (male and female) and neurotype (autistic and allistic).

Comparing the IRT Models: The IRT models fitted to both datasets were compared to evaluate the performance of the AQ items and assess whether the items functioned similarly in both datasets.

Results

Fig. A ICON Rasch Model Output

```
Coefficients:
              value std.err  z.vals
Dffclt.q1   -1.8157  0.7426 -2.4451
Dffclt.q2    2.0153  0.7771  2.5934
Dffclt.q3    0.6983  0.4744  1.4719
Dffclt.q4    0.4237  0.4435  0.9555
Dffclt.q5    0.6983  0.4744  1.4719
Dffclt.q6    0.1752  0.4282  0.4092
Dffclt.q7    1.0176  0.5267  1.9322
Dffclt.q8    0.4236  0.4435  0.9552
Dffclt.q9   -0.2854  0.4309 -0.6622
Dffclt.q10   0.6983  0.4744  1.4719
Dscrmn       2.1102  0.6177  3.4160
```

Fig. B: Control Rasch Model Output

```
Coefficients:
                     value std.err   z.vals
Dffclt.A1_Score    -1.0759  0.0998 -10.7822
Dffclt.A2_Score     0.1994  0.0862   2.3134
Dffclt.A3_Score     0.1799  0.0861   2.0901
Dffclt.A4_Score     0.0055  0.0854   0.0641
Dffclt.A5_Score    -0.0074  0.0854  -0.0866
Dffclt.A6_Score     1.0313  0.1001  10.3060
Dffclt.A7_Score     0.3634  0.0876   4.1484
Dffclt.A8_Score    -0.7049  0.0914  -7.7100
Dffclt.A9_Score     0.8199  0.0950   8.6290
Dffclt.A10_Score   -0.3498  0.0867  -4.0339
Dscrmn              1.0999  0.0498  22.0787
```

ICON Dataset

The coefficients section displays the difficulty of each item and the discrimination parameter. The difficulty parameters are listed as Dffclt.q1 through Dffclt.q10, with lower values indicating easier items. The discrimination parameter, listed as Dscrmn, indicates the degree to which the measure can discriminate between individuals with high and low abilities, with larger values indicating better discrimination.

In the ICON data set, **item 1** (Dffclt.q1) has the lowest difficulty, whereas **item 2** (Dffclt.q2) has the highest. **Item 1** is the easiest item, and **item 2** is the most difficult. **Item 9** (Dffclt.q9) has a negative difficulty value, which indicates that it may be too easy and not contribute meaningfully to the measurement of the latent trait.

The discrimination parameter is 2.1102, indicating that the items can distinguish somewhat well between individuals with high and low autistic traits.

Control Dataset

The coefficients section displays the difficulty of each item and the discrimination parameter. In the control data set, the items are labeled as Dffclt.A1_Score through Dffclt.A10_Score, and the discrimination parameter is listed as Dscrmn. The difficulty values range from -1.0759 (easiest) to 1.0313 (most difficult). **Items 6** and **9** have the highest difficulty values, indicating that they are the most difficult items. **Item 1** has the lowest difficulty value, indicating that it is the easiest item.

The discrimination parameter is 1.0999, which is lower than that of the ICON data set, indicating that the measure functions less effectively at discriminating between individuals with high and low autistic traits. However, the small sample size of the ICON dataset likely resulted in an inflated discrimination parameter, reducing the certainty of this comparison.

DIF across Gender

In the control dataset, the Mantel-Haenszel method detected DIF in one item (**item 7**) with a p-value of 0.0242, which is significant at the 0.05 level. The effect size for this item is 0.9783, which is classified as a moderate effect. This means that there is a moderate difference in the way males and females respond to this item, even after controlling for their overall test score, with females being more likely to endorse **item 7** at the same trait level as males.

In the ICON dataset, the Mantel-Haenszel method detected DIF in six items (**items 2, 3, 5, 6, 7, and 8)**, all with p-values of 0.0000, which are highly significant at the 0.05 level. The effect sizes for these items are not estimable (NaN), except for **item 1** and **item 9**, which have infinite effect sizes. This means that there may be significant and very large differences in the way males and females respond to these items, even after controlling for their overall test score. However, sample size and unequal sampling prevents meaningful interpretation of these values.

DIF across Neurotypes

In the control dataset, the results indicate that two items (**item 4** and **item 7**) have significant DIF between the two groups. The effect size (deltaMH) of these items is large (C), indicating a substantial difference in the item difficulty between the two groups. The other eight items do not show significant DIF.

In the ICON dataset, all ten items show significant DIF between NT and autistic individuals. However, the output does not provide any information about the direction or magnitude of the DIF effect. The effect size (deltaMH) for all items is reported as NaN, indicating that it was not possible to calculate the effect size due to the presence of infinite or missing values.

```
Effect size (ETS Delta scale):

Effect size code:
 'A': negligible effect
 'B': moderate effect
 'C': large effect

           alphaMH deltaMH
A1_Score    1.6145 -1.1257 B
A2_Score    0.9352  0.1574 A
A3_Score    0.9792  0.0495 A
A4_Score    0.9030  0.2399 A
A5_Score    0.9595  0.0971 A
A6_Score    0.8177  0.4730 A
A7_Score    1.6061 -1.1135 B
A8_Score    0.7032  0.8274 A
A9_Score    0.8596  0.3554 A
A10_Score   0.8969  0.2556 A

Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
 (for absolute values of 'deltaMH')
```

Fig. C Control DIF by Gender

```
Effect size code:
 'A': negligible effect
 'B': moderate effect
 'C': large effect

           alphaMH deltaMH
A1_Score    1.2502 -0.5248 A
A2_Score    0.9651  0.0834 A
A3_Score    1.1095 -0.2443 A
A4_Score    2.8538 -2.4644 C
A5_Score    0.7505  0.6745 A
A6_Score    0.8153  0.4798 A
A7_Score    0.4350  1.9562 C
A8_Score    0.7959  0.5365 A
A9_Score    1.4959 -0.9464 A
A10_Score   1.2196 -0.4666 A

Effect size codes: 0 'A' 1.0 'B' 1.5 'C'
 (for absolute values of 'deltaMH')
```

Fig. D Control DIF by Neurotype

Conclusions/Significance

We found that items 1 and 4, pertaining to the attention to detail sub-domain in the AQ had fairly low difficulty in ICON and control samples. A lower difficulty level suggests that these items need to be refined in order to discern the trait they intend to.

- For instance, item 1 stating, "I often notice small sounds when others do not", can be interpreted in two different ways with respect to autism. While there is a high overlap between individuals with sensory processing difficulties and individuals with autism, these difficulties function differentially. In some instances, autistic individuals are hyper-sensitive to particular sensory modalities such as sound, whereas in others they are hypo-sensitive to said sensory modalities.
- Similarly, item 4 stating, "I usually concentrate more on the whole picture rather than the small details." does not possess the requisite clarity to discern attention to detail in an autistic population. Autistic individuals have an exceptionally difficult time with metaphors and as a result typically utilize very literal, explicit language. Without prior exposure to said metaphor, this question can be really confusing and difficult to answer.

Notably, item 7 stating, " I find it easy to work out what someone is thinking or feeling just by looking at their face" yielded a very interesting array of responses across our analyses.

- Within the control group, we found that the conditions of autism and gender significantly increased the likelihood that an individual would endorse a response, indicative of the autistic trait the item intended to discern.
- Consequently, we found low difficulty in item 7, within the control group, but not in the ICON group.

Item 7 corresponds to the social skills subscale- aiming to measure an autistic individual's (diminished) ability to perspective take. However, this question does not take into account that expression of emotion and expectation to perspective-take is heavily influenced by gender norms (ie sociocultural attitudes and expectations regarding the differential presentation of behaviors, preferences, and skill sets between genders). For instance, gender norms may place a greater expectation on women to be more responsive to emotional distress and jointly, responsible for the mitigation or mediation of a conflict.

Discussion

We conducted these analyses in service of the following directives:

1.) We felt that IRT analyses would afford us a better understanding of our sample, as we begin interpreting neuroimaging results in the next phase of our study.

2.) We desired to critically examine the integrity and applicability of the AQ-10, as it functions in a clinical and academic setting

We discerned that items within the attention to detail sub-domain may be too ambiguous to significantly discern the variable presentation of this disorder. Unless reworded, this portion of the AQ-10 may be most relevant as a clinical tool.

Subsequently, our findings suggest that gender and neurotype both inform the likelihood of an individual to endorse items on the Autism Quotient. As such, it is our intention to continue data collection after we complete the analysis of our pilot data, to hopefully obtain a more balanced and applicable dataset.

Bibliography

1. Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. Journal of autism and developmental disorders. 2001;31(1):5–17.
2. Allison C, Auyeung B, Baron-Cohen S. Toward brief "red flags" for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1000 cases and 3000 controls. Journal of the American Academy of Child & Adolescent Psychiatry. 2012;51(2):202–212.
3. Thabtah, F. (2017, May). Autism spectrum disorder screening: machine learning adaptation and DSM-5 fulfillment. In Proceedings of the 1st International Conference on Medical and Health Informatics 2017 (pp. 1-6)