Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Chloe Boynton April 10, 2023

Item Analysis of an Early Social Responsiveness Assessment for Assessing Autism Risk

by

Chloe Boynton

Dr. Elaine Walker
Adviser

Psychology

Dr. Elaine Walker

Adviser

Dr. Opal Ousley

Co-Adviser

Dr. Harold Gouzoules

Committee Member

Dr. Susan Tamasi

Committee Member

2023

Item Analysis of an Early Social Responsiveness Scale for Assessing Autism Risk

By

Chloe Boynton

Dr. Elaine Walker

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Psychology

2023

Abstract

Item Analysis of an Early Social Responsiveness Scale for Assessing Autism Risk
By Chloe Boynton

Early diagnosis of autism spectrum disorder (ASD) is vital for effective intervention and improves long term social and behavioral development. Previous literature has shown that the Early Social Responsiveness (ESR) assessment is effective at detecting ASD risk as early as 13 months (Factor et. al, 2021). However, an item analysis that examines individual scores has not been conducted. In this study, I have analyzed an existing data set containing the individual item responses from the ESR assessment of 120 children (n=61 males, n=59 females; age range= 15-24 months; Factor et. al, 2021). By using item analysis, we determined which ESR items, or item sets, were best at differentiating ASD risk from non-ASD risk. Ease of social engagement measures performed most effectively, scoring the highest in the item discrimination, item discrimination index, and item reliability index analyses. The hat and tickle activities performed the least effectively, with only the ease of social engagement ratings in their respective sections scoring among the highest value items across analyses. The item analysis results may allow the assessment to be shortened which may increase the likelihood of clinical adoption, allowing for earlier intervention and diagnosis.

Item Analysis of an Early Social Responsiveness Scale for Assessing Autism Risk

By

Chloe Boynton

Dr. Elaine Walker

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Psychology

2023

Acknowledgements

Special thanks to Dr. Opal Ousley for providing exceptional support and expertise. I also thank the members of the committee, Dr. Elaine Walker, Dr. Harold Gouzoules, and Dr. Susan Tamasi, who provided their assistance in the editing of my thesis.

Table of Contents

Introduction---------- 1

Methods----------6

Table 1 Demographic Information for the Community Subset Sample----------7

Results---------- 11

Table 2 Descriptive Statistics for the Items in the ESR Assessment----------11

Table 3 Item Analysis Results for Items in the ESR Assessment----------13

Table 4 Item Statistics of the Item Reliability Analysis for the Items in the ESR Assessment-15

Discussion----------16

References----------20

Appendix A. The Early Social Responsiveness Scale----------26

Appendix B. R Codes----------29

Introduction

Autism spectrum disorder (ASD) is characterized by impaired social interaction and communication, as well as repetitive behaviors. Children with ASD face these challenges, which causes a gap between their social and behavioral development in comparison to typically developing children (Elder et al., 2017). Diagnosis becomes most reliable at age two, but parents might observe concerning behaviors that do not yet meet ASD criteria. This makes both clinical intervention and addressing parent's concerns difficult in early years. The Early Social Responsiveness (ESR) assessment has been shown to identify early social delays and differentiate children at high risk for ASD from their peers (Factor et al., 2021). The assessment is interactive, lasts about 3 minutes, and is designed to measure specific social responsive behaviors (e.g., eye contact, smiling, pointing, turn-taking, and ease of social engagement) that are observed in real time across five different activities. Additionally, the assessment uses scripted instructions to provide an easier method for clinicians and pediatricians to adopt in clinic. The total score is then calculated and used to determine ASD risk (Factor et al, 2021). However, we do not yet know the effectiveness of individual items within the assessment to determine ASD risk.

Early diagnosis of ASD, followed by effective intervention, is known to be associated with better long-term development and alerts clinicians to search for any co-existing medical conditions common in individuals with ASD (Posar and Visconti, 2020). Research has shown children who enter intervention programs at earlier ages tend to receive more social and behavioral gains in comparison to their older peers (Corsello, 2005). The American Academy of Pediatrics suggests children be screened for ASD as early as 18 and 24 months but the stability of diagnoses from assessments at this age are not well supported (Guthrie et al., 2013). In their review, Posar and Visconti (2020) acknowledged the variety of assessments available to detect autism early but

emphasized the lack of evidence for their effectiveness. The researchers state the early signs to look for when screening for ASD, which include social communication skills, patterns of behavior, and sensory behaviors or reactions (Posar and Visconti, 2020). These indicators have a range of complexity from simple behaviors to complex interactions. Thus, having an assessment like ESR, which can measure all of these, might be a useful tool for clinicians but barriers to clinical adoption remain.

Not all pediatricians regularly assess for ASD factors in-clinic and wellness check-ups are short, making the adoption of an assessment tool like ESR difficult (Factor et al., 2021). In 2013, the National Institute of Mental Health funded trials to develop innovations designed to improve early ASD screening and diagnosis (Fingert et al., 2018). However, a review of the implementation of these assessments determined there exists multiple barriers preventing their adoption, including inefficient systems of care, provider attitudes, and difficulty changing clinic processes, among others (Fingert et al., 2018). In their literature review, Goetz et al. (2013) investigated numerous studies that aimed to shorten composite measurement scales and found that scales with the highest validity were most useful for clinical practice (Goetz et al., 2013). Another study surveyed caregivers of children with ASD and found that continuous pediatric care and specialty referrals resulted in earlier diagnosis (Mandell et al., 2005). Thus, pediatric care and clinical adoption could have major positive impacts in the early identification and treatment of children with autism. While the barriers mentioned previously are great, studies have shown the implementation of standard assessments (e.g., completion of parent questionnaires) to detect ASD risk during wellness baby checkups is possible and effective (Pierce et al., 2011).

The ESR utilizes some of the earliest indicators of ASD to determine individual risk. It does this by measuring social awareness, responsiveness, and engagement and captures skills that

are present by the first year of life. These include an increase in the presence of smiling and vocalizations, paying more attention to facial expression, and initiating joint attention (Factor et al., 2021). Very young children with ASD tend to exhibit fewer ESR behaviors and have difficulty in timing their social interactions. Especially in activities involving joint attention, infants with ASD struggle significantly more, resulting in poorer performance on the ESR assessment (Factor et al., 2021). There exists well established knowledge of the relationship between ASD and joint attention. The level of joint attention skills is among the earliest signs of autism, and due to its critical role in development, it has become the target for many intervention programs. Behaviors that involve joint attention, such as gaze or attention switching, can be used to detect developmental trajectory and severity of ASD (Charman, 2003). Additionally, joint attention measures have been shown to emphasize early social developmental disturbances, because they are directly related to the cognitive, affective, and neurological processes that play a role in ASD (Mundy and Crowson, 1997).

There are other factors that can help differentiate individual autism risk. Deficits in social and communication responsiveness have been shown to be strongly associated with ASD. Children with autism struggle to imitate a play partner's simple movements, initiate actions with symbolic meaning, and acting upon objects during reciprocal play (Williams, Whiten, & Singh, 2004). Thus, in activities involving the use of reciprocal play with an object (e.g., rolling a ball-back-and-forth) could be particularly difficult for children with ASD. Additionally, research has shown that children with autism show different developmental pathways for face and object processing and that children with autism spend more time looking at objects rather than faces, which could impact social attention during reciprocal play (Pallett, Cohen, & Dobkins, 2014). These findings lead us to predict that object-based activities within the ESR that require coordinating attention between

an object and a person, as well as sustaining engagement and expressing positive emotion, could be more complex for high ASD risk children.

Another major part of the ESR is social engagement ratings. Social engagement is one of the key diagnostic features of ASD (Anagnostou et al., 2014); however, a prior review article found that within ASD, social engagement skills can vary significantly between individuals. Social engagement ratings can depend on age, language, and physical activity level (Pan, 2009). Thus, it could be difficult to rely solely on social engagement as a differentiator, unless these conditions are controlled. The ESR requires no verbal response from children and is administered at a table while the child is sitting on the parent's lap; These procedures provide some element of control for variability in physical activity and language ability.

The current study is an item analysis of the ESR assessment using data previously published (Factor et al., 2021; "parent study"). Item analysis is a statistical tool that can improve the test and item quality of assessments, by investigating individual items, item sets, and the relationship between them (McCowan and McCowan, 1999). Specifically, we calculated the *item difficulty*, *item discrimination, item discrimination index, item reliability index*, and *item validity*. *Item difficulty* is a common statistic used in item analyses, as it measures the proportion of participants who correctly answered each item. Due to our coding of 1 being the absence of expected behavior and 0 indicating the presence, higher item difficulty values indicated more difficult items. To account for this coding scheme, item difficulty measured the proportion of individuals who did not exhibit the behavior on each item. It is represented with a p-value but must not be confused with p-values used in hypothesis testing. Thus, an item with a high p-value was classified as a more difficult question (Desjardins and Bulut, 2020). In terms of ESR difficulty, an easier question was one that had a higher proportion of children who demonstrated a behavior,

while a more ‘difficult’ question had a higher proportion of children who did not exhibit the behavior.

Item discrimination is also widely used and measures an item’s power to distinguish between participants of low and high ability. There are many methods to accomplish this task with the most common being a point-biserial correlation between the total score and the score on an individual item. Another way to calculate item discrimination (i.e., *item discrimination index*) is to split participants into two groups based on their total scores and correlate the group membership variable to individual items (Desjardins and Bulut, 2020). Item difficulty and discrimination are known to be correlated as moderately difficult questions are the best at discriminating between top and bottom scorers (Sim and Rasiah, 2006). In our study, we aimed to identify the items with the highest discriminatory ability to differentiate between high and low ASD risk.

Next, the *item reliability index (IRI)* can provide information on how reliable individual items are. It can be calculated using Cronbach’s alpha given by the formula a=N*c/v+(N-1)*c where N is the number of items, c is the mean covariance between items, and V is the mean item variance. *Item validity* is calculated using the correlation of individual items to an external criterion instead of the total test score. Thus, our item validity analysis used the pass/fail scores from the Modified Checklist for Autism in Toddlers, with follow-up (M-CHAT-23), and correlated the M-CHAT-23 score with individual ESR items (Factor et al., 2021). The M-CHAT-23 measures social relatedness, joint attention, pointing, and pretend play (Robins et al., 2001). Research has shown many of the questions within the M-CHAT-23 to be successful at discriminating children aged 18 to 24 months who have autism from their typically developing peers (Robins et al., 2001; Wong et al., 2004).

The parent study concluded the total ESR score was effective at distinguishing high ASD risk children from their low ASD risk peers. In this study, we ran an item analysis to test which ESR item level scores best detected ASD risk and if a subset of items could be eliminated. We identified the following individual hypotheses. *Hypothesis 1:* Joint attention items will have the highest difficulty values (i.e., the most difficult items) compared to other items; *Hypothesis 2.1*: Every item on the ESR will discriminate between high and low ESR scorers based on the item discrimination analysis; *Hypothesis 2.2:* The strongest discrimination values will involve joint attention items; *Hypothesis 3.1:* Each item on the ESR will discriminate between two scoring groups (i.e., above and below the median ESR total score); *Hypothesis 3.2*: Joint attention items to have a higher discrimination index compared to other items; *Hypothesis 4:* All items will have a high item reliability index (IRI), with the possibility of object based play tasks having higher IRI; *Hypothesis 5:* All items will have high item validity (i.e., correlating with an external ASD risk criterion), with performance on object based play tasks having the highest validity.

Methods

Participants

Statistical analysis for this study was conducted on an existing data set. Participants included a community subset of 120 children (n=61 males [50.8%], n=59 females [49.2]; age range=15-24 months). The sample included the following self-reported demographics: Caucasian (n=63; 52.5%), African American (n=22; 18.3%), Asian (n=14; 11.66%), multiracial (n=12; 10%), Hispanic (n=1; 0.83%), other (n=2; 1.66%), and unknown (n=6; 5%). Maternal education level for the sample included: graduate degree (n=30; 25%), some graduate school (n=10; 8.3%), college degree (n=40; 33.3%), some college (n=23; 19.1%), high school diploma (n=7; 5.8%), some high

school (n=1; 0.83%), other (n=2; 1.7%), and unknown (n=7; 5.8%). Children completed assessments, including the ESR, at the Child Study Lab at Georgia Institute of Technology. Pass/fail results on the M-CHAT-23, with follow-up, revealed 11.7% (n=14) of the sample were at risk for ASD (Table 1). While 40 (33.3%) children attended a follow-up appointment about 4 months after the initial visit (mean 4.6 months; SD 1.8; range 2–8 months), this study only looked at data from the initial visit. Recruitment included advertisements, flyers, an online study portal, and a community of daycares. Data for this subset was collected between March 2011-June 2015.

The data were collected and distributed under the approval of a Georgia Tech Institutional Review Board. Parents provided informed consent prior to entry into the study and clinical referrals were provided as needed.

Table 1

Demographic Information for the Community Subset Sample

Measure	Sample
n	120
Age months, mean (SD)	19.6 (3.0)
Male: female ratio	1.03:1 (61:59)
M-CHAT-23	11.7% ASD-risk (n=14)
Maternal education level	Graduate degree (n=30; 25%), some graduate school (n=10; 8.3%), college degree (n=40; 33.3%), some college (n=23; 19.1%), high school diploma (n=7; 5.8%), some high school (n=1; 0.83%), other (n=2;1.7%), and unknown (n=7; 5.8%)
Race	Caucasian (n=63; 52.5%), African American (n=22; 18.3%), Asian (n=14; 11.66%), multiracial (n=12; 10%), Hispanic (n=1; 0.83%), other (n=2; 1.66%), and unknown (n=6; 5%)

Note: Demographic information is from the community subset in the original data from Factor et al, 2021.

Measures

In the parent study, participants were given multiple assessments that lasted 45 minutes in total. The present study focused on the ESR component, while considering pass/fail results on the M-CHAT-23.

The ESR assessment lasted about 3 minutes and both the child and parent were present in the room. The child sat on the parent's lap at a table while test items were being administered. The test was administered by trained clinicians or research assistants. The same procedures were followed, as scripted language was provided for item administration. The ESR assessment contained five structured play activities that use standardized verbal prompts and periodic pauses during the activities (e.g., rolling a ball, looking at a book, and tickling).

There were 22 individual item scores derived from the ESR including 17 behavioral codes and 5 summary codes as described below. The behavioral codes indicated the presence/absence of pointing, eye contact, smiling, and turn taking. Absence of any behavior was recorded as 1 point and the presence was coded as 0 points. The summary questions measured *ease of social engagement* on a 0 to 2 scale. The engagement of the child was recorded as *very easy to engage, somewhat easy to engage, and hard to engage*. The scores were 0, 1, and 2 respectively. However, the score of 1 was transformed into a score of 2 to achieve a dichotomous score. The total score was the summation of the 17 behavior ratings, as well as the ease-of-social-engagement ratings (total score range=0-27). Thus, higher scores indicated poorer performance while lower scores indicated better performance. The test took 2.5 to 3 minutes and was scored via video tape by two

separate raters. Training for raters included practice administrations after live observations and achieving >80% co-coding reliability. All test administrators were female.

Planned Data Analysis

Item analysis provides important data that allows for the development and refinement of assessments or measures (Desjardins and Bulut, 2020). This item analysis was completed using R and the codes can be found in Appendix B. The following tests were run: *item difficulty, item discrimination, item discrimination index, item reliability index, and the item validity*.

In this section when we refer to a "correct" answer, we are referring to the general interpretation of item analysis results. Regarding this study and ESR, a correct answer would be children with lower scores (i.e., children who exhibit the behavior or are easier to engage).

Item difficulty measures the proportion of participants who correctly answer each item; however, given the direction of coding for each item with 1 being absence of behavior and 0 meaning presence, higher values represent more difficult items. The item difficulty was calculated for each question and scores were rounded to three decimals. The output showed the items with the highest and lowest difficulty.

Item discrimination refers to the ability of an item to distinguish between participants with high versus low ability. A point-biserial correlation between the individual item response and the total score was used. Large, positive values indicated a strong correlation between answering a specific item correctly and doing well on the test (i.e., having a low total ESR score). A value of 0 indicated no correlation. Large, negative values indicated a correct answer to a specific item is correlated with poorer performance on the test (i.e., a larger total ESR score). Thus, if a value is close to zero or negative, the analysis suggests it is not functional in the assessment. The total test score was computed and then the items were correlated to the total score.

The item discrimination index refers to item discrimination with grouping of participants based on their total scores. First, we grouped participants who scored above the median total score and below. Then we calculated the proportion of individuals in each group who answered the item correctly. Next, we calculated the odds ratio. Finally, we calculated the *item discrimination index* (IDI) by multiplying the odds ratio by the difference in the proportion of high performers and low performers who answered the question correctly. IDI is usually used for True/False or Right/Wrong (dichotomous) answers on an exam, thus we recoded our individual item scores to fit within this framework. For ease of social engagement items, all scores of 2 were converted to a score of 1 for this analysis only. Scores of 0 were left as such.

The item reliability index (IRI) can be used to test if a question is reliable. High positive values indicate high reliability. To generate the IRI, the "psych" package in R was used to calculate the "r.drop". R.drop represents the item whole correlation for the particular item against the scale without this item (Revelle, 2022).

Associated values were also generated including the "raw.r" value which represents the correlation of each item with the total score, but which is not corrected for item overlap. Additionally, an "std.r" value was generated that represents the correlation of each item with the total score. Finally, a "r.cor" value was generated. R.cor is the item whole correlation corrected for item overlap and scale reliability (Revelle, 2022).

The item validity index (IVI) is useful when an external criterion is used instead of the total score. In this study, the external criterion was ASD risk measured by M-CHAT-23. A failing M-CHAT-23 score indicates children at high risk for autism. A value of 1 was used to designate a failed M-CHAT-23 (high risk for ASD), while a value of 0 was used for a passing score (low risk

for ASD). Larger IVI values indicate higher validity. The psych package was used for this analysis as well. The validity of the items to the total score (r.cor values) and to the M-CHAT-23 pass/fail scores were examined. Then to convert the correlation to an item validity index we used the Fischer's z transformation. This converts the correlation coefficient to a normal distribution, which can make it easier to compare.

Results

Items 3, 8, 15, 18, and 22 in the ESR were measures of ease of social engagement. Items 4 and 9 measured joint attention skills. Items 5, 11, and 12 involved the child's use of an object during play. Finally, all other items were measuring the presence of absence of social responsive behaviors during non-object play interactions (Appendix A). The mean, standard deviation, and variance for each item on the ESR assessment is shown in Table 2. Ease of social engagement ratings had values of 0 or 2, while other items had values of 0 or 1.

Table 2

Descriptive Statistics for the Items in the ESR Assessment

	Mean	*SD*	*Variance*
1.Greeting Eye Contact	0.167	0.374	0.14
2.Greeting: Smile	0.825	0.382	0.146
3.Greeting: Engagement	0.45	0.672	0.451
4.Ball: Joint Attention	0.225	0.419	0.176
5.Ball: Roll	0.183	0.389	0.151

6.Ball: Smile	0.1	0.301	0.091
7.Ball: Pause	0.258	0.44	0.193
8.Ball: Engagement	0.258	0.615	0.378
9.Book: Joint Attention	0.258	0.44	0.193
10.Book: Smile1	0.567	0.498	0.218
11.Book: Turn Page	0.692	0.464	0.176
12.Book: Point	0.575	0.77	0.246
13.Book: Smile2	0.567	0.091	0.248
14.Book: Pause	0.692	0.49	0.215
15.Book: Engagement	0.608	0.467	0.593
16.Hat: Eye Contact	0.008	0.25	0.008
17.Hat: Smile	0.608	0.382	0.24
18.Hat: Engagement	0.158	0.464	0.218
19.Tickle: Eye Contact	0.067	0.25	0.063
20.Tickle: Smile	0.175	0.382	0.146
21.Tickle: Pause	0.308	0.464	0.215
22.Tickle: Engagement	0.35	0.657	0.431

Results are reported by item number; the item numbers can be found in Appendix A as well as in the tables below. The *item difficulty* analysis showed that item 2 had the highest p-value, with about 80% of children not smiling when greeted (Table 3). Items 10 and 14 also showed high p-values, with about 68% and 69% of children either not smiling (Item 10) or unsuccessfully making eye contact after a pause (Item 14). Items 16 and 19 had the lowest p-values with 99% of children making eye contact during the book as a hat and tickling activities, suggesting they were

the easiest of the assessment (Table 3). Of the highest two p-value items (i.e., the most difficult), none of them comprised joint attention measures, ease of social engagement measures, or items involving the child's use of an object. These findings do not support our hypothesis that joint attention measures would be the most difficult items. However, on joint attention measures, item 4 and 9, about 23% and 26% of participants did not demonstrate joint attention skills (Table 3).

Table 3

Item Analysis Results for Items in the ESR Assessment

Task	Item Difficulty	Item Discrimination	Item Discrimination Index	Item Reliability Index (r.drop)	Item Validity
1.Greeting Eye Contact	0.167	0.509	0.429	0.446	0.261
2. Greeting: Smile	0.825	0.339	0.292	0.265	0.031
3.Greeting: Engagement	0.450	0.568	0.783	0.459	0.184
4. Ball: Joint Attention	0.225	0.495	0.467	0.424	0.381
5. Ball: Roll	0.183	0.419	0.239	0.348	0.165
6. Ball: Smile	0.100	0.425	0.258	0.371	0.052
7. Ball: Eye Contact Aft Pause	0.258	0.436	0.332	0.356	0.331
8. Ball: Engagement	0.258	0.702	0.996	0.626	0.372
9.Book: Joint Attention	0.258	0.416	0.328	0.334	0.142
10. Book: Smile1	0.683	0.242	0.201	0.146	0.137
11. Book: Turn Page	0.225	0.432	0.361	0.355	0.179
12. Book: Point	0.575	0.291	0.225	0.191	0.210
13. Book: Smile2	0.567	0.447	0.555	0.357	0.162
14. Book: Eye Contact Aft Pause	0.692	0.509	0.896	0.431	0.074
15. Book: Engagement	0.608	0.633	0.998	0.519	0.153
16. Hat: Eye Contact	0.008	0.122	0.024	0.103	0.258
17. Hat: Smile	0.608	0.393	0.437	0.300	0.187
18. Hat: Engagement	0.158	0.498	0.352	0.418	0.214
19. Tickle: Eye Contact	0.067	0.320	0.067	0.271	0.111
20. Tickle: Smile	0.175	0.453	0.479	0.386	-0.031
21. Tickle: Eye Contact Aft Pause	0.308	0.534	0.522	0.458	0.210
22.Tickle: Engagement	0.350	0.605	0.781	0.506	0.084

The item discrimination was calculated for each item and correlated with the total score. For items 8, 15, and 22 a large positive correlation was found (Table 3). This suggests these items were best at discriminating high ESR scorers from low. Of these items, all 3 were measures of *ease of social engagement* for the ball, book, and tickling activities. This does not support our hypothesis that every item on the ESR would discriminate, with joint attention items having the strongest discrimination ability. Of the three strongest correlations none were joint attention measures. Additionally, it appears that some items, like item 16, did not discriminate very well.

Next an item discrimination index was conducted to see how the top and bottom scorers performed on each individual item (Table 3). Items 3, 8, 14, and 15 all had very high IDI scores, suggesting they are best able to distinguish between high and low scorers on the ESR assessment. Of these items 3, 8, and 15 are ease of social engagement measures, which could mean that these are the best type of items to discriminate ASD risk. The lowest scores were items 16 and 19, both of which measure eye contact. This could suggest removing these items would be able to shorten the assessment without affecting its ability, as these items also scored low on validity and reliability (Table 3 and Table 4). These findings do not support our hypothesis that joint attention measures are the best discriminators.

The item reliability index showed high values for items 8, 15, and 22 with no zero values for any item (Table 3). This shows not only are the items on the ESR a reliable measure, but also that social engagement ratings scored the highest for reliability. Our hypothesis that object-based play tasks would have the highest reliability was not supported, as items 5, 11, and 12 involved object-based play.

In Table 4, the complete item statistics for the item reliability analysis is shown. The "raw.r" value represents the correlation of each item with the total score, but is not corrected for item overlap (range=0.121-0.702). The "std.r" value represents the correlation of each item with the total score (range=0.196-0.671). The "r.cor" value is the item whole correlation corrected for item overlap and scale reliability (range=0.141-0.686). The "r.drop" is the item whole correlation for the particular item against the scale without this item (range=0.103-0.626) The mean and standard deviations refer to each individual item and their scores (Revelle, 2022).

Table 4

Item Statistics of the Item Reliability Analysis for the Items in the ESR Assessment

	n	raw.r	std.r	r.cor	r.drop	Mean	SD
1.Greeting Eye Contact	120	0.509	0.517	0.496	0.446	0.167	0.374
2.Greeting: Smile	120	0.339	0.354	0.301	0.265	0.825	0.382
3.Greeting: Engagement	120	0.568	0.539	0.529	0.459	0.450	0.672
4.Ball: Joint Attention	120	0.495	0.527	0.498	0.424	0.225	0.419
5.Ball: Roll	120	0.419	0.404	0.380	0.348	0.183	0.389
6.Ball: Smile	120	0.425	0.441	0.403	0.371	0.100	0.301
7.Ball: Pause	120	0.436	0.427	0.384	0.356	0.258	0.440
8.Ball: Engagement	120	0.702	0.671	0.686	0.626	0.258	0.615
9.Book: Joint Attention	120	0.416	0.446	0.414	0.334	0.258	0.440
10.Book: Smile1	120	0.242	0.241	0.180	0.146	0.683	0.467
11.Book: Turn Page	120	0.432	0.441	0.403	0.355	0.225	0.419
12.Book: Point	120	0.291	0.268	0.208	0.191	0.575	0.496
13.Book: Smile2	120	0.447	0.451	0.414	0.357	0.567	0.498
14.Book: Pause	120	0.509	0.492	0.462	0.431	0.692	0.464
15.Book: Engagement	120	0.633	0.591	0.585	0.519	0.608	0.770
16.Hat: Eye Contact	120	0.122	0.196	0.141	0.103	0.008	0.091
17.Hat: Smile	120	0.393	0.389	0.333	0.300	0.608	0.490
18.Hat: Engagement	120	0.498	0.506	0.481	0.418	0.158	0.467
19.Tickle: Eye Contact	120	0.320	0.356	0.318	0.271	0.067	0.250
20.Tickle: Smile	120	0.453	0.453	0.427	0.386	0.175	0.382
21.Tickle: Pause	120	0.534	0.559	0.536	0.458	0.308	0.464
22.Tickle: Engagement	120	0.605	0.589	0.588	0.506	0.350	0.657

Note: "raw.r"= correlation of item with total score; "std.r"=correlation of item with total score if items standardized; "r.cor"=item whole correlation corrected for item overlap and scale reliability; "r.drop"=item whole correlation item against scale without item (Revelle, 2022).

The item validity analysis correlated ESR items with M-CHAT-23 pass/fail scores. Failing scores on the M-CHAT-23 (high autism risk) were coded as a value of 1, while passing scores (low autism risk) were coded as a value of 0. The correlations overall were low, with the highest value of 0.382 for item 4 (range=-0.031-0.382; Table 3). A negative correlation was found for item 20 (r=-0.031), which suggests the item does not predict autism risk based on the M-CHAT-23. However, because this value is so close to zero, it might suggest no correlation. In comparison, the correlation for the individual items to the total test score shows high correlation on some items (Table 4, r.cor values). In this case items 3, 8, 15, 21, and 22 had the higher correlations ranging from about 0.5 to 0.7. Of these items, four of them are ease of social engagement ratings (Items 3, 8, 15, and 22). The item validity results did not support our hypothesis that tasks involving object-based play would have the highest validity.

Discussion

Early diagnosis and intervention are crucial to improve the developmental trajectory of children who have ASD. The ESR assessment was created to detect early delays in social responsive behaviors, but there are many clinical barriers that make its adoption difficult. This study aimed to analyze the individual items within the assessment to see if it could be shortened and improve its chances of clinical use. Prior research showed that joint attention skills are a crucial

milestone in development, that children with autism struggle with tasks involving joint attention, and that the skill level of joint attention can predict the developmental trajectory of children with ASD (Factor et al., 2021; Mundy and Crowson 1997; Charman, 2003). As such, we predicted tasks involving joint attention skills would be the best at differentiating between high and low ASD risk children. Additionally, more complex tasks, such as those that involve the child's use of an object, have been shown to pose a challenge for children with ASD (Williams, Whiten, & Singh, 2004). This could be explained by the fact that children with autism spend more time looking at objects rather than faces, which could make reciprocal play difficult (Pallett, Cohen, & Dobkins, 2014). For these reasons we expected tasks involving objects to have a higher reliability and validity score than other tasks.

Our first hypothesis was that joint attention items would be the most difficult measures compared to other items. In the item difficulty analysis, eye contact in the hat and tickle activities were the easiest items. Conversely, smiling in the greeting activity was the hardest item (i.e., highest item difficulty value). The two joint attention measures were the 9th and 10th easiest out of 22 items on the assessment. Our second hypothesis predicted that every item on the ESR would be able to discriminate with joint attention items having the strongest discriminatory ability. The item discrimination analysis showed that ease of social engagement ratings in the ball, book, and tickling activities had the highest values, while eye contact in the hat activity scored the lowest. The joint attention items scored as the 8th and 15th best discriminators out of the 22 items on the assessment. Our third hypothesis was similar to the second, as it predicted every item on the ESR would be able to discriminate between the top and bottom groups with joint attention items having the highest IDI. The items with the highest IDI consisted of ease of social engagement ratings in the greeting, ball, and book activities. The lowest IDI items measured eye contact in the hat and

tickle activities. The joint attention items scored as the 8th and 14th best discriminators between the groups. Our fourth hypothesis predicted that all items would have a high reliability index and that object based play tasks would associate with the highest reliability values. Similar to previous results, the items with the highest reliability were ease of social engagement in the ball, book, and tickle activities. The items involving object-based play ranked 13th, 14th, and 19th in reliability.

Finally, our last hypothesis was that tasks involving objects would have the highest validity. The item validity analysis showed that joint attention in the ball activity had the highest value. The scores for the item validity analysis, when using the M-CHAT-23 with follow-up as the external criterion, were overall low. Using the M-CHAT-23 with follow-up as the screener for ASD risk (i.e., external criterion) is a limitation of this study. It has been found that the M-CHAT-23 given some variability in reliability as a single measure of ASD risk (Beuker et al., 2013; Charman et al., 2015). In comparison, when correlating the ESR items to the total score, the validity of the items is much higher (Table 2; r.cor). When this correlation is used, the ease of social engagement scores in the ball, book, and tickle activities have the highest validity.

It is well known that children with autism show less social engagement than their typically developing peers. Research has shown that low levels of social engagement is associated with ASD severity, with high-risk children showing significantly less social engagement than their low-risk peers (Campbell et al., 2016). To support this finding, our reliability analysis showed that ease of social engagement ratings scored the highest in reliability. Social engagement is an important target of treatment for young children with ASD. Thus, it is important to measure at the time of screening, but also during follow-up (Knott, Dunlop, & Mackay, 2006).

Overall, there was evidence that the hat and tickle activities performed less effectively, suggesting they could be removed to shorten the assessment. It is very clear that ease of social

engagement rating items should be included in future assessments, but the best type of activity used is not entirely clear. For the ball activity, the ease of social engagement rating was among the top scores across analyses and a ball is readily accessible in most environments. However, the book activity has more items, also had items with top scores across analyses, a book is readily accessible, and has better overall reliability. Thus, we would suggest using the book activity if time constraints allow for only one task.

Implication of findings

The findings of this study suggest that ease of social engagement ratings are the best measure within the ESR for detecting ASD risk. As demonstrated previously, it is well known that deficits in social engagement and behaviors are among the earliest signs of autism in children. Additionally, these skills are common and effective targets in intervention programs for ASD (Vernon et al., 2013). A recent study is developing an AI program that learns from patterns of behavior between children with autism and their typically developing peers (Shahamiri and Thabtah, 2020). The types of behaviors analyzed include social skills, attention switching, and communication skills, which are targeted in the ESR. As ASD screeners improve to become more accessible and accurate, it is important to consider which behaviors of autism are being measured and ensure social engagement is included.

Limitations

A limitation of this study is that we did not conduct statistical analyses for the effectiveness of activity scores (e.g., all ball scores combined; all book scores combined). We acknowledge the possibility that a subset of items could be better at discriminating ASD risk than individual items or the total score. Additionally, we acknowledge the limitation of using the M-CHAT-23 screener as the measure of ASD risk instead of a full diagnostic assessment. As mentioned previously, the

M-CHAT-23 with follow-up has had accuracy issues when discriminating ASD risk on its own (Beuker et al., 2013; Charman et al., 2015). While its sensitivity and specificity has been found to be better than other screening measures, we suggest a different measure of ASD risk should be used (Sunita & Bilszta, 2013). The validity of the items could vary significantly with a different external criterion, so further analysis should be conducted before the consideration of clinical adoption.

Conclusion

Children with ASD experience impaired social interaction and communication, which can negatively impact their social and behavioral development. Diagnosis becomes most reliable at age two, but concerning behaviors can arise before this age, making clinical intervention difficult. While the ESR assessment has been shown to differentiate high risk children from their peers, only the total score was examined. If the assessment was shortened, the possibility of adoption in clinic could increase, allowing for earlier intervention. Through this study's item analysis, the ability of each item and activity is better known, so that future decisions in autism screening using the ESR can be improved and modified.

References

Anagnostou, E., Jones, N., Huerta, M., Halladay, A. K., Wang, P., Scahill, L., Horrigan, J. P., Kasari, C., Lord, C., Choi, D., Sullivan, K., & Dawson, G. (2015). Measuring social communication behaviors as a treatment endpoint in individuals with autism spectrum disorder. *Autism*, *19*(5), 622–636. https://doi.org/10.1177/1362361314542955.

Beuker, K.T., Schjølberg, S., Lie, K.K., Swinkles, S., Rommelse, N.J, & Buitelaar, J.K. (2014).

ESAT and M-CHAT as screening instruments for autism spectrum disorders at 18 months in the general population: issues of overlap and association with clinical referrals. *European Child Adolescent Psychiatry*, **23**, 1081–1091. https://doi.org/10.1007/s00787-014-0561-8.

Campbell, S.B., Leezenbaum, N.B., Mahoney, A.S., Moore, E.L., & Brownell, C.A. (2017). Pretend Play and Social Engagement in Toddlers at High and Low Genetic Risk for Autism Spectrum Disorder. *Journal for Autism and Developmental Disorders*, 46(7), 2305-2316. doi: 10.1007/s10803-016-2764-y.

Cen, C. Q., Liang, Y. Y., Chen, Q. R., Chen, K. Y., Deng, H. Z., Chen, B. Y., & Zou, X. B. (2017). Investigating the validation of the Chinese Mandarin version of the Social Responsiveness Scale in a Mainland China child population. *BMC psychiatry*, *17*(1), 51. https://doi.org/10.1186/s12888-016-1185-y.

Charman, T., Baird, G., Simonoff, E., Chandler, S., Davison-Jenkins, A., Sharma, A., O'Sullivan, T. & Pickles, A. (2016). Testing two screening instruments for autism spectrum disorder in UK community child health services. *Developmental Medicine and Child Neurology*, 58: 369-375. https://doi.org/10.1111/dmcn.12874.

Charman, Tony (2003). Why is joint attention a pivotal skill in autism? *Philosophical Transactions of the Royal Society*,**358,** 315-324. http://doi.org/10.1098/rstb.2002.1199.

Corsello, Christina M. PhD. (2005). Early Intervention in Autism. *Infants & Young Children*, 18(2), 74-85. https://journals.lww.com/iycjournal/Fulltext/2005/04000/Early_Intervention_in_Autism.aspx

Desjardins, C. D., & Bulut, O. (2020). *Handbook of Educational Measurement and Psychometrics Using R*. CRC Press.

Elder, J. H., Kreider, C. M., Brasher, S. N., & Ansell, M. (2017). Clinical impact of early diagnosis of autism on the prognosis and parent-child relationships. *Psychology research and behavior management*, *10*, 283–292. https://doi.org/10.2147/PRBM.S117499.

Factor, R. S., Arriaga, R. I., Morrier, M. J., Mathys, J. B., Dirienzo, M., Miller, C. A., Southerland, A. M., Abowd, A., & Ousley, O. (2021). Development of an interactive tool of early social responsiveness to track autism risk in infants and toddlers. *Developmental Medicine and Child Neurology*, *64*(3), 323-330. https://doi.org/10.1111/dmcn.15035.

Fingert, B.S., Carter, A., Pierce, K., Stone, W. L., Wetherby, A., Scheldrick, C., Smith, C., Bacon, E., James, S. N., Ibañez, L., & Feinberg, E. (2019). Implementing systems-based innovations to improve access to early screening, diagnosis, and treatment services for children with autism spectrum disorder: An Autism Spectrum Disorder Pediatric, Early Detection, Engagement, and Services network study. *Autism*, *23*(3), 653–664. https://doi.org/10.1177/1362361318766238.

Goetz, C., Coste, J., Lemetayer, F., Rat, A.C., Montel, S., Recchia, S., Debouverie, M., Pouchot, S., Spitz,E., & Guillemin, F. (2013). Item reduction based on rigorous methodological guidelines is necessary to maintain validity when shortening composite measurement scales. *Journal of Clinical Epidemiology*, 66(7), 710-718. https://doi.org/10.1016/j.jclinepi.2012.12.015.

Guthrie, W., Swineford, L.B., Nottke, C. and Wetherby, A.M. (2013), Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry*, 54, 582-

590. https://doi.org/10.1111/jcpp.12008

Knott, F., Dunlop, A.-W., & Mackay, T. (2006). Living with ASD: How do children and their parents assess their difficulties with social interaction and understanding? *Autism*, 10(6), 609–617. https://doi.org/10.1177/1362361306068510

Mandell, D.S., Novak, M.M., Zubritsky, C.D. (2005) Factors Associated with Age of Diagnosis Among Children with Autism Spectrum Disorders. *Pediatrics*, 116 (6), 1480–1486. Doi.org/10.1542/peds.2005-0185.

McCowan, R. J., & McCowan, S. C. (1999). Item analysis for criterion referenced tests. *Center of Development for Human Services*. https://files.eric.ed.gov/fulltext/ED501716.pdf .

Mundy, P., Crowson, M. (1997). Joint Attention and Early Social Communication: Implications for Research on Intervention with Autism. *J Autism Dev Disord* **27**, 653–676. https://doi.org/10.1023/A:1025802832021.

OpenAI. (n.d.). ChatGPT [Computer software]. Retrieved January-March, 2023, from https://openai.com/products/chat-gpt-3/

Pallett, P.M., Cohen, S.J. & Dobkins, K.R. (2014). Face and Object Discrimination in Autism, and Relationship to IQ and Age. *J Autism Dev Disord* **44**, 1039–1054. https://doi.org/10.1007/s10803-013-1955-z

Pan, C.Y. (2009). Age, social engagement, and physical activity in children with autism spectrum disorders, *Research in Autism Spectrum Disorders*, 3(1), 22-31, https://doi.org/10.1016/j.rasd.2008.03.002.

Pierce, K., Carter, C., Weinfeld, M., Desmond, J., Hazin, R., Bjork, R., & Gallagher, N. (2011). Detecting, studying, and treating autism early: the one-year well-baby check-up approach. *The Journal of pediatrics*, *159*(3), 458–465.e4656.

https://doi.org/10.1016/j.jpeds.2011.02.036.

Posar, A., & Visconti, P. (2020). Is it autism? Some suggestions for pediatricians. *Turkish Pediatrics archives*, *55*(3), 229–23. https://doi.org/10.14744/TurkPediatriArs.2020.59862.

Revelle, W. (2022). *Procedures for Psychological, Psychometric, and Personality Research*. RDocumentation. https://www.rdocumentation.org/packages/psych/versions/2.2.9

Robins, D.L., Fein, D., Barton, M.L. & Green, J.A. (2001). The Modified Checklist for Autism in Toddlers: An Initial Study Investigating the Early Detection of Autism and Pervasive Developmental Disorders. *Journal of Autism Developmental Disorders*, **31**, 131–144. https://doi.org/10.1023/A:1010738829569.

Shahamiri, S.R., Thabtah, F. (2020). Autism AI: a New Autism Screening System Based on Artificial Intelligence. *Cognitive Computation*, **12**, 766–777. https://doi.org/10.1007/s12559-020-09743-3

Sim, S.M., & Rasiah, R.I. (2006). Relationship Between Item Difficulty and Discrimination Indices in True/False-Type Multiple Choice Questions of a Para-Clinical Multidisciplinary Paper. *Annals Academy of Medicine Singapore*, 35(2), 67-71. https://pubmed.ncbi.nlm.nih.gov/16565756/.

Sunita, and Bilszta, J.L. (2013), Early identification of autism: A comparison of the Checklist for Autism in Toddlers and the Modified Checklist for Autism in Toddlers. *Journal of Pediatrics and Child Health*, 49: 438-444. https://doi.org/10.1111/j.1440-1754.2012.02558.x

Vernon, T.W., Koegel, R.L., Dauterman, H., & Stolen, K. (2013). An Early Social Engagement

Intervention for Young Children with Autism and their Parents. *Journal of Autism and Developmental Disorders,* 42(12), 2702-2717. doi: 10.1007/s10803-012-1535-7

Williams, J.H.G., Whiten, A. & Singh, T. (2004). A Systematic Review of Action Imitation in Autistic Spectrum Disorder. *Journal of Autism and Developmental Disorders,* **34**, 285–299. https://doi.org/10.1023/B:JADD.0000029551.56735.3a

Wong, V., Lai-Hing, S.H., Wing-Cheong, L., Lok-Sum, J.L., Po-Ki, P.H., Wai-Ling, C.L., Fung, C.W., & Chung, B. (2004). A Modified Screening Tool for Autism (Checklist for Autism in Toddlers [CHAT-23]) for Chinese Children. *Pediatrics,* 114 (2), DOI: e166–e176. 10.1542/peds.114.2.e166.

Appendix A

The Early Social Responsiveness Scale

Instructions for the interactive assessment of early social responsiveness **Step 1: Present each task in a standardized way.** For each task, follow the instructions in the shaded boxes. For Task 2, use a ball about 3 inches across. For Task 3, use a board book between 2 and 4 inches. **Step 2: Rate the target behaviors.** Answer the questions about the child's behavior, as you complete each task. **0 = Yes,** the child showed the behavior. **1 = No,** the child did not show the behavior. **Step 3: Rate ease of engagement for each task.** Complete the engagement ratings, as you complete each task. **0 = Yes, very easy to engage** -- this rating indicates that the child is attuned to the examiner's actions, is readily available for interaction, and attends to the examiner with *anticipation* and *expectancy*, requiring minimal effort from the examiner **1 = Somewhat easy to engage** **2 = Hard to engage** **Step 4: Determine the Total Score.** In the far-right column, circle the appropriate score or converted score that corresponds to each rating. Sum all scores in the far-right column to determine a **Total Score**.		**Scores**
TASK 1 Smiling and saying "hello"	Instructions: When you are ready to start, smile and say in a playful tone, "**Hi (insert child's name).**" **PAUSE** for **2 seconds** and say, "**Are you ready to play with some new toys?**" Lean in and keep smiling for **2 seconds**.		
	1. Did the child look you in the eyes?	**0=Yes** **1=No**	**0 1**
	2. Did the child smile?	**0=Yes** **1=No**	**0 1**
	3. **Engagement rating:** Overall, during Task 1, was the child easy to engage, taking little to no effort from you?	**0=Yes, very easy to engage; child attends to the examiner with *anticipation* and *expectancy*** **1 = Somewhat easy to engage (→ converted to a score of 2)** **2 = Hard to engage**	**0 2**
TASK 2 Ball play	Instructions: Hold the ball to the right, about 12 inches from your head, at your eye level. Say "**Look at my ball.**" Watch to see if child looks at the ball then back to your eyes.		
	4. Did the child look at the ball and then back at your eyes?	**0=Yes** **1=No**	**0 1**
	Instructions: Say "**Let's play ball. Ready, set, GO!**" See if child will roll or throw the ball back to you, then repeat at least **2 times, but not more than 4**.		
	5. Did the child roll or throw the ball back to you one or more times?	**0=Yes** **1=No**	**0 1**
	6. Did the child smile?	**0=Yes** **1=No**	**0 1**
	Instructions: On the **3rd roll** say, "**Ready, set…..**" **PAUSE** for **5 seconds** "**GO!**"		
	7. Did the child look at your eyes after the PAUSE?	**0=Yes** **1=No**	**0 1**
	8. **Engagement rating:** Overall, during Task 2, was the child easy to engage, taking little to no effort from you?	**0=Yes, very easy to engage; child attends to the examiner with *anticipation* and *expectancy*** **1 = Somewhat easy to engage (→ converted to a score of 2)** **2 = Hard to engage**	**0 2**

TASK 3 Book	Instructions: Hold the book to the right, about 12 inches from your head, at your eye level. Say, "**Look at my book.**"		
	9. Did the child look at the book and then back at your eyes?	**0=Yes** **1=No**	**0 1**
	10. Did the child smile?	**0=Yes** **1=No**	**0 1**
	Instructions: Present the book, within 6 inches in front of child, as you read the 1st page. Modeling turning the page, read the 2nd page, then say "**Let's see what's next**" (**wait for the child to turn the page**). If the child does not turn the page, turn the page and say, "**Where is the (insert name of picture in the book)?**" only once. Then say, "**Can you turn the page?**" at least once, but not more than twice. Then continue turning the pages at least **3 times consecutively** (see below for PAUSE instructions).		
	11. "Did the child turn one or more pages?"	**0=Yes** **1=No**	**0 1**
	12. Did the child point to or tap on a picture in the book?	**0=Yes** **1=No**	**0 1**
	13. Did the child smile?	**0=Yes** **1=No**	**0 1**
	Instructions: After 3 consecutive page turns, say "**Let's see what's next….." PAUSE** for **5 seconds** as you hold the page with your thumb, preventing the child from turning the page….. then turn the page.		
	14. Did the child look at your eyes after the PAUSE?	**0=Yes** **1=No**	**0 1**
	15. **Engagement rating:** Overall, during Task 3, was the child easy to engage, taking little to no effort from you?	**0=Yes, very easy to engage; child attends to the examiner with *anticipation* and *expectancy*** **1 = Somewhat easy to engage (→ converted to a score of 2)** **2 = Hard to engage**	**0 2**
TASK 4 Putting book on your head as a hat	Instructions: As you playfully put the book on your head, ***gasp*** while smiling. Then say, "**Where's the book?**" Wait **2 seconds**, then say, "**It's on my head, it's a hat**!"		
	16. Did the child look you in the eyes?	**0=Yes** **1=No**	**0 1**
	17. Did the child smile?	**0=Yes** **1=No**	**0 1**
	18. **Engagement rating:** Overall, during Task 4, was the child easy to engage, taking little to no effort from you?	**0=Yes, very easy to engage; child attends to the examiner with *anticipation* and *expectancy*** **1 = Somewhat easy to engage (→ converted to a score of 2)** **2 = Hard to engage**	**0 2**
TASK 5 Smiling and tickling	Instructions: Hold your hands up in front of you, wiggle your fingers, and say "**I'm gonna tickle you**." Wait **2 seconds**, then say "**I'm gonna get you, I'm gonna get you, I'm gonna get you**" while ***slowly*** leaning in toward the child. ***Gently*** tickle the child on the belly or arms, saying "**tickle, tickle, tickle**" Repeat this sequence for **a 2nd time,** starting with **I'm gonna get you, … etc".**		
	19. Did the child look you in the eyes?	**0=Yes** **1=No**	**0 1**
	20. Did the child smile?	**0=Yes** **1=No**	**0 1**
	Instructions: On the **3rd time** begin again, as above, and say, "**I'm gonna get you…..**" then **PAUSE** for **5 seconds** before saying, "**I'm gonna get you, I'm gonna get you**" and then gently tickling, etc.		

	21. Did the child look at your eyes after the PAUSE?	**0=Yes** **1=No**	**0 1**
	22. **Engagement rating:** Overall, during Task 5, was the child easy to engage, taking little to no effort from you?	**0=Yes, very easy to engage; child attends to the examiner with *anticipation* and *expectancy*** **1 = Somewhat easy to engage (→ converted to a score of 2)** **2 = Hard to engage**	**0 2**
	TOTAL SCORE: Higher scores indicate poorer early social responsiveness skills *(range=0-27)*		

Note: Originally from Factor et al., 2014

Appendix B

R Codes

Item Difficulty

```
"item_diff <- colMeans(ESR_Item_Scores, na.rm = TRUE)".

round(item_diff, 3
```

Item Discrimination

```
total_score <- rowSums (ESR_Item_Scores, na.rm = TRUE

item_discr <- cor(ESR_Item_Scores, total_score)
```

Item Discrimination Index

```
high_performers <- subset(ESR_Item_Scores, total_score > median(total_score))

low_performers <- subset(ESR_Item_Scores, total_score <= median(total_score))

p_high <- mean(high_performers$item_response)

p_low <- mean(low_performers$item_response)

model <- glm(item_response ~ performance_score, data = your_data, family =

binomial())

coefficients <- coef(model)

odds_ratio <- exp(coefficients[2])

log_odds_high <- coefficients[1] + coefficients[2] *median(performance_score)

log_odds_low <- coefficients[1]

idi <- odds_ratio * (p_high - p_low)
```

Item Reliability Index

```
reliability <- alpha(ESR_Item_Scores)

reliability.index <- item.stats$alpha.drop
```

Item Validity

```
item.stats <- alpha(ESR_Test_Items)$item.stats

validity.index <- item.stats$r.cor

itemval <- cor(ESR_Item_Scores, ESR_Item_Scores_With_MCHAT$MCHAT

valindex <- fischerz(itemval)
```

Descriptive Statistics

```
Sapply(ESR_Item_Scores, var)

Sapply(ESR_Item_Scores, sd)

Sapply(ESR_Item_Scores, mean)

Summary(ESR_item_Scores)
```

Note: The "psych" package in R was downloaded and used for some analyses.