Published in final edited form as:

J Autism Dev Disord. 2024 May ; 54(5): 1820–1833. doi:10.1007/s10803-023-05931-w.

Differential Performance of Social Communication Questionnaire Items in African American/Black vs. White Children

Ethan Dahl[1], **Eric J. Moody**[2], **Brian Barger**[3], **Steven Rosenberg**[4], **Carolyn DiGuiseppi**[5], **M. Daniele Fallin**[6], **Li-Ching Lee**[7], **Lisa Wiggins**[8]

[1]Department of Education, Health, & Behavior Studies, College of Education & Human Development, University of North Dakota, 231 Centennial Dr Stop 7189, Grand Forks, ND 58202-7189, USA

[2]Wyoming Institute for Disabilities, University of Wyoming, Laramie, WY, USA

[3]Center for Leadership in Disabilities, Georgia State University, Atlanta, Georgia

[4]Department of Psychiatry, University of Colorado School of Medicine, Aurora, CO, USA

[5]Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[6]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[7]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA

[8]National Center on Birth Defects and Developmental Disabilities, Centers for Disease Control and Prevention, Atlanta, Georgia

Abstract

Screening for autism spectrum disorder (ASD) is an essential early step in the identification process and inaccurate screening may lead to significant delays in the onset of treatment. Past research has highlighted discrepancies in the performance of ASD screening tools such as the Social Communication Questionnaire (SCQ) among certain racial and ethnic groups. The current study explored the functioning of the SCQ among African American/Black and White respondents based on item level performance on the measure. Differential Item Functioning (DIF) analyses showed that 16 (41%) items of the SCQ functioned differently for African American/Black respondents when compared to White respondents. Implications, such as the potential for delayed diagnosis and treatment, and the influence on downstream outcomes, are discussed.

✉Ethan Dahl, Ethan.Dahl@und.edu.

Author Contributions ED, EJM, BB, and SR were involved in conceptualization. ED, EJM, and BB were involved in methodology. ED and EJM were invloved in formal analysis. ED and EJM were involved in writing—original draft preparation. ED, EJM, BB, SR, CD, MDF, LL, and LW were involved in writing—review and editing.

Conflict of interest The authors have not conflicts of interest to state regarding this manuscript.

Keywords

Autism spectrum disorder (ASD); Social communication questionnaire (SCQ); Differential item functioning (DIF); Racial disparities; Autism identification

Introduction

Despite increased societal awareness of autism spectrum disorder (ASD), most children are not diagnosed until over four years of age, and African American/Black[1] children are diagnosed significantly later than White children (Maenner et al., 2020). The historical disparity in ASD prevalence for African American/Black children has recently become less prominent (Maenner et al., 2020; Mandell et al., 2002, 2009), as diagnostic age between African American/Black children and their peers has become more equivalent among 4 year-olds (Shaw et al., 2021) and 8 year-olds (Maenner et al., 2020). However, variation does still exist between geographic sites (Maenner et al., 2020), suggesting racial disparities may persist. Delays in diagnosis have continued despite evidence that overall prevalence of ASD is similar between African American/Black and White children in recent years (Maenner et al., 2020). Further, prevalence of ASD has also been shown to be associated with socioeconomic status (Durkin et al., 2017; Kelly et al., 2019). Lack of diagnosis is more likely when medical records are incomplete or missing, which is especially true for African American/Black children (Imm et al., 2019). This delay in ASD diagnosis for African American/Black children ranges from a few months up to 7 years and potentially leads to years of missed opportunities for early intervention (EI), services, and supports known to improve child outcomes (Daniels & Mandell, 2014; Maenner et al., 2020). Moreover, past literature on ASD highlights significant racial and ethnic disparities in service use (Liptak et al., 2008; Magana et al., 2012; Magaña et al., 2013). For instance, African American/Black and Hispanic mothers report less access to specialty services, educational services, community services, and acute care for their child, and lower quality of care in comparison to White individuals (Smith et al., 2020). Additionally, African American/Black preschool children with ASD tend to have greater use of psychotropic medication paired with less use of behavioral therapy (Wiggins et al., 2021), suggesting a lack of a complete services and supports plan.

The reasons for these disparities are complex and due to numerous factors. One area that has gained interest recently is the intricate community-based monitoring, screening, and referral processes that span multiple systems of care (Barger, Rice, & Roach, 2018; Barger, Rice, Simmons, et al., 2018). This aspect of the diagnostic process encompasses the first critical steps in identifying children with ASD and ensuring equitable opportunities for diagnosis, services, and supports. However, there is growing evidence that the tools used to screen for ASD perform differently based on race and ethnicity. One study on the Modified Checklist for Autism in Toddlers (M-CHAT), a popular ASD screening tool for children 18–30 months of age, found that Hispanic families tended to score higher (indicating increased

---

[1] For the purposes of this manuscript, the term African American/Black will be used to represent descriptions of African American, Black, and non-Hispanic Black.

presence of ASD characteristics) but had fewer successful referrals for follow up testing than Non-Hispanic families (Windham et al., 2014). Other studies suggest that African American/ Black children tend to be screened with the M-CHAT later than White children (Herlihy, 2014), and that M-CHAT results are less accurate in African American/Black children, even when the instrument was nearly universally implemented within a healthcare system (Guthrie et al., 2019). Specifically, Guthrie et al. (2019) found that African American/Black children did not differ from other non-white children but that all children of color and those of lower-income showed 2–3 times the rate of positive screens in comparison to white, suburban, higher-income children and those with private insurance. However, Guthrie and et al, (2019), did not speculate as to why there was a higher positive screening rate in these groups.

Racial disparities in the performance of other common ASD screening tools have also been reported. Two studies used a well characterized and diverse community-based sample to examine the performance of the Social Communication Questionnaire (SCQ) and Social Responsiveness Scale (SRS) across racial and ethnic groups (Moody et al., 2017; Rosenberg et al., 2018). Both the SCQ and SRS were validated for preschool children older than 30 months (the upper age limit of the M-CHAT), and the SCQ was adapted from a gold-standard diagnostic parent interview for ASD. These studies revealed that both tools had specificity rates well below acceptable levels across multiple socio-demographic populations (Moody et al., 2017) and that the SCQ was less accurate as a measure for African American/ Black children (Rosenberg et al., 2018). Original psychometric data on the SCQ (Rutter et al., 2003) indicated that the measure has high internal consistency ($\alpha = 0.87$), high sensitivity (96%), and moderate specificity (80%). More recent work has found that internal consistency of the SCQ varies by sample. For instance, Rosenberg et al. (2018) found high internal consistency for children with ASD or other developmental disorders (recruited from clinical and educational sources) ($\alpha = 0.89$) but lower internal consistency for children recruited from a random sample of state birth records ($\alpha = 0.77$). It was additionally noted though that SCQ data were less accurate because of higher rates of false positive screenings, similar to the MCHAT in this sample (Rosenberg et al., 2018). Snow and Lecavalier (2008) also found slightly reduced internal consistency ($\alpha = 0.81$) for preschool children. Together with data on the M-CHAT, these findings suggest that there may be fundamental differences in how these instruments function across groups, including racial and ethnic groups, indicating that further investigation is warranted into these tools.

One reason for discrepancies in ASD screening tool performance could be differential item response patterns by racial and ethnic group (Herlihy, 2014; Reyes et al., 2021). Different item response patterns refer to the ways in which groups such as racial or ethnic groups may respond on an instrument or a specific item in a way that is different in comparison to another group. For instance, this could include one group tending to respond with one end of the scale, or the extremes, while another group may tend to respond with middle responses. These differential response patterns may contribute to systematic variation in screening tool performance within specific groups that is masked when only broader measures of performance are examined (e.g., sensitivity or specificity; Moody et al., 2017). If so, identifying individual items that perform differently for different children is an important first step to improving ASD screening tools, particularly for African American/Black children who may continue to be at higher risk of under-detection. One such approach to evaluating item response patterns is the statistical approach of Differential Item Functioning (DIF).

DIF is a psychometric method used to evaluate whether different subgroups respond differently to each item within a measure (Swaminathan & Rogers, 1990). It is particularly useful in identifying potential bias among the items of the instrument. Specifically, DIF analyzes individual items of a measure and compares groups on their probability of endorsing the item. For binary items such as those of the SCQ, DIF determines if one group is more likely to score a 1 (vs. 0) on that item. DIF analyses were initially used as part of educational testing to help understand if test items were biased for certain groups (e.g., male vs. female, or different ethnicities). More recently DIF has been applied in several areas including education screening (Barger et al., 2020; Goodrich et al., 2019; Hope et al., 2018), psychological testing (Choi et al., 2019; Cicero et al., 2019; Lin et al., 2018; Murphy et al., 2019; Rice et al., 2020; Wei et al., 2015), and healthcare evaluation (Peipert et al., 2018; Reyes et al., 2021; Scott et al., 2010). Ideally, DIF analyses would be performed in the process of instrument development; however more often it is applied after an instrument has already been established. DIF can be either uniform or nonuniform. Uniform DIF indicates that there is a consistently higher or lower probability for one group to endorse an item in comparison to another group at all levels of the underlying latent trait. Nonuniform DIF indicates that there is an interaction between the likelihood to endorse the item and group membership by total score on the measured trait.

The Study to Explore Early Development (SEED; Schendel et al., 2012) presents a unique opportunity to explore the psychometric performance of the SCQ at the item level. SEED constitutes the largest sample of preschool children with ASD, other developmental delays and disabilities, and children from the general population in the United States. The SEED sample is large, diverse, and well characterized, including over 7000 children across study phases 1 and 2, ~ 50% non-White respondents, over 1400 children with ASD, and families of diverse educational backgrounds. For a detailed description of the SEED 1 sample see (Diguiseppi et al., 2016; Wiggins et al., 2015), and Bradley et al. (2018) for details on the rigorous clinical characterization of the study. The SCQ was administered to every child who enrolled in the SEED study regardless of ASD diagnosis, and steps were taken in the second phase of SEED to increase recruitment of African American/Black children and families. Therefore, the combined SEED Phase 1 and SEED Phase 2 samples were used to explore differential item level performance of the SCQ across African American/Black and White groups.

Method

Participants

Overall SEED Sample—Data for this analysis come from the first two phases of The Study to Explore Early Development (SEED; Phase 1 enrollment: 2007–2011; Phase 2 enrollment: 2012–2016), a multi-site study that evaluated children 30–68 months of age with ASD, with non-ASD developmental delays, and from the general population (N = 7271). Participants in SEED do not meaningfully differ between Phase 1 and Phase 2 other than enrollment date as procedures were consistent across both phases. There were six study sites, located in California, Colorado, Georgia, North Carolina, Pennsylvania, and Maryland. See Schendel et al. (2012) for additional details about the SEED study, recruitment, and protocol details. Families were recruited from three sources. First, children with ASD could be referred by their family (i.e., self-referred; Phase 1 only). Second, each site worked with local community service providers, including on-site clinical services at five sites, to recruit children with a broad range of developmental delays and disabilities including ASD. Finally, each site worked with their state's health department to recruit a random sample of children from the community via birth certificate records. These children were mostly typically developing, but some children with ASD or another developmental delay or disability (DD) were recruited from this source due to the random sampling procedure. All children completed an in-person clinical evaluation, after which they were categorized as having ASD, having another DD, or being in the population comparison (POP). The final classification of ASD, DD, or POP was based on the SEED final classification algorithm (Wiggins et al., 2015). This algorithm took into account results from the ADOS and ADIR as well as best practice guidelines for diagnosis of ASD. In particular, classification of all children was based on clinical judgement of highly experienced clinicians from in-person administration of gold-standard diagnostic instruments. See Wiggins et al. (2015) for additional details of the classification process.

Study Sample—The total sample size of the combined SEED 1 and SEED 2 studies who received a final study group classification was 4799. From the total SEED 1 and 2 samples, a subset of children was selected for the current study ($n = 3376$). Specifically, the sample analyzed here consisted of only those children who were non-Hispanic African American/ Black or non-Hispanic White (see Demographic Variables, below), had completed the SCQ and received a classification of ASD, DD, or POP. Additionally, the sample was restricted to children ≥ 4 years of age to fit within the recommended age usage of the SCQ. The DD and POP SEED study classification groups were combined to form a Non-ASD group for the analyses conducted here. Specifically, the analyses were used to compare children in the SEED sample with an ASD classification and those without an ASD classification who were labeled as Non-ASD. Total sample sizes by demographic variables are outlined in Table 1.

Measures and Procedures

Social Communication Questionnaire (SCQ)—The key measures for this study are the Social Communication Questionnaire-Current (SCQ; Rutter et al., 2003) and demographic variables. The SCQ is a 40-item screening tool designed to evaluate a child's current risk for ASD among children who have been referred for additional assessment. The SCQ asks caregivers to report on the presence of child behaviors that are characteristic of ASD, such as inconsistent eye contact and minimal use of gestures. The SCQ takes about 10 min to complete. In the SEED study, the SCQ was the first standardized measure collected and was administered by phone interview to all families to determine risk for ASD and subsequent assessment procedures. A knowledgeable parent or guardian was the respondent. All items are scored as either the presence (1) or absence (0) of the described behavior. A total score is calculated based on items 2 through 40 (maximum score—39).

A positive screen was defined as an SCQ total score of 11 or higher. While the typical SCQ cutoff is 15, previous research has found that a cutoff of 11 maximizes sensitivity and specificity for children ≤ 4 years (Wiggins et al., 2007) and was used in SEED. Moreover, it was found that a cutoff of 11 in SEED was optimal while accounting for race, education, and income (Rosenberg et al., 2018). As the SCQ was originally designed as a questionnaire for children ≥ 4 years, the measure performs better among those children but still maintains good discriminative validity for distinguishing between younger children with ASD and those without an ASD diagnosis (Berument et al., 1999). Children at risk for ASD (i.e., SCQ score ≥ 11 at enrollment, a previous ASD diagnosis, or child observed showing ASD symptoms during the basic clinical examination), regardless of source population, were given additional developmental assessments, including the Autism Diagnostic Observation-Schedule (ADOS; Lord et al., 2000) and the Autism Diagnostic Interview-Revised (ADI-R; Gotham et al., 2007).

Demographic Variables—The demographic variables for these analyses came from a Maternal Interview that was administered after SEED study enrollment. The Maternal Interview is a structured phone interview conducted by trained interviewers at each site, with ongoing reliability assessment (Bradley et al., 2018; Schendel et al., 2012). All interviewers underwent a rigorous, multi-day training process prior to being approved for interview implementation which included coding a standardized validation set of interviews, cross-site evaluation, and on-site reliability assessments. Ongoing reliability assessment was conducted throughout data collection to ensure reliability and validity of these data. Additional details on the Maternal Interview procedure can be found in Schendel et al. (2012). For SEED 1, the respondent could be any knowledgeable primary caregiver, although 98% were biological mothers. For SEED 2, only biological mothers were respondents. Race and ethnicity were self-reported, and the respondent was permitted to select as many categories as desired. In both SEED phases, the respondent reported the categories of both the biological mother and father. If race was unknown in SEED 1 it was categorized as unknown and excluded in the current analysis. The biological mother's race was used to indicate whether the child was African American/Black, White, another race, or mixed race, since data on biological mother's race were more complete and are more likely to be accurate since they were self-reported. Respondents who selected African American/Black plus any other racial category were not included in the African American/Black category. Similarly, respondents who chose White plus any other racial category were not included in the White category. Respondents who selected only White or only African American/Black were categorized as such. This analysis did not examine ethnicity or races other than African American/Black and White to simplify the interpretation of the group comparisons.

Analysis Plan

All analyses were conducted in SPSS v. 26. Analysis of these data was conducted via three steps. First, we computed descriptive demographics (Table 1) of the two primary racial groups reported on here. Chi-square tests for independence were performed to examine if there were significant differences between African American/Black and White respondents in key variables of interest including ASD classification (2 levels), maternal education (4 levels), or current household income (7 levels). Variables that were significantly different were considered as covariates. However, correlations between potential covariates were first computed to determine if any of the variables were collinear. In such cases, only one of the variables would be included in the model as a covariate. Second, given that prior research has documented reduced internal reliability in some groups, we computed Cronbach's alpha for the overall sample, and each racial group, and compared them with a Feldt test (Feldt, 1969). The Feldt test compares two alpha coefficients from independent samples to assess the equality of the alpha coefficients. Finally, to determine if individual items performed differently in the White vs. African American/Black sample, we used DIF.

Logistic regression was used to examine uniform and nonuniform DIF simultaneously (Swaminathan & Rogers, 1990). Separate models were computed for each SCQ item with race as a fixed factor (White vs African American/Black), SCQ total score, and child's SCQ total score minus the item included to predict item endorsement. White children were used as the reference group for the race factor based on higher sensitivity and specificity of the SCQ in this group. Inclusion of the total score minus the item being modeled allows for the comparison of White versus African American/Black children on performance of each item while matching them on total score, thereby factoring out the contribution of that item to the total score. Uniform DIF is indicated if there is a significant main effect for race with no interaction effect; nonuniform DIF is indicated if the interaction term between item endorsement and race by total score is statistically significant ($p < 05$). Due to the need to run multiple models for these analyses significance levels were adjusted using the Benjamini–Hochberg procedure (Benjamini & Hochberg, 1995) to correct for Type I error. To define statistical significance of each model, the Benjamini–Hochberg corrected p value was compared to the unadjusted p value. Any model where the Benjamini–Hochberg p value is less than the unadjusted p value is considered significant. This correction was done for each separate model and for both the main effect of race (White vs. African American/ Black), and the interaction between race and total SCQ score. The covariate (total SCQ score minus individual item contribution) was not evaluated.

Results

Demographic results are displayed in Table 1. There were significant differences between African American/Black and White participants on ASD classification χ^2 (1, $N = 3376$) = 11.69, $p < 0.001$, maternal education χ^2 (3, $N = 3373$) = 446.87, $p < 0.001$, and household income χ^2 (6, $N = 3260$) = 861.62, $p < 0.001$. Because of the high degree of correlation between these variables and African American/Black and White race ($p < 0.001$), the subsequent analysis will only use race to represent the potential differences in item level response to avoid collinearity. The combined SEED 1 and 2 reliability was $\alpha = 0.90$ for all White and African American/Black participants ($N = 3376$). Reliability differed when stratifying by White ($N = 2496$, $\alpha = 0.90$) and African American/Black ($N = 880$, $\alpha = 0.88$) respondents ($W = 0.78$, $p < 0.001$). However, internal reliability levels were extremely high for both groups. Therefore, this statistical difference may have little clinical meaning as discussed below, under Limitations.

DIF analysis results indicated that 16 SCQ items showed DIF; 13 items showed non-uniform DIF (Table 2) while 3 items showed only uniform DIF (Table 3), leaving the remaining 23 items of the SCQ showing no DIF (excluding item 1 which is not used to calculate the total score). All 13 items that showed non-uniform DIF simultaneously showed uniform DIF however, when this occurs only non-uniform DIF is interpreted. The 13 non-uniform DIF items showed significant differences in the likelihood of item endorsement by race (White vs. African American/Black); however, item endorsement was inconsistent in the way that they differed. These items included odd or repetitive phrases (item 3), pronoun mixing (item 5), particular ordering or rituals (item 8), appropriate facial expressions (item 9), odd preoccupying interests (item 11), particular friends or a best friend (item 19), pointing (item 22), gestures other than pointing or pulling (item 23), head nodding for ‘yes’ (item 24), head nodding for ‘no’ (item 25), looking and using gestures and sounds to get attention in order to get an object (item 32), pretend play (item 35), and imaginative games (item 39). The interpretation of item endorsement is beyond the scope of the current analysis but includes interactions such as greater levels of endorsement for White respondents at higher total SCQ scores for the particular item. The 3 uniform DIF items showed significant differences in the likelihood of item endorsement by race (White vs. African American/Black). These items included socially inappropriate questions (item 3), insistence or use of repetitive language (item 7), and unusual interest in sensations (item 14). All items were significantly less likely to be endorsed by African American/Black respondents than White respondents (Item 4—*OR* = 0.52, Item 7—*OR* = 0.39, Item 14—*OR* = 0.57).

Discussion

We found that the SCQ had high internal consistency for both White (α = 0.91) and African American/Black respondents (α = 0.88). However, the alpha coefficients of White and African American/Black respondents differed significantly. Additionally, DIF analyses identified that 16 (41%) of the 39 items analyzed showed differential responses by racial group. These findings highlight variance in SCQ screen performance based on race that could influence interpretation of the total score and lead to inconsistent referral for diagnosis, services, and supports, potentially slowing identification and early intervention opportunities.

In terms of differential item performance, African American/Black respondents were significantly less likely to endorse three items across all levels of total SCQ score. These items ask about socially inappropriate questions or statements, repetitive speech, and the focus on certain sensations. The 13 that showed nonuniform DIF indicate that the relationship between the overall score on the SCQ and the performance on the specific items was not constant for the racial groups considered in this analysis. For these items, certain items display significantly higher likelihood of African American/Black respondents endorsing the item at lower total SCQ scores but significantly lower likelihood of African American/Black respondents endorsing the item at higher total SCQ scores, or vice versa. These items asked parents about odd phrases, pronoun misuse, particular phrases or rituals, repetitive phrases, inappropriate facial expressions, odd interests, friendships, pointing and hand gestures, head shaking and nodding, conventional gestures, and make-believe/imaginative games. Because most of the differential responses are inconsistent within group (i.e., nonuniform), these differences may be harder to detect at the instrument level, highlighting the need to conduct additional evaluation of the SCQ and other ASD screening tools to ensure equitable performance across racial groups. While it is beyond the scope of the current manuscript to evaluate each nonuniform DIF item, it is the hope that the current work will encourage future analyses to evaluate the details of the interactions between each item endorsement, racial group, and total SCQ score.

We did not examine reasons for the differential item performance on the SCQ. Some plausible explanations could be due to sociodemographic differences noted in Table 1 (i.e., African American/Black mothers had lower education and income than White mothers in this sample). Specifically, these sociodemographic differences could lead to conflicting interpretations of some items due to comprehension level or environmental distractions for low-income families. Some other possible explanations are cultural differences that lead to different response patterns or regional variations in the interpretations of some questions. Additional research is needed to understand the factors that lead to these differences, which could lead to the development and revision of ASD screening tools that perform more consistently across racial and other demographic groups. Solutions may include the construction of new item wordings, unique cutoff scores for differing groups, or the removal/ addition of scale items.

The differential item performance identified here could have some degree of impact on downstream steps in the diagnostic, services, and supports processes. While screening tools are only one part of the identification process (Filipek et al., 2000), the consequences for systematic differences in how these screening tools perform across groups can be substantial (Moody et al., 2017). For instance, decreased sensitivity among specific groups can lead to increased disparities in the diagnosis of ASD. Decreased specificity among specific groups can lead to increased public health burden due to unnecessary tests, increased parental anxiety, and concern. Given that there is often high distrust of healthcare systems by traditionally underserved communities (Alpers, 2018; Kennedy et al., 2007), these added disparities may further exacerbate healthcare system distrust (Armstrong et al., 2013), and lead families to avoid care. Given the critical role of screening in ASD identification, the potential for decreased sensitivity of screening tools in some groups could exacerbate racial differences in the presence and timing of an ASD diagnosis, which may impact referral for specialized services and supports and thereby influence the building of an environment of informed support individuals around the child (Fernell et al., 2013). On the other hand, the potential for decreased specificity in some groups could place additional burden on an already taxed healthcare system and increase parent anxiety, concern, and distrust in the healthcare system. Overall, these findings suggest that additional research is needed to understand how screening tools perform across racial, ethnic, and socioeconomic groups (Soto et al., 2015), and they may need to be adapted to increase accuracy across certain groups if psychometric problems such as lower reliability or DIF are found.

The findings of the current study suggest that clinicians should use caution when using the SCQ to decide whether to refer a child for evaluation. This may especially be true when working with African American/Black children. This is consistent with other findings that have found that this, and other screening tools perform differently among racial and ethnic groups. Overall, more work must be done to improve ASD screening tools to foster additional confidence among individuals using them.

Research outlining differential psychometric properties of screening tools across African American/Black populations relative to White children is limited in scope. Most research focuses on Hispanic versus White comparisons and tends to rely on the M-CHAT for infants and toddlers (Guthrie et al., 2019; Windham et al., 2014), or small clinical samples with few details related to clinical phenotypes or how diagnoses were rendered (Albores-Gallo et al., 2012; Guthrie et al., 2019; Herlihy, 2014; Scarpa et al., 2013; Soto et al., 2015; Windham et al., 2014). The near exclusive focus of this literature on the M-CHAT is particularly problematic as the median age of ASD identification is currently 51 months, i.e., 21 months after the M-CHAT's upper age cut-offs; this suggests that most youth are currently identified with screening tools other than the M-CHAT, like the SCQ (Daniels & Mandell, 2014; Maenner et al., 2020). Thus, there is a need to investigate racial/ethnic disparities in the psychometric properties of screening tools developed for identification in older children, especially for African American/Black children.

Item level differential functioning could be the result of several factors, such as unique interpretation of the content of items, completing the tool in distracting environments, or wording that is too complex or poorly translated, to name a few. Regardless of the causes of these psychometric differences, the potential impact on health outcomes for African American/Black families is substantial. Therefore, it is important that ASD screening tools are equally valid among those with diverse backgrounds so that they do not contribute to persistent racial disparities. More research is needed to ensure that the screening tools perform similarly at the item and test level, and to determine if DIF is the result of true group differences or the result of the test's construction.

Limitations

While the current study was based on a large and diverse sample, some limitations should be noted. First, given the large sample size, it is important not to rely on p-values as the sole indicator of a meaningful outcome. For instance, while we did find a difference in the Cronbach's alpha between African American/Black children and White children, both levels were quite high. This statistical difference may not be meaningful at a population level and should, therefore, be interpreted in that context. Moreover, this limitation applies to all item level models as well. Specifically, DIF and other psychometric analyses are able to detect smaller and smaller differences as sample sizes increase. Given the statistical power of SEED, it will be important to further study the practical impact of the differential functioning found here to better understand to what extent these differences lead to downstream health disparities. Similarly, while this study cannot determine the reasons for these differences, it will be important to determine the reasons why these items do not perform consistently.

The use of the SCQ in children under the age of 5 is also a potential concern. Prior research such as (Wei et al., 2015) have highlighted how the SCQ Current form shows certain measurement issues (e.g., lower item discrimination and internal consistencies, weaker factor structure, subscale-level bias) when used in this population. The SEED sample used as part of these analyses did include children < 5 years old, however the SEED study used a cutoff of 11, rather than the traditional 15, in line with previous research that found that the 11 cutoff maximizes sensitivity and specificity for children ≤ 4 years old (Wiggins, et al, 2007). It is possible that the DIF results found here were due to the combination of children both 4–5 and > 5 years old. Future work should be conducted to more closely evaluate the impact of child age ranges on the functioning of the SCQ. The SCQ was also administered by phone as part of the SEED study. It should be noted that the SCQ was not intended to be administered by phone and that the results of this measure could have been impacted by this form of implementation.

The results of this study could also be influenced by unique factors associated with participants in this sample. While the SEED sample is large and drawn from several different states, the current study only compared respondents of African American/Black and White children. Additional research will be needed to determine if similar differences exist for other racial and ethnic groups, as well as how generalizable these findings are across cultural contexts. Also the ASD/DD samples within the SEED study are known to be more diverse in terms of race and income than the POP group (Diguiseppi et al., 2016). This may impact the comparisons of these groups in the present analysis, and additional work should be conducted to replicate this finding more generally.

Similarly, there are known to be several associations between diagnosis of ASD and demographic factors such as the family's socioeconomic status, degree of parental concern, race and ethnicity, but with many conflicting findings (Daniels & Mandell, 2014). It is also likely that many of these factors interact such as language, ethnicity, income and education (Reyes et al., 2021). Indeed, there were high correlations between race and many of these factors in the SEED study, and therefore, we cannot rule out that the differences discovered here are due to maternal education or income. Additionally, the intersections between race and other cultural factors like ethnicity and language may have a different impact than income and education (and vice-versa). However, this suggests that the performance of the SCQ, as well as other screening is more complex than previously thought, and additional research is needed to better understand the reasons for these disparities, and how to overcome them.

Conclusion

We found that 41% of the SCQ items perform differently across African American/Black respondents and White respondents. This suggests that results of the SCQ may not be interpreted similarly for African American/Black and White children. Due to 3 items being less likely to be endorsed by African American/Black respondents (uniform DIF) and 13 items varying in terms of the likelihood of endorsement between the two groups (non-uniform DIF), the SCQ could be both over- and underrepresenting potential ASD symptomology, and this could lead to further disparities in diagnosis, access to services, and later outcomes. More work is needed to understand the reasons for these differences and to develop ASD screening tools that are psychometrically sound across all racial and demographic groups. Until that time, referral of a child for evaluation should not simply rely on the SCQ but rather, a combination of appropriate tools and sound clinical judgement.

Acknowledgments

This journal article was supported by several Grant/Cooperative Agreements (**See Funding**) from the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention. We thank the SEED Data Coordinating Center team at the Clinical and Translational Sciences Institute of Michigan State University for their support throughout this study.

Funding

This research is supported by the Centers for Disease Control and Prevention, Centers for Autism and Developmental Disabilities Research, Study to Explore Early Development through six cooperative agreements: Cooperative Agreement Number U10DD000180, Colorado Department of Public Health/University of Colorado School of Medicine; Cooperative Agreement Number U10DD000181, Kaiser Foundation Research Institute (CA); Cooperative Agreement Number U10DD000182, University of Pennsylvania; Cooperative Agreement Number U10DD000183, Johns Hopkins University; Cooperative Agreement Number U10DD000184, University of North Carolina at Chapel Hill; and Cooperative Agreement Number U10DD000498, Michigan State University. Additional support came in part from core grants awarded to the Wyoming Institute for Disabilities, University of Wyoming, the University Center for Excellence in Developmental Disabilities at the University of Wyoming, College of Health Sciences from the U.S. Department of Health and Human Services, through the Administration on Developmental Disabilities Grant #90DDUC0011. The findings and conclusions in this paper are those of the authors and do not necessarily reflect the views of the Centers for Disease Control and Prevention. This study was also supported by six cooperative agreements from the Centers for Disease Control and Prevention (CDC): Cooperative Agreement Number U01000750, University of Colorado Denver; Cooperative Agreement Number U01000748, Kaiser Foundation Research Institute (CA); Cooperative Agreement Number U01000752, University of Pennsylvania; Cooperative Agreement Number U01000746, Johns Hopkins University; Cooperative Agreement Number U01000749, University of North Carolina at Chapel Hill; and Cooperative Agreement Number U10DD000901, Michigan State University.

References

Albores-Gallo L, Roldán-Ceballos O, Villarreal-Valdes G, Betan-zos-Cruz BX, Santos-Sánchez C, Martínez-Jaime MM, & Hilton CL (2012). M-CHAT Mexican version validity and reliability and some cultural considerations. International Scholarly Research Notices. 10.5402/2012/408694

Alpers LM (2018). Distrust and patients in intercultural healthcare: A qualitative interview study. Nursing Ethics, 25(3), 313–323. 10.1177/0969733016652449 [PubMed: 27284017]

Armstrong K, Putt M, Halbert CH, Grande D, Schwartz JS, Liao K, Marcus N, Demeter MB, & Shea JA (2013). Prior experiences of racial discrimination and racial differences in health care system distrust. Medical Care, 51(2), 144–150. 10.1097/mlr.0b013e31827310a1 [PubMed: 23222499]

Barger B, Rice C, & Roach A (2018). Commentary: Response to Foreman's commentary on detecting unmet mental health needs in preschool children (2018). Child and Adolescent Mental Health, 23(3), 217–219. [PubMed: 32677300]

Barger B, Rice C, Simmons CA, & Wolf R (2018). A systematic review of part C early identification studies. Topics in Early Childhood Special Education, 38(1), 4–16. 10.1177/0271121416678664 [PubMed: 28479651]

Barger B, Graybill E, Roach A, & Lane K (2022). Differential item and test functioning of the SRSS-IE12 across race, ethnicity, gender, and elementary level. Assessment for Effective Intervention. 10.1177/1534508420976830

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.

Berument R, Lord P, & Bailey A (1999). Autism screening questionnaire: Diagnostic validity. British Journal of Psychiatry, 175, 444–451.

Bradley CB, Browne EN, Alexander AA, Collins J, Dahm JL, DiGuiseppi CG, & Daniels JL (2018). Demographic and operational factors predicting study completion in a multi-site case-control study of preschool children. American Journal of Epidemiology, 187(3), 592–603. 10.1093/aje/kwx262 [PubMed: 29506195]

Choi YJ, Han H, Dawson KJ, Thoma SJ, & Glenn AL (2019). Measuring moral reasoning using moral dilemmas: evaluating reliability, validity, and differential item functioning of the behavioural defining issues test (bDIT). European Journal of Developmental Psychology, 16(5), 622–631.

Cicero DC, Martin EA, & Krieg A (2019). Differential item functioning of the full and brief wisconsin schizotypy scales in Asian, white, hispanic, and multiethnic samples and between sexes. Assessment, 26(6), 1001–1013. 10.1177/1073191117719509 [PubMed: 28705020]

Daniels AM, & Mandell DS (2014). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. Autism, 18(5), 583–597. 10.1177/1362361313480277 [PubMed: 23787411]

DiGuiseppi CG, Daniels JL, Fallin DM, Rosenberg SA, Schieve LA, Thomas KC, & Schendel DE (2016). Demographic profile of families and children in the study to explore early development (SEED): Case-control study of autism spectrum disorder. Disability and Health Journal, 9(3), 544–551. 10.1016/j.dhjo.2016.01.005 [PubMed: 26917104]

Durkin MS, Maenner MJ, Baio J, Christensen D, Daniels J, Fitzgerald R, & Yeargin-Allsopp M (2017). Autism Spectrum disorder among US children (2002–2010): Socioeconomic, racial, and ethnic disparities. American Journal of Public Health, 107(11), 1818–1826. 10.2105/ajph.2017.304032 [PubMed: 28933930]

Feldt LS (1969). A test of the hypothesis that cronbach's alpha or kuder-richardson coefficent twenty is the same for two tests. Psychometrika, 34(3), 363–373. 10.1007/bf02289364

Fernell E, Eriksson MA, & Gillberg C (2013). Early diagnosis of autism and impact on prognosis: a narrative review. Clinical Epidemiology, 5, 33–43. [PubMed: 23459124]

Filipek A, Ashwal B, Cook D, Gordon G, Johnson K, Levy M, Ozonoff P, Rapin R, Stone T, & Tuchman, & Volkmar. (2000). Practice parameter: Screening and diagnosis of autism. Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society, 55(4), 468–479. 10.1212/wnl.55.4.468

Goodrich JM, Lonigan CJ, & Alfonso SV (2019). Measurement of early literacy skills among monolingual English-speaking and Spanish-speaking language-minority children: A differential item functioning analysis. Early Childhood Research Quarterly, 47, 99–110. 10.1016/j.ecresq.2018.10.007

Gotham K, Risi S, Pickles A, & Lord C (2007). The autism diagnostic observation schedule: revised algorithms for improved diagnostic validity. Journal of Autism and Developmental Disorders, 37(4), 613–627. 10.1007/s10803-006-0280-1 [PubMed: 17180459]

Guthrie W, Wallis K, Bennett A, Brooks E, Dudley J, Gerdes M, Pandey J, Levy SE, Schultz RT, & Miller JS (2019). Accuracy of autism screening in a large pediatric network. Pediatrics, 144(4), e20183963. 10.1542/peds.2018-3963 [PubMed: 31562252]

Herlihy. (2014). Racial/ethnic and socioeconomic differences in screening toddlers for autism spectrum disorders using the M-CHAT [Dissertation, Univesity of Connecticut]. https://opencommons.uconn.edu/dissertations/454

Hope D, Adamson K, McManus IC, Chis L, & Elder A (2018). Using differential item functioning to evaluate potential bias in a high stakes postgraduate knowledge based assessment. BMC Medical Education, 18(1), 1–7. 10.1186/s12909-018-1143-0 [PubMed: 29291730]

Imm P, White T, & Durkin MS (2019). Assessment of racial and ethnic bias in autism spectrum disorder prevalence estimates from a US surveillance system. Autism, 23(8), 1927–1935. 10.1177/1362361319827510 [PubMed: 30892923]

Kelly B, Williams S, Collins S, Mushtaq F, Mon-Williams M, Wright B, & Wright J (2019). The association between socioeconomic status and autism diagnosis in the United Kingdom for children aged 5–8 years of age: Findings from the Born in Bradford cohort. Autism, 23(1), 131–140. 10.1177/1362361317733182 [PubMed: 29113453]

Kennedy BR, Mathis CC, & Woods AK (2007). African Americans and their distrust of the health care system: Healthcare for diverse populations. Journal of Cultural Diversity, 14(2), 56–60. [PubMed: 19175244]

Lin CY, Griffiths MD, & Pakpour AH (2018). Psychometric evaluation of Persian Nomophobia Questionnaire: Differential item functioning and measurement invariance across gender. Journal of Behavioral Addictions, 7(1), 100–108. 10.1556/2006.7.2018.11 [PubMed: 29444607]

Liptak GS, Benzoni LB, Mruzek DW, Nolan KW, Thing-voll MA, Wade CM, & Fryer GE (2008). Disparities in diagnosis and access to health services for children with autism: data from the National Survey of Children's Health. Journal of Developmental Behavioral Pediatrics, 29(3), 152–160. 10.1097/DBP.0b013e318165c7a0 [PubMed: 18349708]

Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, & Rutter M (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism & Developmental Disorders, 30(3), 205. [PubMed: 11055457]

Maenner S, Baio W, DiRienzo Patrick, Christensen W, Pettygrove A, Lopez H, Baroud S, White R, Lee H, Huston H, Esler H-L, Poynter H-M, Constantino F, Zahorodny S, Daniels W, Vehorn S, & Dietz PM (2020). Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring Network. Morbidity and Mortalith Weekly Report, 69(4), 1–12.

Magaña S, Parish SL, Rose RA, Timberlake M, & Swaine JG (2012). Racial and ethnic disparities in quality of health care among children with autism and other developmental disabilities. Intellectual and Developmental Disabilities, 50(4), 287–299. 10.1352/1934-9556-50.4.287 [PubMed: 22861130]

Magaña S, Lopez K, Aguinaga A, & Morton H (2013). Access to diagnosis and treatment services among latino children with autism spectrum disorders. Intellectual and Developmental Disabilities, 51(3), 141–153. [PubMed: 23834211]

Mandell DS, Listerud J, Levy SE, & Pinto-Martin JA (2002). Race differences in the age at diagnosis among medicaid-eligible children with autism. Journal of the American Academy of Child & Adolescent Psychiatry, 41(12), 1447–1453. 10.1097/00004583-200212000-00016 [PubMed: 12447031]

Mandell DS, Wiggins LD, Carpenter LA, Daniels J, DiGuiseppi C, Durkin MS, & Kirby RS (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. American Journal of Public Health, 99(3), 493–498. 10.2105/ajph.2007.131243 [PubMed: 19106426]

Moody EJ, Reyes N, Ledbetter C, Wiggins L, DiGuiseppi C, Alexander A, & Rosenberg SA (2017). Screening for Autism with the SRS and SCQ: Variations across Demographic, Developmental and Behavioral Factors in Preschool Children. Journal of Autism and Developmental Disorders, 47(11), 3550–3561. 10.1007/s10803-017-3255-5 [PubMed: 28856480]

Murphy S, Elklit A, Chen YY, Ghazali SR, & Shevlin M (2019). Sex differences in PTSD symptoms: A differential item functioning approach. Psychological Trauma: Theory, Research, Practice, and Policy, 11(3), 319. [PubMed: 29723027]

Peipert JD, Bentler P, Klicko K, & Hays RD (2018). Negligible impact of differential item functioning between Black and White dialysis patients on the Kidney Disease Quality of Life 36-item short form survey (KDQOLTM-36). Quality of Life Research, 27(10), 2699–2707. 10.1007/s11136-018-1879-3 [PubMed: 29761347]

Reyes NM, Moody E, Hightshoe K, Davidon S, Rosenberg S, Dahl E, & Kubicek LF (2021). Factors that may influence social communication questionnaire (SCQ) scores: An examination of the Spanish and English SCQ in school-aged students. School Psychology Review. 10.1080/2372966x.2021.1927830

Rice SM, Parker AG, Mawren D, Clifton P, Harcourt P, Lloyd M, & Purcell R (2020). Preliminary psychometric validation of a brief screening tool for athlete mental health among male elite athletes: The athlete psychological strain questionnaire. International Journal of Sport and Exercise Psychology, 18(6), 850–865. 10.1080/1612197x.2019.1611900

Rosenberg SA, Moody EJ, Lee LC, DiGuiseppi C, Windham GC, Wiggins LD, & Fallin MD (2018). Influence of family demographic factors on social communication questionnaire scores. Autism Research, 11(5), 695–706. 10.1002/aur.1935 [PubMed: 29424959]

Rutter M, Bailey A, & Lord C (2003). The social communication questionnaire: Manual. Western Psychological Services.

Scarpa A, Reyes NM, Patriquin MA, Lorenzi J, Hassenfeldt TA, Desai VJ, & Kerkering KW (2013). The modified checklist for autism in toddlers: Reliability in a diverse rural American sample. Journal of Autism and Developmental Disorders, 43(10), 2269–2279. 10.1007/s10803-013-1779-x [PubMed: 23386118]

Schendel DiGuiseppi, Croen F, Reed S, Wiggins D, Grether L, Miller N, Pinto-Martin R, Windham A, Aylsworth B, Bonner B, Bradley C, Ferretti F, Giarelli H, Hepburn H, Kaparich L, Lee L, Meyerer R, Ratchford R, Rosenberg R, … Young, & Yeargin-Allsopp. (2012). The study to explore early development (SEED): A multisite epidemiologic study of autism by the centers for autism and developmental disabilities research and epidemiology (CADDRE) network. Journal of Autism and Developmental Disorders, 42(10), 2121–2140. 10.1007/s10803-012-1461-8 [PubMed: 22350336]

Scott F, Aaronson B, De. Graeff, Groenvold G, Koller P, & Sprangers MA (2010). Differential item functioning (DIF) analyses of health-related quality of life instruments using logistic regression. Health and Quality of Life Outcomes. 10.1186/1477-7525-8-81

Shaw KA, Maenner MJ, Bakian AV, Bilder DA, Durkin MS, Furnier SM, & Cogswell ME (2021). Early identification of autism spectrum disorder among children aged 4 years—autism and developmental disabilities monitoring network. MMWR Surveillance Summaries, 70(10), 1.

Smith KA, Gehricke JG, Iadarola S, Wolfe A, & Kuhlthau KA (2020). Disparities in service use among children with autism: A systematic review. Pediatrics, 145(1), S35–S46. 10.1542/peds.2019-1895g [PubMed: 32238530]

Snow AV, & Lecavalier L (2008). Sensitivity and specificity of the Modified Checklist for Autism in Toddlers and the Social Communication Questionnaire in preschoolers suspected of having pervasive developmental disorders. Autism, 12(6), 627–644. 10.1177/1362361308097116 [PubMed: 19005032]

Soto S, Linas K, Jacobstein D, Biel M, Migdal T, & Anthony BJ (2015). A review of cultural adaptations of screening tools for autism spectrum disorders. Autism, 19(6), 646–661. 10.1177/1362361314541012 [PubMed: 25008216]

Swaminathan H, & Rogers HJ (1990). Detecting differential item functioning using logistic regression procedures. Journal of Educational Measurement, 27(4), 361–370.

Wei T, Chesnut SR, Barnard-Brak L, & Richman D (2015). Psychometric analysis of the social communication questionnaire using an item-response theory framework: Implications for the use of the lifetime and current forms. Journal of Psychopathology and Behavioral Assessment, 37(3), 469–480. 10.1007/s10862-014-9468-4

Wiggins LD, Bakeman R, Adamson LB, & Robins DL (2007). The utility of the Social Communication Questionnaire in screening for autism in children referred for early intervention. Focus on Autism and Other Developmental Disabilities, 22(1), 33–38.

Wiggins LD, Reynolds A, Rice CE, Moody EJ, Bernal P, Blaskey L, & Levy SE (2015). Using standardized diagnostic instruments to classify children with autism in the study to explore early development. Journal of Autism and Developmental Disorders, 45(5), 1271–1280. 10.1007/s10803-014-2287-3 [PubMed: 25348175]

Wiggins LD, Nadler C, Rosenberg S, Moody E, Reyes N, Reynolds A, & Levy SE (2021). Many young children with autism who use psychotropic medication do not receive behavior therapy: A Multisite case-control study. The Journal of Pediatrics, 232, 264–271. 10.1016/j.jpeds.2021.01.043 [PubMed: 33493493]

Windham GC, Smith KS, Rosen N, Anderson MC, Grether JK, Coolman RB, & Harris S (2014). Autism and developmental screening in a public, primary care setting primarily serving Hispanics: challenges and results. Journal of Autism and Developmental Disorders, 44(7), 1621–1632. 10.1007/s10803-014-2032-y [PubMed: 24408091]

Table 1

Study to explore early development participant demographic characteristics

Characteristic	**Full sample**		**African American/Black**		**White**	
	n	**%**	**n**	**%**	**n**	**%**
Full study sample	3376	100	879	26.1	2494	73.9
SEED study phase						
SEED 1, 2007–2011	2003	59.3	432	49.1	1571	62.9
SEED 2, 2012–2016	1373	40.7	448	50.9	925	37.1
SEED site CA	478	14.2	50	5.7	428	17.1
CO	414	12.3	11	1.3	403	16.1
GA	906	26.8	437	49.7	469	18.8
NC	502	14.9	114	13.0	388	15.5
PA	536	15.9	99	11.3	437	17.5
MA	540	16.0	169	19.2	371	14.9
ASD diagnosis						
ASD	914	27.1	277	31.5	637	25.5
Non-ASD	2462	72.9	603	68.5	1859	74.5
Maternal education						
High school or less	615	18.2	314	35.7	301	12.1
Some college	839	24.9	322	36.6	517	20.7
Bachelor’s degree	1119	33.2	157	17.9	962	38.6
Post graduate degree	800	23.7	86	9.8	714	28.6
Maternal annual income						
< $10,000	265	8.1	190	22.3	75	3.1
$10,000–$30,000	437	13.4	265	31.1	172	7.1
$30,000–$50,000	401	12.3	153	18.0	248	10.3
$50,000–$70,000	423	13.0	88	10.3	335	13.9
$70,000–$90,000	423	13.0	54	6.3	369	15.3
$90,000–$110,000	381	11.7	42	4.9	339	14.1
> $110,000	930	28.5	60	7.0	870	36.1
Sex						
Male	1140	33.8	284	32.3	856	34.4
Female	2139	63.5	558	65.2	1580	62.9

Table 2

SCQ Non-uniform DIF for White vs. African American/Black in the study to explore early development

SCQ Item	**β**	**OR**	**Sig (*p*)**	**Benjamini-Hochberg critical value**	**Lower 95% CI**	**Upper 95% CI**
2. Do you have a to and fro “conversation” with her/him that involves taking turns or building on what you have said?	0.02	1.02	.420	0.03	0.973	1.068
*3. Does she/he ever use odd phrases or say the same thing over and over in almost exactly the same way (either phrases that she/he hears other people use or ones that she/he makes up)?	0.10	1.10	< .001	< .001	1.050	1.155
4. Does she/he ever use socially inappropriate questions or statements? For example, does she/he ever regularly ask personal questions or make personal comments at awkward times?	0.01	1.01	.413	0.03	0.983	1.041
*5. Does she/he ever get her/his pronouns mixed up (e.g. saying you or she/he for I)?	0.05	1.05	.003	0.01	1.015	1.079
6. Does she/he ever use words that she/he seems to have invented or made up her/himself; put things in odd, indirect ways; or use metaphorical ways of saying things (e.g., saying hot rain for steam)?	0.01	1.01	.573	0.04	0.980	1.038
7. Does she/he ever say the same thing over and over in exactly the same way or insist that you say the same thing over and over again?	0.05	1.05	.023	0.02	1.007	1.099
*8. Does she/he ever have things that she/he seems to have to do in a very particular way or order or rituals that she he/insists that you go through?	0.04	1.04	.015	0.02	1.007	1.068
*9. Does her/his facial expression usually seem appropriate to the particular situation, as far as you can tell?	0.05	1.05	.006	0.01	1.014	1.089
10. Does she/he ever use your hand like a tool or as if it were a part of her/his own body (e.g., pointing with your finger or putting your hand on a doorknob to get you to open the door)?	0.02	1.02	.276	0.03	0.986	1.050
*11. Does she/he ever have any interests that preoccupy her/him and might seem odd to other people (e.g., traffic lights, drainpipes, or timetables)?	0.05	1.05	.004	0.01	1.017	1.093
12. Does she/he ever seem to be more interested in parts of a toy or an object (e.g., spinning the wheels of a car), rather than in using the object as it was intended?	−0.02	0.98	.311	0.03	0.947	1.018
13. Does she/he ever have any special interests that are unusual in their intensity but otherwise appropriate for her/his age and peer group (e.g., trains or dinosaurs)?	0.01	1.01	.442	0.04	0.982	1.043
14. Does she/he ever seem to be unusually interested in the sight, feel, sound, taste, or smell of things or people?	0.03	1.03	.046	0.02	1.001	1.065
15. Does she/he ever have any mannerisms or odd ways of moving her/his hands or fingers, such as flapping or moving her/his fingers in front of her/his eyes?	0.001	1.00	.940	0.05	0.966	1.039
16. Does she/he ever have any complicated movements of her/his whole body, such as spinning or repeatedly bouncing up and down?	0.01	1.01	.676	0.04	0.975	1.041
17. Does she/he ever injure her/himself deliberately, such as by biting her/his arm or banging her/his head?	0.02	1.02	.296	0.03	0.985	1.053
18. Does she/he ever have any objects (other than a soft toy or comfort blanket) that she/he has to carry around?	−0.01	1.00	.726	0.04	0.966	1.024
*19. Does she/he have any particular friends or a best friend?	0.04	1.04	.006	0.01	1.011	1.068
20. Does she/he ever talk with you just to be friendly (rather than to get something)?	0.02	1.02	.421	0.03	0.979	1.053

SCQ Item	β	OR	Sig (*p*)	Benjamini-Hochberg critical value	Lower 95% CI	Upper 95% CI
21. Does she/he ever spontaneously copy you (or other people) or what you are doing (such a vacuuming, gardening, or mending things)?	0.01	1.01	.293	0.03	0.989	1.037
*22. Does she/he ever spontaneously point at things around her/him just to show you things (not because she/he wants them)?	0.05	1.05	<.001	<.001	1.024	1.078
*23. Does she/he ever use gestures, other than pointing or pulling your hand, to let you know what she/he wants?	0.04	1.04	<.001	<.001	1.017	1.063
*24. Does she/he nod her/his head to indicate yes?	0.07	1.07	<.001	0.01	1.038	1.103
*25. Does she/he shake her/his head to indicate no?	0.07	1.07	<.001	0.01	1.036	1.098
26. Does she/he usually look at you directly in the face when doing things with you or talking with you?	0.002	1.00	.926	0.05	0.969	1.035
27. Does she/he smile back if someone smiles at her/him?	0.03	1.03	.150	0.03	0.989	1.073
28. Does she/he ever show you things that interest her/him to engage your attention?	0.001	1.00	.965	0.05	0.947	1.059
29. Does she/he ever offer to share things other than food with you?	0.01	1.01	.646	0.04	0.971	1.049
30. Does she/he ever seem to want you to join in her/his enjoyment of something?	0.01	1.01	.792	0.04	0.958	1.058
31. Does she/he ever try to comfort you if you are sad or hurt?	−0.01	0.99	.758	0.04	.954	1.035
*32. If she he wants something or wants help, does she/he look at you and use gestures when sounds or words to get your attention?	0.05	1.05	.011	0.02	1.011	1.085
33. Does she/he show a normal range of facial expressions?	−0.05	0.95	.073	0.02	0.903	1.005
34. Does she/he ever spontaneously join in and try to copy the actions in social games, such as The Mulberry Bush or London Bridge is Falling Down?	0.04	1.04	.023	0.02	1.006	1.080
*35. Does she/he play any pretend or make-believe games?	0.04	1.05	.004	0.01	1.014	1.077
36. Does she/he seem interested in other children of approximately the same age whom she/he does not know?	0.001	1.00	.950	0.05	0.966	1.038
37. Does she/he respond positively when another child approaches her/him?	−0.03	0.97	.083	0.02	0.930	1.004
38. If you come into a room and start talking to her/him without calling her/his name, does she/he usually look up and pay attention to you?	0.03	1.03	.132	0.02	0.992	1.066
*39. Does she/he ever play imaginative games with another child in such a way that you can tell that each child understands what the other is pretending?	0.05	1.05	.004	0.01	1.016	1.090
40. Does she/he play cooperatively in games that need some form of joining in with a group of other children, such as hide-and-seek or ball games?	0.01	1.01	.692	0.04	0.969	1.049

SCQ Social Communication Questionnaire; *DIF* Differential Item Functioning; *OR* Odds Ratio; *Sig (p)* Significance Level; *CI* Confidence Interval

Item 1 of the SCQ not included in table as is not used in calculation of total SCQ Score

Uniform DIF indicates that there is a consistently higher or lower probability for one group to endorse an item in comparison to another group at all levels of the underlying latent trait. Nonuniform DIF indicates that there is an interaction between the likelihood to endorse the item and group membership by total score on the measured trait

[*] Indicate statistically significant based on the BH correction. To determine significance corrected for inflation of alpha, the p-values are compared to the BH critical value. p-values that are less than the BH critical value are considered significant at the .05 level while correcting for the false discovery rate

Table 3

SCQ uniform DIF for White vs. African American/Black in the study to explore early development

SCQ Item	β	OR	Sig (*p*)	Benjamini-Hochberg critical value	Lower 95% CI	Upper 95% CI
2. Do you have a to and fro “conversation” with her/him that involves taking turns or building on what you have said?	−0.13	0.88	.745	0.04	0.409	1.896
#3. Does she/he ever use odd phrases or say the same thing over and over in almost exactly the same way (either phrases that she/he hears other people use or ones that she/he makes up)?	−1.23	0.29	< .001	< .001	0.203	0.424
*4. Does she/he ever use socially inappropriate questions or statements? For example, does she/he ever regularly ask personal questions or make personal comments at awkward times?	−0.65	0.52	< .001	< .001	0.363	0.745
#5. Does she/he ever get her/his pronouns mixed up (e.g. saying you or she/he for I)?	−0.51	0.60	.001	0.02	0.440	0.818
6. Does she/he ever use words that she/he seems to have invented or made up her/himself; put things in odd, indirect ways; or use metaphorical ways of saying things (e.g., saying hot rain for steam)?	−0.01	0.99	.947	0.05	0.716	1.367
*7. Does she/he ever say the same thing over and over in exactly the same way or insist that you say the same thing over and over again?	−0.94	0.39	< .001	< .001	0.257	0.594
8. Does she/he ever have things that she/he seems to have to do in a very particular way or order or rituals that she he/insists that you go through?	−0.30	0.74	.065	0.02	0.541	1.018
#9. Does her/his facial expression usually seem appropriate to the particular situation, as far as you can tell?	−1.50	0.22	< .001	0.01	0.121	0.411
10. Does she/he ever use your hand like a tool or as if it were a part of her/his own body (e.g., pointing with your finger or putting your hand on a doorknob to get you to open the door)?	−0.45	0.64	.041	0.02	0.417	0.983
#11. Does she/he ever have any interests that preoccupy her/him and might seem odd to other people (e.g., traffic lights, drainpipes, or timetables)?	−0.54	0.59	.023	0.02	0.369	0.930
12. Does she/he ever seem to be more interested in parts of a toy or an object (e.g., spinning the wheels of a car), rather than in using the object as it was intended?	0.07	1.07	.753	0.04	0.701	1.635
13. Does she/he ever have any special interests that are unusual in their intensity but otherwise appropriate for her/his age and peer group (e.g., trains or dinosaurs)?	−0.12	0.89	.478	0.03	0.630	1.241
*14. Does she/he ever seem to be unusually interested in the sight, feel, sound, taste, or smell of things or people?	−0.56	0.57	.005	0.02	0.388	0.842
15. Does she/he ever have any mannerisms or odd ways of moving her/his hands or fingers, such as flapping or moving her/his fingers in front of her/his eyes?	0.15	1.16	.585	0.04	0.685	1.954
16. Does she/he ever have any complicated movements of her/his whole body, such as spinning or repeatedly bouncing up and down?	0.07	1.07	.754	0.05	0.700	1.637
17. Does she/he ever injure her/himself deliberately, such as by biting her/his arm or banging her/his head?	−0.12	0.89	.677	0.04	0.516	1.537
18. Does she/he ever have any objects (other than a soft toy or comfort blanket) that she/he has to carry around?	−0.28	0.75	.197	0.03	0.490	1.158
#19. Does she/he have any particular friends or a best friend?	−0.86	0.42	< .001	0.01	0.306	0.583
20. Does she/he ever talk with you just to be friendly (rather than to get something)?	−0.25	0.78	.396	0.03	0.432	1.394

SCQ Item	β	OR	Sig (*p*)	Benjamini-Hochberg critical value	Lower 95% CI	Upper 95% CI
21. Does she/he ever spontaneously copy you (or other people) or what you are doing (such a vacuuming, gardening, or mending things)?	−0.05	0.95	.762	0.05	0.699	1.300
#22. Does she/he ever spontaneously point at things around her/him just to show you things (not because she/he wants them)?	−0.82	0.44	< .001	0.01	0.309	0.630
#23. Does she/he ever use gestures, other than pointing or pulling your hand, to let you know what she/he wants?	−0.77	0.46	< .001	0.01	0.354	0.608
#24. Does she/he nod her/his head to indicate yes?	−0.92	0.40	< .001	0.01	0.254	0.624
#25. Does she/he shake her/his head to indicate no?	−0.91	0.40	< .001	0.01	0.259	0.623
26. Does she/he usually look at you directly in the face when doing things with you or talking with you?	0.04	1.04	.889	0.05	0.630	1.705
27. Does she/he smile back if someone smiles at her/him?	−0.40	0.67	.280	0.03	0.325	1.385
28. Does she/he ever show you things that interest her/him to engage your attention?	−0.23	0.79	.679	0.04	0.261	2.403
29. Does she/he ever offer to share things other than food with you?	−0.61	0.54	.044	0.02	0.298	0.984
30. Does she/he ever seem to want you to join in her/his enjoyment of something?	−0.72	0.49	.136	0.03	0.188	1.256
31. Does she/he ever try to comfort you if you are sad or hurt?	0.59	1.80	.099	0.02	0.896	3.599
#32. If she he wants something or wants help, does she/he look at you and use gestures when sounds or words to get your attention?	−1.17	0.31	< .001	0.01	0.176	0.551
33. Does she/he show a normal range of facial expressions?	0.69	1.99	.176	0.03	0.735	5.413
34. Does she/he ever spontaneously join in and try to copy the actions in social games, such as The Mulberry Bush or London Bridge is Falling Down?	−0.35	0.71	.248	0.03	0.390	1.275
#35. Does she/he play any pretend or make-believe games?	−1.24	0.29	< .001	0.01	0.184	0.460
36. Does she/he seem interested in other children of approximately the same age whom she/he does not know?	−0.12	0.88	.659	0.04	0.509	1.533
37. Does she/he respond positively when another child approaches her/him?	0.31	1.37	.323	0.03	0.736	2.537
38. If you come into a room and start talking to her/him without calling her/his name, does she/he usually look up and pay attention to you?	0.18	1.20	.487	0.04	0.723	1.977
#39. Does she/he ever play imaginative games with another child in such a way that you can tell that each child understands what the other is pretending?	−0.76	0.47	.001	0.02	0.300	0.730
40. Does she/he play cooperatively in games that need some form of joining in with a group of other children, such as hide-and-seek or ball games?	−0.10	0.91	.746	0.04	0.496	1.654

SCQ Social Communication Questionnaire; *DIF* Differential Item Functioning; *OR* Odds Ratio; *Sig (p)* Significance Level; *CI* Confidence Interval

Item 1 of the SCQ not included in table as is not used in calculation of total SCQ Score

Uniform DIF indicates that there is a consistently higher or lower probability for one group to endorse an item in comparison to another group at all levels of the underlying latent trait. Nonuniform DIF indicates that there is an interaction between the likelihood to endorse the item and group membership by total score on the measured trait

[*] Indicate statistically significant based on the BH correction

[#] Indicate presence of Uniform and Non-Uniform DIF simultaneously. To determine significance corrected for inflation of alpha, the p-values are compared to the BH critical value. p-values that are less than the BH critical value are considered significant at the .05 level while correcting for the false discovery rate