Original Research Article

Assessment
1–13

DOI: 10.1177/10731911241306360
journals.sagepub.com/home/asm
Sage

Measurement Invariance of the First Years Inventory (FYIv3.1) Across Age and Sex for Early Detection of Autism in a Community Sample of Infants

Yun-Ju Chen[1,2,3], John Sideris[2,3], Linda R. Watson[2,3], Elizabeth R. Crais[2,3], and Grace T. Baranek[2,3]

Abstract

The use of parent-report screeners for early detection of autism is time- and cost-efficient in clinical settings but their utility may vary by respondent characteristics. This study aimed to examine the degree to which infants' age and sex impacted parental reports of early behavioral signs of autism captured by the First Years Inventory Version 3.1 (FYIv3.1). The current sample included 6,454 caregivers of infants aged 6 to 16 months recruited through the North Carolina vital records. Using moderated nonlinear factor analysis for each of the seven FYIv3.1, we identified differential item functioning in small to medium effect sizes across 18 out of 69 items, with the majority of biases associated with infants' age (e.g., object mouthing, walking, pretend, and imitation), while sex-related biases were minimal. This indicates that differential scoring algorithms by infants' age and more closely spaced monitoring may be needed for these constructs for more accurate identification of autism in infancy.

Keywords

differential item functioning, measurement invariance, early autism screening, infant sensory, motor and social communication development, community sample, behavioral risk markers

Autism is a neurodevelopmental condition that encompasses difficulties in various behavioral domains, such as perceptual processing, sensory reactivity, and motor coordination, in addition to social communication and restrictive/repetitive behavior (Canu et al., 2021; Micai et al., 2020). The manifestations of these behaviors are heterogeneous, moderated by child and family demographics and child clinical characteristics (Kim & Lord, 2013), and tend to vary over time, particularly during the first years of life (Elsabbagh & Johnson, 2007; Ozonoff et al., 2014). Such individual and developmental variability of behavioral signs in early life make it challenging to use them to reliably predict a later diagnosis of autism by 18 months (Towle & Patrick, 2016; Zwaigenbaum et al., 2015). To date, one of the most time- and cost-efficient methods for developmental surveillance in primary care settings is collecting information from parents or caregivers via questionnaires or interviews about their child's development, through which valuable aspects of a child's behavior beyond clinical contexts are captured (Eapen et al., 2014; Glascoe, 2000). However, the utility of parent-report screening tools may be influenced by parents' sensitivity to and interpretations of their child's behavioral signs of autism (Towle & Patrick, 2016), which may hinge on factors such as child characteristics (e.g., sex, chronological, or developmental age) and proxy reporters' backgrounds (e.g., education, child-rearing experiences, and cultural differences; Bennetts et al., 2016; Dubay & Watson, 2019). Proxy-report or respondent biases may occur when caregivers or other respondents rate their child's behavior differently not because of what is intended to be measured (e.g., level of autistic traits, likelihood of autism) but due to other factors (e.g., respondent or child characteristics) at the time of the assessment. To address this, measurement invariance (MI) testing can be used to empirically test whether an instrument assesses an unobserved construct (or latent trait) in the same manner across all respondents (Meredith, 1993). If

[1]Chang Gung University, Taoyuan, Taiwan
[2]University of Southern California, Los Angeles, USA
[3]The University of North Carolina at Chapel Hill, Chapel Hill, USA

Corresponding Author:
Yun-Ju (Claire) Chen, Department of Occupational Therapy and Graduate Institute of Behavioral Sciences, Chang Gung University, Guishan District, Taoyuan City, 333 Taiwan.
Email: chenyunj@cgu.edu.tw

MI is not met, the measure may be biased because it is not functioning equivalently across the subpopulations of interest. The assessment of MI has been suggested as a critical practice prior to observed mean comparisons across groups or measurement occasions (Martinková et al., 2017; Walker, 2011). If there are systematic differences across respondents with different characteristics (i.e., measurement non-invariance) for a screening tool, certain subgroups of the targeted population may be over- or under-identified, thus affecting the screening performance (Gonzalez & Pelham, 2021). Because such respondent bias may impact the accuracy of screening tools, it should be properly addressed before any inferences about group differences and predictive utility can be made.

Among the many factors that may impact parental reports of early signs of autism, a child's age is particularly important for an early screening tool targeting the first years of life, a period in which infants and toddlers develop rapidly and variably. Previous studies that applied parent-report screening tools in a wider age window for early detection of autism have reported varying utility by child's age. A validation study of the Infant-Toddler Checklist (ITC) targeting a community sample aged 6 to 24 months has shown higher false positive rates for infants in the age range of 6 to 8 months (Wetherby et al., 2008). Previous validation studies of the Modified Checklist for Autism in Toddlers (M-CHAT) found that the screening accuracy was lower in the younger (e.g., <24 months) age groups (Guthrie et al., 2019; Pandey et al., 2008; Sturner et al., 2017). The Social Communication Questionnaire (SCQ) was also reported to show particularly lower specificities among children aged below 4 years (Corsello et al., 2007; Oosterling et al., 2010). A previous prospective study has shown that parental concerns were predictive of a later autism diagnosis among infant siblings, but the predictive utility of different types of concerns (or behavioral domains) was found to vary between 6 and 24 months of age (Sacrey et al., 2015). It is important to note, however, that all these findings were based on the observed scores that do not take MI into account. Without the examination of MI, it remains unknown whether the differences in screening accuracy related to age were due to parent-report biases and/or true differences in behavioral manifestation across children's development. Several important behavioral markers of autism, such as playing pretend games, gaze following, and declarative or imperative pointing, have been reported to emerge at later stages of infancy (Carpenter et al., 1998; Crais et al., 2004; Inada et al., 2010). In addition, preverbal symbolic behaviors might be too subtle for parents to notice in infancy (Johnson, 2008). It is possible that parents do not report these behaviors because they have not emerged or could not be stably observed in early infancy (Towle & Patrick, 2016). Thus, there is a need to examine whether parents report these behaviors without being systematically biased by their child's age across the infancy period when the assessment is taken, as such information may have important implications for improving the utility of early screeners (Schjølberg et al., 2022).

Sex is another potential source of reporting bias given the fact that autism varies in its prevalence and behavioral manifestation across males and females (Christensen et al., 2018; Lai et al., 2015). Concerns have been raised regarding the development and validation process of autism measures that predominantly relied on male samples, which may exclude behaviors that are more representative of autistic females (Goldman, 2013; Kreiser & White, 2014; Lai & Szatmari, 2020). Although sex differences in autistic traits have been widely investigated across studies varying by samples and methodology, there are fewer studies examining sex-related MI on measures of autistic traits or symptoms, especially in the prodromal period. Previous studies have reported sex-related measurement bias in some items across measures and formats, including self-reports (Autism-Spectrum Quotient; Murray et al., 2017), parent reports (SCQ Current form; Wei et al., 2015), and clinical observation (Autism Diagnostic Observation Schedule, ADOS; Kalb et al., 2022; Ronkin et al., 2022; Tien et al., 2024) among autistic and non-autistic populations. Interestingly, most of the items with significant sex biases reported in these studies were about restricted and repetitive behaviors (RRBs), while the study that only examined a social communication construct of ADOS-2 toddler module found no evidence of sex difference in item functioning (Ronkin et al., 2022). According to the studies that targeted younger populations but did not directly test MI, parents of male toddlers and preschoolers, regardless of diagnosis or likelihood of autism, tended to report more autism-related concerns (Messinger et al., 2015; Øien et al., 2017; Ramsey et al., 2018). All these findings highlight the need for a systematic examination of sex-related MI to clarify the source of sex differences among very young populations, such as infants and toddlers. Moreover, sex differences in some behavioral domains seemed to emerge at different ages (McDonnell et al., 2021; Ramsey et al., 2018), suggesting potential sources of bias by age and sex interaction that merit further investigations.

To elucidate whether age- and sex-related biases are present when evaluating early behavioral signs of autism during infancy, the current study focused on the YIv 3.1 (Baranek et al., 2013, 2022), a parent-report questionnaire for early detection of autism among general populations of infants aged 6 to 16 months. This version builds upon a previous version that targeted 12-month-olds (FYIv2.0; Baranek et al., 2003), whose utility has been validated in community samples (Reznick et al., 2007; Turner-Brown et al., 2013; Watson et al., 2007) and samples of infants with an elevated likelihood for autism due to familial risk (e.g., Macari et al., 2018; Meera et al., 2021). To capture the early behavioral signs that may emerge by the first year, the FYIv3.1 adapted and expanded the previous version to include a broad range of autism-related behavioral constructs, such as social communication, sensory-regulatory functions, and motor development (Baranek et al., 2022). A recent study demonstrated the structural validity of the current version and its ability to differentiate children with various clinical outcomes by age 3 (Baranek et al., 2022). Leveraging moderated nonlinear factor analysis (MNLFA), we aimed to examine whether the FYIv3.1 items and constructs can be used to estimate the likelihood of a later diagnosis of autism without being biased by demographic factors such as infants' age and sex in a heterogeneous community sample. MNLFA is a novel and more flexible method for evaluating MI and differential item functioning (DIF) that combines the strengths of multi-group and multiple-indicator multiple-cause (MIMIC) confirmatory factor analysis (Bauer, 2017; Schiltz & Magnus, 2021) and allows for the simultaneous inclusion of multiple continuous or categorical covariates (i.e., potential sources of DIF), including interaction terms (e.g., Age $\times$ Sex). It was expected that more items related to social communication would show DIF across ages, given the later emergence of several social communication behaviors as mentioned above. We also expected that FYIv3.1 items related to RRBs (e.g., sensory interests and repetitions) might show more significant sex-related DIF, as reported by previous MI studies using other autism screeners or diagnostic measures at older ages (Kalb et al., 2022; Murray et al., 2017; Tien et al., 2024; Wei et al., 2015). Aside from identifying DIF items, we also evaluated the cumulative DIF impact at the construct level, which may be negligible if the item-level biases in opposite directions cancel out each other (Chalmers et al., 2016). This would provide clinically relevant insights on whether caution should be used when making group comparisons or developing scoring algorithms for autism screening based on construct scores.

Method

Participants and Procedures

We used age-stratified recruitment from a North Carolina community sample of 40,000 families with children ages 6 to 16 months, born between January 1 and December 31, 2013, ascertained through the state birth registry. To reduce the response time burden, the FYIv3.1 was split into two complementary forms (Forms A and B), each consisting of 48 questions with 27 items in common. Each family randomly received either an A or B form in the recruitment packet and was asked to answer multiple-choice questions about the frequency of their child's behaviors. A total of 6,636 caregivers returned the questionnaire, resulting in a response rate of around 17%. Of these respondents, 60% were mothers, 3% were fathers, 2% were multiple respondents, and 35% did not provide this information. Regarding their racial backgrounds, 77% were non-Hispanic White, 11% Black, 3% Asian or American Indian/Hawaiian, and 8% multiracial or other. Children's ages were adjusted for premature babies born below 36 weeks of gestation. After removing duplicates ($N = 23$), incomplete responses (i.e., $<75\%$ of items completed; $N = 142$), and those with corrected chronological ages outside the 6 to 16 month range ($N = 17$), the final sample for analysis included 6,454 infants (see Table 1 for the sample demographics). Those who received an A form or a B form did not differ in demographic characteristics. All procedures were approved by the Institutional Review Board of the University of North Carolina at Chapel Hill (IRB #13-2648).

Measures

The FYIv3.1 is a parent-report questionnaire with a total of 69 items designed to identify infants aged 6 to 16 months who may have an elevated likelihood of autism among the general population. It measures the frequency of children's behaviors with a 5-point Likert-type scale (from 0 = *never* to 4 = *always*), comprising seven constructs derived from a validated factor structure (Baranek et al., 2022): *Communication, Imitation & Play (CIP), Social Attention & Affective Engagement (SAE), Sensory Hyperresponsiveness (HYPER), Sensory Hyporesponsiveness (HYPO), Self-regulation in Daily Routines (SREG), Sensory Interests, Repetitions, & Seeking Behaviors (SIRS)*, and *Motor Coordination & Milestones (MCM)*. Since the items assess both typical and atypical behaviors, the responses of items about typical behavior were reversely coded so that higher scores indicate a higher likelihood of autism. The item-level endorsement rates by sex and age for the current sample are shown in Table S1 (Supplementary Materials).

Data Analyses

All factor analyses, including MNLFA, were performed in Mplus 8.6 (Muthén & Muthén, 2018) with robust

Table 1. Sample Demographics by the Completed Version of FYIv3.1 Form (Total N = 6,454).

Child and Parent Demographics	FYIv3.1-A Form (N = 3,213)	FYIv3.1-B Form (N = 3,241)
Infants' age in months [M (SD)]	12.1 (2.2)	12.0 (2.2)
Infants' sex (male)	1,669 (52.0%)	1,603 (49.5%)
Infants' race		
White	2,505 (78.0%)	2,480 (76.5%)
Black	339 (10.5%)	399 (12.3%)
Asian	75 (2.3%)	78 (2.4%)
American Indian/Hawaiian	25 (.8%)	21 (.7%)
Multi-racial/Other	269 (8.4%)	263 (8.1%)
Parents' education (6% missing)		
Both parents have a college degree (or beyond)	1,300 (40.5%)	1,251 (38.6%)
One of the parents has a college degree (or beyond)	645 (20.0%)	629 (19.4%)
None of the parents has a college degree (or beyond)	1,007 (31.3%)	1,106 (34.1%)

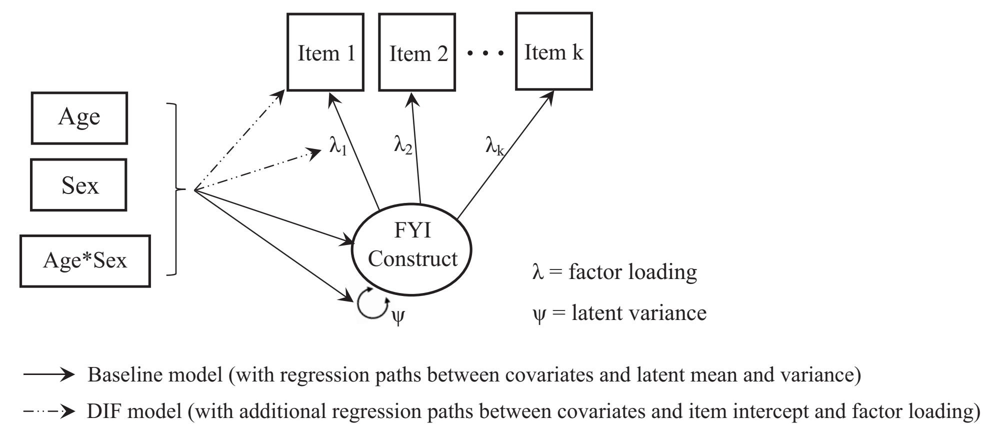

Figure 1. Model Specification of Moderated Nonlinear Factor Analysis (MNLFA).
Baseline model (with regression paths between covariates and latent mean and variance). DIF model (with additional regression paths between covariates and item intercept and factor loading).

maximum likelihood estimation. Single-factor models were fitted to each of the seven FYIv3.1 constructs for testing the assumption of unidimensionality. Comparative fit index (CFI) $\geqslant$ .95 and root mean square error of approximation (RMSEA) < .08 were used to determine the factor model fit (Hu & Bentler, 1999). A set of MNLFA was then performed for each construct with the child's corrected chronological age (continuous: between 6 months 0 days and 16 months 30 days), sex (categorical: 0 = male, 1 = female), and the age $\times$ sex interaction term as covariates. Infants' age was mean-centered to facilitate interpretation of the results. To begin with, a baseline model was specified by allowing all covariates to predict the latent mean and variance (see Figure 1). Next, covariate effects were introduced on the item parameters (i.e., item intercept and factor loading) for one item at a time while treating other items as anchor items (i.e., items without DIF). If the introduction of the covariate effects on the item parameters caused a significant change in model fit compared to the baseline model as indicated by likelihood ratio test (LRT) statistics, the item would be considered as having intercept and/or loading DIF. This process was repeated for each of the items under the same latent factor. Finally, all the items flagged with DIF were estimated simultaneously while any non-significant covariate effect was removed from the model (Bauer, 2017). As the identification of DIF items relies on the chi-square test (Kleinman & Teresi, 2016), which is sensitive to sample size, effect size (Pearson's r for age-related DIF and Cohen's d for sex-related DIF) was calculated for the intercept and loading DIF effects in the final model to determine their salience. Based on effect size benchmarks for behavioral research (Brydges, 2019; Gignac &

Szodorai, 2016), r = .10, .20, and .30 for continuous covariates (i.e., age and Age $\times$ Sex) and d = .15, .40, and .75 for categorical covariates (i.e., sex) are, respectively, considered small, medium, and large in magnitude.

To assess the cumulative DIF impact at the construct level for each individual in our sample, we compared differences between the latent factor scores derived from the baseline model (accounting for impacts on latent mean and variance) and those from the final MNLFA model (accounting for impacts on latent mean and variance as well as DIF). Differences larger than the median standard error of the latent factor estimates were considered as indicating salient cumulative DIF impact (Kleinman & Teresi, 2016). The percentage of participants flagged with salient DIF impact for each construct was examined.

Results

Preliminary Tests of Unidimensionality

Single-factor models revealed adequate fit across the seven constructs (CFI = .901–.973, RMSEA = .015–.045; see Table S2), indicating that the assumption of unidimensionality was met for MNLFA.

Impact on Item Intercepts and Factor Loadings (DIF)

The MNLFA revealed that 48 out of 69 FYIv3.1 items (70%) were flagged with intercept and/or loading DIF given the significant LRT results ($p < .05$) across all constructs. However, the DIF effects in most of the items were negligible ($r < .10$ or $d < .15$). A total of 18 items (9 CIP, 3 SAE, 2 SREG, 2 SIRS, and 2 MCM items) were flagged with intercept/loading DIF in small to medium effect sizes (Table 2; for complete MNLFA results, see Tables S3 to S5). Among these 18 items, 14 items were flagged with salient intercept DIF by age. That is, when holding constant the latent construct scores, the response categories that indicate a higher likelihood of autism were more likely to be endorsed by parents of younger infants for the 10 items with negative intercept DIF across the constructs of CIP, SAE, SREG, SIRS, and MCM (Items 16, 34, 39, 42, 54, 57, 59, 63, 66, and 69; $\beta = -.17$ to $-.04$, SE = .01, all $p < .001$, r =.11 to .27) and by parents of older infants for the three CIP items (Items 14, 43, and 65) with positive intercept DIF (β = .08 to .09, all SE = .01, $p < .001$, r = .12 to .18). There were three CIP and one MCM items (Items 13, 41, 53, and 54) showing salient loading DIF by age ($\beta = -.09$ to $-.06$, SE = .01, $p < .001$, r = .14 to .22), which suggests that they are weaker indicators of the latent factor for older infants. There were only 2 CIP items and 1 SAE item (Items 13, 45, and 58) flagged with salient intercept DIF by sex (β = .11 to .28, SE = .02 to .03, all $p < .001$, d = .16 to .26), which means that the parents of girls were more likely to endorse the response categories that indicate a higher likelihood of autism for these items when holding the construct scores constant. None of the items was flagged with salient loading DIF by sex as well as intercept/loading DIF by age and sex interaction.

Impact on Latent Means and Variances

As shown in Table 3, younger infants had significantly higher latent factor means for the CIP, SAE, SIRS, and MCM constructs ($\beta = -.37$ to $-.04$, SE = .01–.02, all $p < .001$, r = .05–.42) but a lower factor mean for HYPER (β = .06, SE = .01, $p < .001$, r = .07) before accounting for DIF. Larger latent variances were observed for younger infants in the CIP and MCM constructs (i.e., more individual variability around the estimated latent means of CIP and MCM; $\beta = -.10$ to $-.36$, SE = .02–.08, both $p < .001$, r = .06). After accounting for DIF, the age effects on the latent mean of SAE and the latent variance of CIP became non-significant. Regarding sex effects, boys showed higher latent means across all constructs ($\beta = -.23$ to $-.07$, all SE = .03, $p < .05$, d = .0–.19) except for HYPER. Larger latent variances in the CIP, SAE, HYPO, and MCM constructs were also observed for boys ($\beta = -.57$ to $-.12$, SE = .05–.14, all $p < .05$, d = .05–.10). After accounting for DIF, the sex effects on the latent mean of SREG and the latent variance of HYPO became non-significant. No significant effect of age and sex interaction was found for latent means and variances across constructs. The individual DIF-adjusted factor scores by age and sex are visualized in Figure S1.

Cumulative DIF Impact at the Construct Level

To examine the cumulative DIF impact on the individual latent factor estimates, the differences between factor scores unadjusted for DIF (from the baseline model) and adjusted for DIF (from the final MNLFA model) were calculated for each individual and plotted against age and sex (see Figure 2). None of the participants were identified with salient changes in the factor estimates of SAE, HYPER, SREG, and SIRS as the differences were smaller than the median standard error of latent factor estimates across constructs. However, salient changes in factor estimates were found in a small portion of participants: a total of 53 participants (.8%) for CIP, 8 participants (.1%) for HYPO, and 103 participants (1.6%) for MCM.

Table 2. FYIv3.1 Items Flagged With DIF in Small to Medium Effect Sizes.

	Age				Sex			
	Intercept		Loading		Intercept		Loading	
Abbreviated items	β (*SE*)	ES (*r*)	β (*SE*)	ES (*r*)	β (*SE*)	ES (*d*)	β (*SE*)	ES (*d*)
Communication, Imitation, & Play (CIP)								
13[a]. Get attention to play games	.03 (.01)**	.05	**−.06 (.01)*****	**.17**	**.14 (.03)*****	**.16**		
14[a]. Repeat after imitation	**.09 (.01)*****	**.18**			.08 (.03)**	.10		
39[a]. Point to communicate	**−.10 (.01)*****	**.14**	−.03 (.00)***	.09				
41[a]. Social clap	−.07 (.01)***	.10	**−.09 (.01)*****	**.22**	−.09 (.03)**	.10		
43[a]. Get attention by making sounds & looking	**.08 (.01)*****	**.12**			.09 (.03)**	.10		
45[a]. Typical play with toys			−.04 (.01)***	.09	**.28 (.03)*****	**.26**	.05 (.02)*	.06
53[a]. Use gestures			**−.06 (.00)*****	**.18**	−.06 (.02)**	.08	−.05 (.02)*	.06
57[a]. Simple pretend actions	**−.11 (.01)*****	**.14**	−.03 (.01)**	.05	−.13 (.04)***	.09		
65[a]. Copy sounds or noises	**.09 (.01)*****	**.16**						
Social Attention & Affective Engagement (SAE)								
34[a]. Response to sadness	**−.06 (.01)*****	**.11**			.11 (.04)**	.11		
58[a]. Laugh without physical games					**.11 (.02)*****	**.16**		
66[a]. Stop on command	**−.06 (.01)*****	**.14**	−.03 (.01)**	.06				
Self-regulation in Daily Routines (SREG)								
16. Choke or gag	**−.04 (.01)*****	**.11**	.03 (.01)*	.04				
42. Wake up two or more times	**−.06 (.01)*****	**.11**	−.02 (.01)*	.03	−.11 (.03)***	.09		
Sensory Interests, Repetitions, & Seeking Behaviors (SIRS)								
59. Object mouthing	**−.14 (.01)*****	**.27**	.06 (.01)***	.10				
63. Repeatedly flap hands or arms	**−.10 (.01)*****	**.19**	−.04 (.01)***	.07				
Motor Coordination & Milestones (MCM)								
54[a]. Pincer grasp on small objects	**−.07 (.01)*****	**.12**	**−.07 (.01)*****	**.14**	−.10 (.03)***	.13		
69[a]. Walk	**−.17 (.01)*****	**.25**						

Note. Sex was coded as 0 = male, 1 = female. DIF = differential item functioning, ES = effect size. The DIF effects with small to medium effect sizes are bolded. The results for two FYIv3.1 constructs (Sensory Hyper- and Hyporesponsiveness) and Age × Sex DIF effect are omitted here given their negligible effect sizes.
[a]Item responses were reversely coded.
*$p < .05$. **$p < .01$. ***$p < .001$.

Discussion

In this study, we assessed whether parents exhibited differential endorsement of FYIv3.1 items when administered to infants aged 6 to 19 months and of different sexes, while controlling for behavioral trait levels. As a result, 48 out of 69 (70%) of the FYIv3.1 items showed statistically detectible amounts of item-level biases related to infants' age and/or sex given the large sample size, but most of these biases were of negligible effect size. Among the 48 items, there were only 18 items showing DIF in small to medium effect sizes. Two items *"object mouthing"* and *"walking"* showed the largest magnitude of intercept DIF by age, with parents of younger infants more likely to endorse the response categories that indicate a higher likelihood of behaviors associated with autism. The negative intercept DIF by age observed in *"walking"* and other items tapping adaptive behaviors or abilities, such as *"simple pretend actions," "point to communicate," "response to sadness," "stop on command,"* and *"pincer grasp on small objects"* may be due to that these items assess developmental skills typically not mastered at younger ages. When examining the endorsement rates (as shown in Table S1), large discrepancies across infants' ages were found for the above-mentioned six items. Taking the item *"simple pretend actions"* as an example, less than 30% of the parents of 11-month-olds reported that their child always or almost always showed this behavior, while this rate surged to around 80% for infants aged ⩾ 13 months, resonating

Table 3. Age and Sex Effects on Latent Means and Variances Before and After Accounting for DIF.

	Not accounting for DIF				Accounting for DIF			
	Age		Sex		Age		Sex	
	β (*SE*)	ES (*r*)	β (*SE*)	ES (*d*)	β (*SE*)	ES (*r*)	β (*SE*)	ES (*d*)
Communication, Imitation, & Play (CIP)								
Mean	−.37 (.01)***	.42	−.23 (.03)***	.19	−.37 (.01)***	.42	−.26 (.03)***	.22
Variance	−.10 (.02)***	.06	−.12 (.05)*	.06	−.02 (.02)	.01	−.12 (.05)*	.06
Social Attention & Affective Engagement (SAE)								
Mean	−.04 (.01)***	.05	−.14 (.03)***	.12	−.02 (.01)	.02	−.14 (.03)***	.12
Variance	−.04 (.02)	.02	−.20 (.07)**	.07	−.04 (.02)	.02	−.17 (.06)**	.07
Sensory Hyperresponsiveness (HYPER)								
Mean	.06 (.01)***	.07	.04 (.04)	.02	.06 (.01)***	.07	.04 (.04)	.02
Variance	.02 (.02)	.01	−.02 (.08)	.00	.02 (.02)	.01	−.01 (.08)	.00
Sensory Hyporesponsiveness (HYPO)								
Mean	.00 (.01)	.00	−.13 (.03)***	.11	.01 (.02)	.01	−.13 (.03)***	.11
Variance	.03 (.05)	.01	−.29 (.14)*	.05	−.04 (.04)	.01	−.26 (.14)	.05
Self-regulation in Daily Routines (SREG)								
M	−.02 (.01)	.02	−.07 (.03)*	.02	−.00 (.01)	.00	−.03 (.03)	.02
Variance	.00 (.02)	.00	.03 (.06)	.01	.01 (.02)	.01	.03 (.06)	.01
Sensory Interests, Repetitions, & Seeking Behaviors (SIRS)								
Mean	−.07 (.01)***	.09	−.09 (.03)**	.08	−.04 (.01)**	.05	−.13 (.04)***	.08
Variance	.00 (.02)	.00	−.03 (.06)	.01	.02 (.02)	.01	.03 (.06)	.01
Motor Coordination & Milestones (MCM)								
Mean	−.15 (.02)***	.09	−.22 (.03)***	.15	−.20 (.03)***	.08	−.18 (.03)***	.15
Variance	−.36 (.08)***	.06	−.57 (.14)***	.10	−.25 (.03)***	.10	−.43 (.11)***	.10

Note. Sex was coded as 0 = male, 1 = female. DIF = differential item functioning, ES = effect size. Higher latent mean estimates indicate greater difficulties or more atypical behaviors for that construct.
*$p < .05$. **$p < .01$. ***$p < .001$.

with the previously reported pass rates that vary across ages for several social communication and play items on the M-CHAT (Inada et al., 2010). The current analysis demonstrated that these endorsement differences across age remained substantial even when controlling for the level of trait, indicating age biases on the item thresholds. On the other hand, items that assess potentially maladaptive behaviors flagged by negative intercept DIF by age—such as *"object mouthing," "repeatedly flap hands or arms," "choke or gag,"* and *"wake up two or more times"*—are more common in early infancy and are therefore more likely to be endorsed by parents of younger infants. It is noteworthy that motor stereotypies are commonly observed at the early stages of motor development (Thelen, 1979; Wolff et al., 2016) and therefore behaviors such as hand flapping might be difficult for parents of younger infants to differentiate whether they are adaptive or maladaptive.

Interestingly, a few items in the construct of *Communication, Imitation & Play* showed significant intercept DIF by age in the opposite direction: parents of older infants were more likely to report that their children less frequently "*repeat (a sound) after imitation,"*"*get attention by making sounds and looking,"* and *"copy sounds or noises,"* given the same trait level. Since these three items are all related to making sounds, such preverbal behavior is likely replaced by more complex verbal skills as children grow older and thus less frequently endorsed by parents. In addition, four items (*"get attention to play games," "social clap," "use gestures,"* and *"pincer grasp on small objects"*) were flagged with negative loading DIF by age in small to medium effect size, indicating that they contribute less to the construct of *Communication, Imitation & Play* or *Motor Coordination & Milestones* for the older infants. As these items tend to reflect developmental milestones, most children at older ages may have mastered these behaviors earlier in their development and/or acquired more advanced skills to replace them, resulting in lower discrimination performance for older infants in the general population.

As for the sex-related biases, there were only three items about the play (*"typical play with toys," "get attention to play games,"* and *"laugh without physical games"*) showing intercept DIF by sex in small effect size. That is, parents of girls tended to report that their child performed these behaviors less frequently when holding the level of trait constant. Girls were reported to develop more variations or complexities in toy manipulations earlier than boys (Barbu et al., 2011; Cherney et al.,

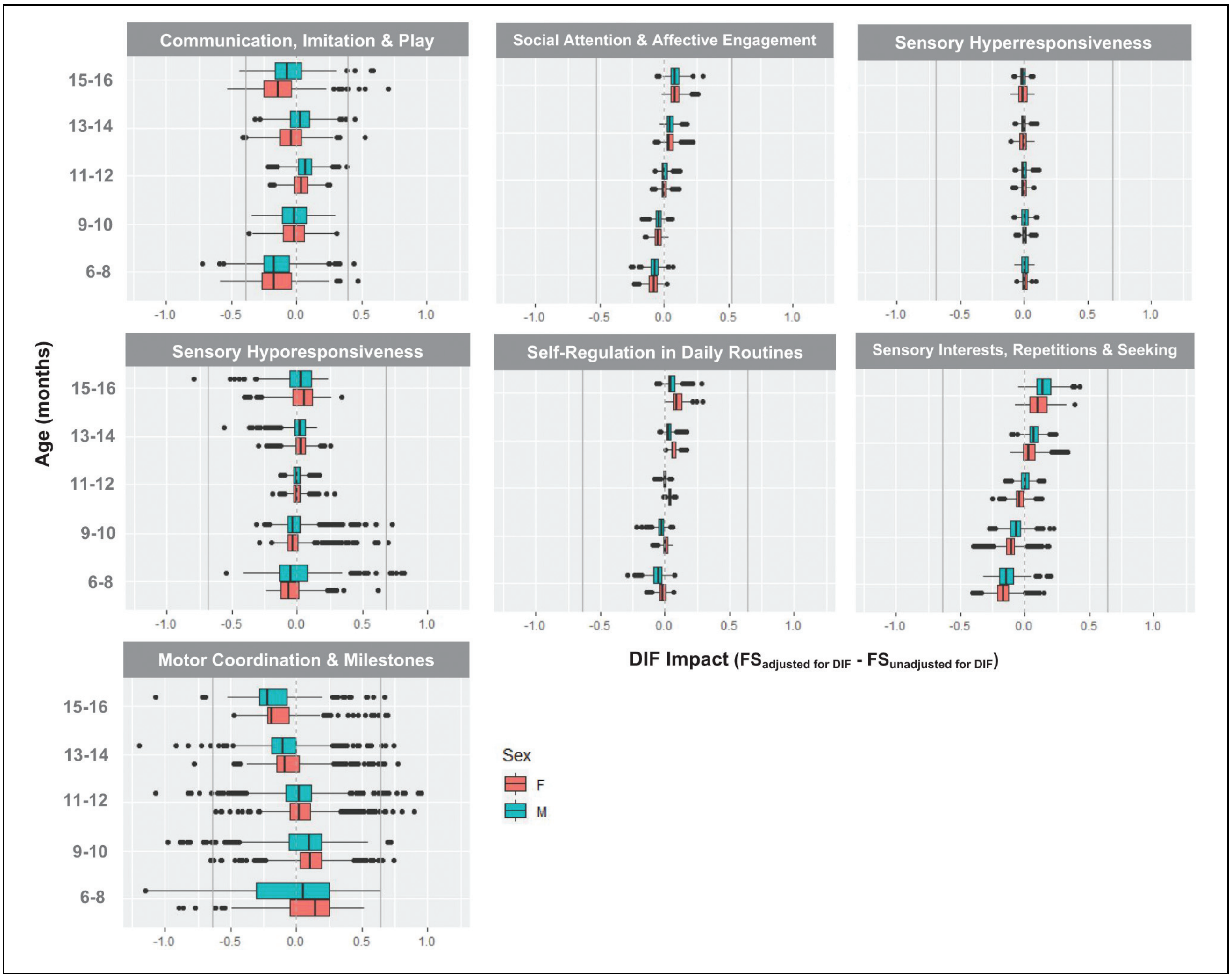

Figure 2 Individual-Level DIF Impact by Infants' Age and Sex Groups

Note. DIF = differential item functioning; FS = factor scores. The x-axes denote the differences between factor scores after versus before adjusting for DIF, which were plotted by age (grouped by 2-month intervals as shown on y-axes) and sex (indicated by colors; males are shown at the top of each age range row). The gray vertical lines denote the ± median standard errors; differences in individual factor scores beyond the range were considered salient.

2003; Goldberg & Lewis, 1969) and thus parents of girls may be prone to report that their child does not play with toys in typical ways. As for the items *"get attention to play games,"* and *"laugh without physical games,"* the examples given for these items, such as peek-a-boo and taking turns to roll a ball, may predominantly reflect physical play activities that are potentially preferred by boys (Pellegrini & Nathan, 2011). As previous evidence has shown sex differences in toy preference among preschool-aged autistic children and play complexity that differed from non-autistic peers (Harrop et al., 2017), further investigations may be merited on how such sex-related bias for early play behaviors observed in the general population would further impact early detection of autism. Overall, while prior studies on older children or adults revealed potential sex-related biases for RRB items (Kalb et al., 2022; Murray et al., 2017; Tien et al., 2024; Wei et al., 2015), our results indicated an absence of such concerns among FYIv3.1 items applied to 6- to 16-month-olds.

When examining the latent means and variances as a function of age and sex, most of the age and sex effects remained significant after accounting for DIF (as shown in Table 3). Particularly, the age difference in latent means remained large for the construct of *Communication, Imitation & Play*, indicating that age-specific cutoffs for this behavioral construct may be needed for a more accurate estimation of the likelihood of autism in infants ages 6 to 16 months. This finding may speak to why varying utility across age bands has been commonly observed in several autism screening tools (e.g., M-CHAT, ITC, and SCQ) that focus predominantly on social communication behaviors (Corsello et al., 2007; Guthrie et al., 2019; Oosterling et al., 2010; Pandey et al., 2008; Sturner et al., 2017; Wetherby et al., 2008). Sex differences in latent means were overall minimal, indicating that sex-specific cutoffs may not be necessary for parent-report screeners administered during infancy. However, relatively large sex differences were observed in the constructs of *Communication, Imitation & Play* and *Motor Coordination and Milestones*, indicating that boys are less likely to show expected behaviors measured by these constructs, consistent with the previous evidence of slower developmental gains in key behavioral domains related to autism in male toddlers among the general population (Øien et al., 2017). An interesting finding is that the latent variance of *Motor Coordination & Milestones* scores was observed to be larger for boys (irrespective of age) as well as infants at younger ages, indicating more heterogeneous manifestations of motor-related behaviors on the FYIv3.1 and potentially more challenges of using them to assess the likelihood of autism for these subgroups. This finding agrees with the prior observation that parents of autistic boys were less consistent in their first concerns about motor delays (Dillon et al., 2021), thus stressing the importance of parsing the individual variability of motor behaviors in infancy to improve the utility of this construct for early identification of autism (Hudry et al., 2020).

Further, we examined the cumulative impact of item-level biases by age and sex on the latent factor estimates for each participant. Clinically salient differences were observed in the construct scores of *Communication, Imitation & Play, Sensory Hyporesponsiveness*, and *Motor Coordination & Milestones* before versus after accounting for DIF in a small number (less than 2%) of our participants. Despite this small portion of our sample with clinically significant DIF, more accurate DIF-adjusted estimates are still recommended to be used for further statistical inferences (Gonzalez & Pelham, 2021; Putnick & Bornstein, 2016). Decisions such as whether to drop the items with salient DIF and whether to use DIF-adjusted scores for further measurement validation need to be weighed against the practical utility of the items and other qualitative factors, such as the clinical significance of behaviors measured by these items and feasibility for applying different scoring algorithms or cutoffs in clinical practices across subpopulations (Jones, 2019; Teresi et al., 2012). As the examination of DIF has been suggested as a routine part of instrument development and validation (Martinková et al., 2017; Walker, 2011), the current study serves as an important initial step toward more accurate early identification of the likelihood of autism in the general population through parent-report screeners such as the FYIv3.1.

Although the large sample size in this study is a strength, findings should be interpreted cautiously considering the following limitations. First, the data collection approach of splitting into two forms resulted in some analytic challenges. Although there is no difference in demographics between participants who received either form, and robust maximum likelihood estimation was applied to address the split-form missingness for factor analyses, the item-level analyses (i.e., DIF testing) for items that are unique to either form might be subject to more uncertainty in estimation. Another caveat is that the current sample may not accurately represent the general population of the United States, as recruitment was limited to one state and 77% of the families identified as White race/non-Hispanic ethnicity, which exceeds the estimated 64% proportion in the state's general population at that time (Tippett, 2014). Also, around 35% of our participants did not indicate who was the respondent in the current study, and thus we did not examine parent education level as a potential source of bias. Future research may be necessary to ensure that the estimation of the likelihood of autism is not biased by parent education level and other factors such as race/ethnicity, sibling status, and child's comorbidity with other conditions. Another important direction for future work is to include long-term outcome data (e.g., whether the child is later diagnosed with autism) and to develop a refined scoring algorithm for predicting a later diagnosis, informed by the current findings. Incorpo-rating statistical learning methods, such as Classification and Regression Trees (CART), may offer the potential for developing finer-tuned algorithms that account for DIF in predictive models (Classe & Kern, 2024; Finch et al., 2016).

Conclusion

The current study aimed to investigate potential respondent biases related to infants' age and sex when applying the FYIv3.1, a parent-report screener for early detection of autism, to a general population aged 6 to 16 months. While the DIF analyses revealed minimal concerns for sex-related biases, relatively large age-related biases were observed in multiple items, particularly those related to sensory seeking or repetitions (e.g., object mouthing), motor development (e.g., walking), play (e.g., imitation, pretend), and preverbal communication (e.g., pointing). This is consistent with our expectation that many of these behaviors emerge or mature at different age points during the infancy period and may thus influence parents' reports of these behaviors. However, the item-level biases in the opposite directions might potentially cancel out at the construct level, as reflected by merely a small portion of our sample with salient differences in latent factor scores before and after accounting for DIF for some constructs. Nevertheless, the age differences in DIF-adjusted scores remained evident for the constructs of *Communication, Imitation & Play* and *Motor Coordination & Milestones*, indicating that age-specific scoring algorithms or cutoffs and more close-spaced monitoring may be needed for these constructs. These findings have important implications for understanding age and sex differences in the manifestation of autism-related behaviors in infancy among the general population, and for improving the utility of the FYIv3.1 as well as other early autism screeners when being applied to a developmentally sensitive period. Applying psychometric strategies, such as MNLFA, to parent-report measures can help address potential respondent biases and enhance the accuracy of estimating the likelihood of autism or the level of autistic traits. This, in turn, can contribute to the development of more effective early identification and screening measures.

Acknowledgments

The authors would like to thank the participated families and the staff who helped with data collection and coding as part of the Program for Early Autism Research, Leadership, and Service (PEARLS) team. We also appreciate the valuable feedback from the Innovations in Neurodevelopmental Sensory Processing Research (insp!re) lab members at the University of Southern California. We are especially grateful to our departed colleague, Dr. J. Steven Reznick from the Department of Psychology and Neuroscience at the UNC-Chapel Hill, for his leadership, friendship, in the early stages of the project.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: GTB is the lead author of the First Years Inventory used in the current study, and L.R.W. and E.R.C. are the co-authors of the First Years Inventory; there are no financial conflicts with the use of this measure as it is freely available. Y-J.C. and J.S. declared no conflict of interest.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was funded by the Autism Speaks Foundation (Grant #5946) and the Ireland Family Foundation.

ORCID iD

Yun-Ju Chen https://orcid.org/0000-0002-0659-2110

Supplemental Material

Supplemental material for this article is available online.

References

Baranek, G. T., Sideris, J., Chen, Y. J., Crais, E., Turner-Brown, L., & Watson, L. R. (2022). Early measurement of autism risk constructs in the general population: A new factor structure of the First Years Inventory (FYIv3.1) for ages 6-16 months. *Autism Research*, *15*(5), 915–928. https://doi.org/10.1002/aur.2691

Baranek, G. T., Watson, L. R., Crais, E., & Reznick, S. (2003). *First Year Inventory (FYI) 2.0*. University of North Carolina at Chapel Hill.

Baranek, G. T., Watson, L. R., Crais, E., Turner-Brown, L., & Reznick, S. (2013). *First Years Inventory (FYI) 3.1*. University of North Carolina at Chapel Hill.

Barbu, S., Cabanes, G., & Le Maner-Idrissi, G. (2011). Boys and girls on the playground: Sex differences in social development are not stable across early childhood. *PLOS ONE*, *6*(1), e16407. https://doi.org/10.1371/journal.pone.0016407

Bauer, D. J. (2017). A more general model for testing measurement invariance and differential item functioning. *Psychological Methods*, *22*(3), 507–526. https://doi.org/10.1037/met0000077

Bennetts, S. K., Mensah, F. K., Westrupp, E. M., Hackworth, N. J., & Reilly, S. (2016). The agreement between parent-reported and directly measured child language and parenting behaviors. *Frontiers in Psychology*, *7*, 1710. https://doi.org/10.3389/fpsyg.2016.01710

Brydges, C. R. (2019). Effect size guidelines, sample size calculations, and statistical power in gerontology. *Innovation in Aging*, *3*(4), igz036. https://doi.org/10.1093/geroni/igz036

Canu, D., Van der Paelt, S., Canal-Bedia, R., Posada, M., Vanvuchelen, M., & Roeyers, H. (2021). Early non-social behavioural indicators of autism spectrum disorder (ASD) in siblings at elevated likelihood for ASD: A systematic review. *European Child & Adolescent Psychiatry*, *30*(4), 497–538. https://doi.org/10.1007/s00787-020-01487-7

Carpenter, M., Nagell, K., Tomasello, M., Butterworth, G., & Moore, C. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monographs of the Society for Research in Child Development*, *63*(4), i-174. https://doi.org/10.2307/1166214

Chalmers, R. P., Counsell, A., & Flora, D. B. (2016). It might not make a big DIF: Improved differential test functioning statistics that account for sampling variability. *Educational and Psychological Measurement*, *76*(1), 114–140. https://doi.org/10.1177/0013164415584576

Cherney, I. D., Kelly-Vance, L., Gill Glover, K., Ruane, A. M. Y., & Oliver Ryalls, B. (2003). The effects of stereotyped toys and gender on play assessment in children aged 18-47 months. *Educational Psychology*, *23*(1), 95–106. https://doi.org/10.1080/01443410303222

Christensen, D. L., Braun, K. V. N., Baio, J., Bilder, D., Charles, J., Constantino, J. N., Daniels, J., Durkin, M. S., Fitzgerald, R. T., Kurzius-Spencer, M., Lee, L.-C., Pettygrove, S., Robinson, C., Schulz, E., Wells, C., Wingate, M.

S., Zahorodny, W., & Yeargin-Allsopp, M. (2018). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2012. *MMWR Surveillance Summaries*, *65*(13), 1–23. https://doi.org/10.15585/mmwr.ss6513a1

Classe, F., & Kern, C. (2024). Detecting differential item functioning in multidimensional graded response models with recursive partitioning. *Applied Psychological Measurement*, *48*(3), 83–103. https://doi.org/10.1177/01466216241238743

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H. Jr., Leventhal, B. L., & Lord, C. (2007). Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *The Journal of Child Psychology and Psychiatry*, *48*(9), 932–940. https://doi.org/10.1111/j.1469-7610.2007.01762.x

Crais, E., Douglas, D., & Campbell, C. (2004). The intersection of gestures and intentionality. *Journal of Speech, Language, and Hearing Research*, *47*(3), 678–694. https://doi.org/10.1044/1092-4388(2004/052)

Dillon, E. F., Kanne, S., Landa, R. J., Annett, R., Bernier, R., & Bradley, C., SPARK Consortium. (2021). Sex differences in autism: Examining intrinsic and extrinsic factors in children and adolescents enrolled in a national ASD cohort. *Journal of Autism and Developmental Disorders*, *53*, 1305–1318. https://doi.org/10.1007/s10803-021-05385-y

DuBay, M., & Watson, L. R. (2019). Translation and cultural adaptation of parent-report developmental assessments: Improving rigor in methodology. *Research in Autism Spectrum Disorders*, *62*, 55–65. https://doi.org/10.1016/j.rasd.2019.02.005

Eapen, V., Črnčec, R., Woolfenden, S., & Blackmore, R. (2014). Screening for Autism Spectrum Disorders using the PEDS and M-CHAT. *Journal of Mental Health Research in Intellectual Disabilities*, *7*(1), 1–13. https://doi.org/10.1080/19315864.2012.704489

Elsabbagh, M., & Johnson, M. H. (2007). Infancy and autism: Progress, prospects, and challenges. *Progress in Brain Research*, *164*, 355–383. https://doi.org/10.1016/S0079-6123(07)64020-5

Finch, W. H., Hernández Finch, M. E., & French, B. F. (2016). Recursive partitioning to identify potential causes of differential item functioning in cross-national data. *International Journal of Testing*, *16*(1), 21–53. https://doi.org/10.1080/15305058.2015.1039644

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Glascoe, F. P. (2000). Early detection of developmental and behavioral problems. *Pediatrics in Review*, *21*(8), 272–280. https://doi.org/10.1542/pir.21.8.272

Goldberg, S., & Lewis, M. (1969). Play behavior in the year-old infant: Early sex differences. *Child Development*, *40*(1), 21–31. https://doi.org/10.2307/1127152

Goldman, S. (2013). Opinion: Sex, gender and the diagnosis of autism—A biosocial view of the male preponderance. *Research in Autism Spectrum Disorders*, *7*(6), 675–679. https://doi.org/10.1016/j.rasd.2013.02.006

Gonzalez, O., & Pelham, W. E III. (2021). When does differential item functioning matter for screening? A method for empirical evaluation. *Assessment*, *28*(2), 446–456. https://doi.org/10.1177/1073191120913618

Guthrie, W., Wallis, K., Bennett, A., Brooks, E., Dudley, J., Gerdes, M., Pandey, J., Levy, S. E., Schultz, R. T., & Miller, J. S. (2019). Accuracy of autism screening in a large pediatric network. *Pediatrics*, *144*(4), e20183963. https://doi.org/10.1542/peds.2018-3963

Harrop, C., Green, J., & Hudry, & PACT Consortium. (2017). Play complexity and toy engagement in preschoolers with autism spectrum disorder: Do girls and boys differ? *Autism*, *21*(1), 37–50. https://doi.org/10.1177/1362361315622410

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Hudry, K., Chetcuti, L., & Hocking, D. R. (2020). Motor functioning in developmental psychopathology: A review of autism as an example context. *Research in Developmental Disabilities*, *105*, 103739. https://doi.org/10.1016/j.ridd.2020.103739

Inada, N., Kamio, Y., & Koyama, T. (2010). Developmental chronology of preverbal social behaviors in infancy using the M-CHAT: Baseline for early detection of atypical social development. *Research in Autism Spectrum Disorders*, *4*(4), 605–611. https://doi.org/10.1016/j.rasd.2009.12.003

Johnson, C. P. (2008). Recognition of autism before age 2 years. *Pediatrics in Review*, *29*(3), 86–96. https://doi.org/10.1542/pir.29.3.86

Jones, R. N. (2019). Differential item functioning and its relevance to epidemiology. *Current Epidemiology Reports*, *6*(2), 174–183. https://doi.org/10.1007/s40471-019-00194-5

Kalb, L. G., Singh, V., Hong, J. S., Holingue, C., Ludwig, N. N., Pfeiffer, D., Reetzke, R., Gross, A. L., & Landa, R. (2022). Analysis of race and sex bias in the Autism Diagnostic Observation Schedule (ADOS-2). *JAMA Network Open*, *5*(4), e229498. https://doi.org/10.1001/jamanetworkopen.2022.9498

Kim, S. H., & Lord, C. (2013). The behavioral manifestations of autism spectrum disorders. In J. D. Buxbaum & P. R. Hof (Eds.), *The neuroscience of autism spectrum disorders* (pp. 25–37). Elsevier. https://doi.org/10.1016/B978-0-12-391924-3.00002-8

Kleinman, M., & Teresi, J. A. (2016). Differential item functioning magnitude and impact measures from item response theory models. *Psychological Test and Assessment Modeling*, *58*(1), 79–98.

Kreiser, N. L., & White, S. W. (2014). ASD in females: Are we overstating the gender difference in diagnosis? *Clinical Child and Family Psychology Review*, *17*(1), 67–84. https://doi.org/10.1007/s10567-013-0148-9

Lai, M. C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/gender differences and autism: Setting the scene for future research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(1), 11–24. https://doi.org/10.1016/j.jaac.2014.10.003

Lai, M. C., & Szatmari, P. (2020). Sex and gender impacts on the behavioural presentation and recognition of autism. *Current Opinion in Psychiatry*, *33*(2), 117–123. https://doi.org/10.1097/YCO.0000000000000575

Macari, S. L., Wu, G. C., Powell, K. K., Fontenelle, S., Macris, D. M., & Chawarska, K. (2018). Do parents and clinicians agree on ratings of autism-related behaviors at 12 months of age? A study of infants at high and low risk for ASD. *Journal of Autism and Developmental Disorders*, *48*(4), 1069–1080. https://doi.org/10.1007/s10803-017-3410-z

Martinková, P., Drabinová, A., Liaw, Y. L., Sanders, E. A., McFarland, J. L., & Price, R. M. (2017). Checking equity: Why differential item functioning analysis should be a routine part of developing conceptual assessments. *CBE—Life Sciences Education*, *16*(2), rm2. https://doi.org/10.1187/cbe.16-10-0307

McDonnell, C. G., DeLucia, E. A., Hayden, E. P., Penner, M., Curcin, K., Anagnostou, E., Nicolson, R., Kelley, E., Georgiades, S., Liu, X., & Stevenson, R. A. (2021). Sex differences in age of diagnosis and first concern among children with autism spectrum disorder. *Journal of Clinical Child & Adolescent Psychology*, *50*(5), 645–655. https://doi.org/10.1080/15374416.2020.1823850

Meera, S. S., Donovan, K., Wolff, J. J., Zwaigenbaum, L., Elison, J. T., Kinh, T. & IBIS Network. (2021). Towards a data-driven approach to screen for autism risk at 12 months of age. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(8), 968–977. https://doi.org/10.1016/j.jaac.2020.10.015

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543. https://doi.org/10.1007/BF02294825

Messinger, D. S., Young, G. S., Webb, S. J., Ozonoff, S., Bryson, S. E., Carter, A., Carver, L., Charman, T., Chawarska, K., Curtin, S., Dobkins, K., Hertz-Picciotto, I., Hutman, T., Iverson, J. M., Landa, R., Nelson, C. A., Stone, W. L., Tager-Flusberg, H., & Zwaigenbaum, L. (2015). Early sex differences are not autism-specific: A Baby Siblings Research Consortium (BSRC) study. *Molecular Autism*, *6*(1), 1–12. https://doi.org/10.1186/s13229-015-0027-y

Micai, M., Fulceri, F., Caruso, A., Guzzetta, A., Gila, L., & Scattoni, M. L. (2020). Early behavioral markers for neurodevelopmental disorders in the first 3 years of life: An overview of systematic reviews. *Neuroscience & Biobehavioral Reviews*, *116*, 183–201. https://doi.org/10.1016/j.neubiorev.2020.06.027

Murray, A. L., Allison, C., Smith, P. L., Baron-Cohen, S., Booth, T., & Auyeung, B. (2017). Investigating diagnostic bias in autism spectrum conditions: An item response theory analysis of sex bias in the AQ-10. *Autism Research*, *10*(5), 790–800. https://doi.org/10.1002/aur.1724

Muthén, L. K., & Muthén, B. (2018). *Mplus. The Comprehensive Modelling Program for Applied Researchers: User's Guide*, 5.

Øien, R. A., Hart, L., Schjølberg, S., Wall, C. A., Kim, E. S., Nordahl-Hansen, A., Eisemann, M. R., Chawarska, K., Volkmar, F. R., & Shic, F. (2017). Parent-endorsed sex differences in toddlers with and without ASD: Utilizing the M-CHAT. *Journal of Autism and Developmental Disorders*, *47*(1), 126–134. https://doi.org/10.1007/s10803-016-2945-8

Oosterling, I., Rommelse, N., De Jonge, M., Van Der Gaag, R. J., Swinkels, S., Roos, S., Visser, J., & Buitelaar, J. (2010). How useful is the Social Communication Questionnaire in toddlers at risk of autism spectrum disorder? *The Journal of Child Psychology and Psychiatry*, *51*(11), 1260–1268. https://doi.org/10.1111/j.1469-7610.2010.02246.x

Ozonoff, S., Young, G. S., Belding, A., Hill, M., Hill, A., Hutman, T., Johnson, S., Miller, M., Rogers, S. J., Schwichtenberg, A. J., Steinfeld, M., & Iosif, A. M. (2014). The broader autism phenotype in infancy: When does it emerge? *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(4), 398–407. https://doi.org/10.1016/j.jaac.2013.12.020

Pandey, J., Verbalis, A., Robins, D. L., Boorstein, H., Klin, A., Babitz, T., Chawarska, K., Volkmar, F., Green, J., Barton, M., & Fein, D. (2008). Screening for autism in older and younger toddlers with the Modified Checklist for Autism in Toddlers. *Autism*, *12*(5), 513–535. https://doi.org/10.1177/1362361308094503

Pellegrini, A. D., & Nathan, P. E. (Eds.). (2011). *The Oxford handbook of the development of play*. Oxford Library of Psychology.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Ramsey, R. K., Nichols, L., Ludwig, N. N., Fein, D., Adamson, L. B., & Robins, D. L. (2018). Brief report: Sex differences in parental concerns for toddlers with autism risk. *Journal of Autism and Developmental Disorders*, *48*(12), 4063–4069. https://doi.org/10.1007/s10803-018-3583-0

Reznick, J. S., Baranek, G. T., Reavis, S., Watson, L. R., & Crais, E. R. (2007). A parent-report instrument for identifying one-year-olds at risk for an eventual diagnosis of autism: The first year inventory. *Journal of Autism and Developmental Disorders*, *37*(9), 1691–1710. https://doi.org/10.1007/s10803-006-0303-y

Ronkin, E., Tully, E. C., Branum-Martin, L., Cohen, L. L., Hall, C., Dilly, L., & Tone, E. B. (2022). Sex differences in social communication behaviors in toddlers with suspected autism spectrum disorder as assessed by the ADOS-2 toddler module. *Autism*, *26*(5), 1282–1295. https://doi.org/10.1177/13623613211047070

Sacrey, L. A. R., Zwaigenbaum, L., Bryson, S., Brian, J., Smith, I. M., Roberts, W., Szatmari, P., Roncadin, C., Garon, N., Novak, C., Vaillancourt, T., McCormick, T., MacKinnon, B., Jilderda, S., & Armstrong, V. (2015). Can parents' concerns predict autism spectrum disorder? A prospective study of high-risk siblings from 6 to 36 months of age. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(6), 470–478. https://doi.org/10.1016/j.jaac.2015.03.014

Schiltz, H. K., & Magnus, B. E. (2021). Differential item functioning based on autism features, IQ, and age on the screen for child anxiety related disorders (scared) among youth on the autism spectrum. *Autism Research*, *14*(6), 1220–1236. https://doi.org/10.1002/aur.2481

Schjølberg, S., Shic, F., Volkmar, F. R., Nordahl-Hansen, A., Stenberg, N., Torske, T., Larsen, K., Riley, K., Sukhodolsky, D. G., Leckman, J. F., Chawarska, K., & Øien, R. A. (2022). What are we optimizing for in autism screening? Examination of algorithmic changes in the M-CHAT. *Autism Research*, *15*(2), 296–304. https://doi.org/10.1002/aur.2643

Sturner, R., Howard, B., Bergmann, P., Morrel, T., Landa, R., Walton, K., & Marks, D. (2017). Accurate autism screening at the 18-month well-child visit requires different strategies than at 24 months. *Journal of Autism and Developmental Disorders*, *47*(10), 3296–3310. https://doi.org/10.1007/s10803-017-3231-0

Teresi, J. A., Ramirez, M., Jones, R. N., Choi, S., & Crane, P. K. (2012). Modifying measures based on differential item functioning (DIF) impact analyses. *Journal of Aging and Health*, *24*(6), 1044–1076. https://doi.org/10.1177/0898264312436877

Thelen, E. (1979). Rhythmical stereotypies in normal human infants. *Animal Behaviour*, *27*, 699–715. https://doi.org/10.1016/0003-3472(79)90006-X

Tien, I. S., Kim, J., Johnson, A. R., & Wood, J. J. (2024). Gender differences: Confirmatory factor analysis of the ADOS-II. *Journal of Autism and Developmental Disorders*. Epub ahead of print. PMID: 38652372. https://doi.org/10.1007/s10803- 024-06346-x

Tippett, R. (2014, June 30). 2013 County population estimates: Race & ethnicity. *Carolina Demography*. https://carolinademography.cpc.unc.edu/2014/06/30/2013-county-population-estimates-race-ethnicity

Towle, P. O., & Patrick, P. A. (2016). Autism spectrum disorder screening instruments for very young children: A systematic review. *Autism Research and Treatment*, *2016*, 4624829. https://doi.org/10.1155/2016/4624829

Turner-Brown, L. M., Baranek, G. T., Reznick, J. S., Watson, L. R., & Crais, E. R. (2013). The First Year Inventory: A longitudinal follow-up of 12-month-old to 3-year-old children. *Autism*, *17*(5), 527–540. https://doi.org/10.1177/1362361312439633

Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psychoeducational Assessment*, *29*(4), 364–376. https://doi.org/10.1177/0734282911406666

Watson, L. R., Baranek, G. T., Crais, E. R., Reznick, J. S., Dykstra, J., & Perryman, T. (2007). The first year inventory: Retrospective parent responses to a questionnaire designed to identify one-year-olds at risk for autism. *Journal of Autism and Developmental Disorders*, *37*(1), 49–61. https://doi.org/10.1007/s10803-006-0334-4

Wei, T., Chesnut, S. R., 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. https://doi.org/10.1007/s10862-014-9468-4

Wetherby, A. M., Brosnan-Maddox, S., Peace, V., & Newton, L. (2008). Validation of the Infant—Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism*, *12*(5), 487–511. https://doi.org/10.1177/1362361308094501

Wolff, J. J., Boyd, B. A., & Elison, J. T. (2016). A quantitative measure of restricted and repetitive behaviors for early childhood. *Journal of Neurodevelopmental Disorders*, *8*(1), 1–10. https://doi.org/10.1186/s11689-016-9161-x

Zwaigenbaum, L., Bauman, M. L., Fein, D., Pierce, K., Buie, T., Davis, P. A., Newschaffer, C., Robins, D. L., Wetherby, A., Choueiri, R., Kasari, C., Stone, W. L., Yirmiya, N., Estes, A., Hansen, R. L., McPartland, J. C., Natowicz, M. R., Carter, A., Granpeesheh, D., . . .Wagner, S. (2015). Early screening of autism spectrum disorder: Recommendations for practice and research. *Pediatrics*, *136*(Supplement 1), S41–S59. https://doi.org/10.1542/peds.2014-3667D