

# Factor Analysis of the Childhood Autism Rating Scale in a Sample of Two Year Olds with an Autism Spectrum Disorder

Emily Moulton<sup>1</sup>  · Kathryn Bradbury<sup>1</sup> · Marianne Barton<sup>1</sup> · Deborah Fein<sup>1</sup>

Published online: 14 October 2016  
© Springer Science+Business Media New York 2016

**Abstract** The Childhood Autism Rating Scale (CARS), (Schopler et al. in *J Autism Dev Disord* 10(1):91–103, 1980) is a 15-item observation-based rating scale that yields a total score reflective of autism symptom severity. This study investigated the factor structure of the CARS in a sample of 2-year-old children with DSM-IV-TR (American Psychiatric Association in *Diagnostic and statistical manual of mental disorders*, 4th edn. American Psychiatric Publishing, Washington, 2000) diagnoses of AD or PDD-NOS. Following a preliminary internal cross-validation, principal axis factor analysis was completed ( $N=282$ ). The results indicate a three-factor solution: Social Communication, Stereotyped Behaviors and Sensory Sensitivities, and Emotional Reactivity. The factors are meaningful, with the first two reflective of DSM-5 symptom domains. This study supports the continued relevance of the CARS in ASD assessment, and extends its utility in 2-year-old children.

**Keywords** CARS · Factor analysis · ASD assessment

## Introduction

Autism Spectrum Disorders (ASDs) are a group of neuro-developmental disorders characterized by persistent deficits in social communication and interaction, and by the

presence of restricted, repetitive behaviors and/or interests, which may include sensory sensitivities. In addition to these core deficits, individuals with ASD often experience a number of comorbid deficits including cognitive delays/intellectual disabilities, motor delays, adaptive skill deficits, anxiety and aggressive/destructive behavior (Charman et al. 2011; Johnson and Myers 2007; Kerns and Kendall 2012; Levy et al. 2009; Lloyd et al. 2013; Macdonald et al. 2013; McDougle et al. 2003; Volkmar et al. 2004). The Center for Disease Control (CDC 2016) reports an overall prevalence rate for ASDs of one in 68, with boys affected at greater rates than girls (4.5:1). Given increases in the understanding of the early behavioral profiles of ASD, the American Academy of Pediatrics recommends routine ASD screening at 18 and 24 months of age (Johnson and Myers 2007; Zwaigenbaum et al. 2015). It is recommended that children that screen positive on screening measures, such as the Modified Checklist for Autism in Toddlers with Follow Up (M-CHAT/F; Robins et al. 1999), be immediately referred for evaluation (Zwaigenbaum et al. 2015). Through gold standard developmental and diagnostic evaluations, reliable and stable diagnoses can often be made in early childhood, at around 24 months or even earlier (Chawarska et al. 2009; Eaves and Ho 2004; Guthrie et al. 2013; Kim et al. 2015; Kleinman et al. 2008; Lord 1995; Moulton et al. 2016; Ozonoff et al. 2015; Sutura et al. 2007; Steiner et al. 2012; Turner and Stone 2007).

## ASD Diagnostic Procedures for Toddlers

Gold standard developmental and diagnostic evaluations of children presenting with ASD concerns in the toddler years assess functioning in multiple domains, including early cognitive abilities, social communication and interaction

✉ Emily Moulton  
Emily.Moulton@uconn.edu

<sup>1</sup> Department of Psychological Sciences, University of Connecticut, Storrs, CT, USA

abilities, adaptive skills, and the presence of atypical motor and/or sensory behaviors (Steiner et al. 2012). It is recommended that ASD assessments utilize multiple measures and methodologies (e.g., structured and semi-structured measures, parent report, direct observation), and that final diagnosis be based on expert clinical opinion (Steiner et al. 2012). Commonly utilized ASD-specific measures include the Autism Diagnostic Observation Schedule (Lord et al. 2000), the Autism Diagnostic Interview—Revised (Rutter et al. 2003), and the Childhood Autism Rating Scale (Schopler et al. 1980). Clinical judgment in the assignment of ASDs has been shown to have high inter-rater reliability and is considered best practice in the field of ASDs (Klin et al. 2000; Steiner et al. 2012).

### Role of the Childhood Autism Rating Scale

The Childhood Autism Rating Scale (CARS, Schopler et al. 1980) is a 15-item observation-based rating scale designed to accurately differentiate children with autism from those with developmental delays without features of autism. The CARS is intended for use by highly trained raters in the context of a wider multi-method approach that includes behavioral observations, interview of primary caregivers, assessment of intellectual functioning, and detailed developmental and family history (Schopler et al. 1980). Raters are to base their ratings on the frequency, intensity, duration and atypicality of the specified behavior, while considering the chronological age of the child. Each of the 15 items is rated on a seven-point scale (1, 1.5, 2...4) ranging from “within normal limits for that age,” which is coded as one, to “severely abnormal for that age,” which is coded as four (Schopler et al. 1995). A total score is determined by summing the ratings on all 15 items. CARS total scores range from a low of 15 (within normal limits on all items) to a high of 60 (severely abnormal on all items). In 2010, Schopler and colleagues released the CARS, Second Edition (CARS2), which includes a Standard Form (CARS2-ST, previously named the CARS), a High-Functioning Version (CARS2-HF) and a Questionnaire for Parents or Caregivers (CARS2-QPC). The CARS2-ST is identical to the original CARS in the rating scale used and the items included, but includes updated formatting to enhance its ease of use (Schopler et al. 2010). Given that CARS and CARS2-ST items are the same, investigations of the CARS remain applicable to current practice.

By comparing CARS total scores of 1520 children to corresponding expert clinical assessment, Schopler et al. (1995) created a diagnostic categorization system in which CARS total scores below 30 indicate that an individual is “non-autistic,” while scores of 30 or above indicate that an individual is “autistic.” Individuals with scores above 30 are

further subdivided into having “mild to moderate autism” (30–36.5) or “severe autism” (37–60).

The psychometric properties of the CARS are generally strong, with some variability among specific populations and age groups. A meta-analysis of research utilizing the CARS between 1980 and 2012 found good inter-rater reliability (0.796) and good internal consistency (0.896; Breidbord and Croudace 2013). Chlebowski et al. (2010) also found a high degree of internal consistency, with a coefficient alpha of 0.90 for a 2-year-old sample and 0.93 for a 4-year-old sample. Inter-rater reliability has been reported to be somewhat lower in older children (0.79), and in adolescent and young adult samples (0.73), but this may be improved by eliminating the “level and consistency of intellectual response” item (Garfin and McCallon 1988). Test–retest reliability is complicated by expected changes following intervention; however, it has generally been found to be adequate [0.77 over 3 months (Perry et al. 2005) and 0.88 over 1 year (Schopler et al. 1995)].

The validity of diagnoses based on CARS total scores has been investigated by comparing CARS classifications to other well-established diagnostic measures [e.g., the Autism Diagnostic Interview, Revised (ADI-R) and the Autism Diagnostic Observation Schedule (ADOS)], to DSM-IV(-TR) diagnoses, and to clinical judgment. Across studies, findings vary based on the CARS cutoff score selected, and the age of participants included. Using a CARS cutoff of 30, correlations between the CARS and the ADI-R completed by independent raters based on separately collected information, have revealed acceptable convergent validity (66.7–85.7% agreement; Saemundsen et al. 2003; Pilowsky et al. 1998). The highest rate of agreement was found in the sample with the highest mean age (87.5% in Pilowsky et al. 1998). When using a CARS cutoff of 30 in children 16–30 months, similar rates of agreement have been found between the CARS and the ADOS ( $\kappa=0.691$ ), and between the CARS and clinical judgment ( $\kappa=0.691$ ; Ventola et al. 2006); however, in these instances, the clinician completing the CARS was present for the ADOS and determined diagnoses with knowledge of the CARS. In a study of 3 year olds with ASD, in which the CARS and ADOS were administered by different clinicians in separate sessions, the correlation between CARS total scores and ADOS total Calibrated Severity Scores (ADOS CSS) has been found to be medium to large (0.432; Reszka et al. 2014).

Agreement between CARS diagnostic classification (i.e., scores of 30+ indicating that a child is “autistic”) and DSM-IV(-TR) Autistic Disorder (AD) is strong (100% in Rellini et al. 2004; 88% in Perry et al. 2005). Relatively weaker agreement has been found between the CARS diagnostic cutoff score of 30 and DSM-IV(-TR) Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS;

Rellini et al. 2004; Perry et al. 2005). In order to reflect the understanding of ASD as a broad spectrum including PDD-NOS, and to extend the clinical utility of the CARS to younger children, additional cutoffs have been recommended. Chlebowski et al. (2010) derived the following suggested cutoffs for 2 and 3 year olds, based on data in their large sample: 25.5–32 for PDD-NOS, 32+ for AD. For 4 year olds, the following cutoffs are recommended: 25.5–30 for PDD-NOS, 30+ for AD (Chlebowski et al. 2010).

In order to directly assess the utility of the CARS given the recent diagnostic changes in the DSM-5 (APA 2013), Mayes et al. (2014) examined the diagnostic agreement between the CARS and DSM-5. Using a range of CARS cutoff scores based upon age and intellectual ability, diagnostic agreement between the CARS and DSM-5 was 84% (Mayes et al. 2014). This finding is similar to findings of diagnostic agreement between the CARS and DSM-IV(-TR) diagnostic criteria (86%; Chlebowski et al. 2010). Additional work has shown that diagnostic sensitivity of the CARS2-ST based on the DSM-5 was 0.84, compared with 0.81 for the DSM-IV-TR (Dawkins et al. 2016).

While updating cutoff scores has helped to keep the CARS relevant to a broadened conceptualization of ASD, some critical limitations remain. ASDs are highly heterogeneous in their symptom presentation, with significant variability in the severity of core deficits within individuals (Johnson and Myers 2007). For example, a child may show severe deficits in the domain of social communication/interaction, with relatively less severe restricted, repetitive behaviors or interests. Therefore, a single total score may not best represent the severity of a child’s ASD, and in turn, may not be the most useful in determining diagnosis, or in

the subsequent development of treatment recommendations. In order to address this limitation, several factor structure analyses of the CARS have been conducted (see Table 1).

Previously conducted exploratory factor analyses of the CARS have resulted in mixed findings, with factor solutions ranging from three (DiLalla and Rogers 1994) to five factors (Stella et al. 1999). The proportion of variance accounted for by factors has been generally adequate, but somewhat variable (42% in Magyar and Pandolfi 2007; 69% in; DiLalla and Rogers 1994 and; Stella et al. 1999). Each existing factor analysis has utilized a distinct methodology, which likely contributes to the variability seen in findings across studies (see Table 1 for study characteristics).

With regard to the diagnostic inclusion criteria utilized, two studies included only children with Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised (DSM-III(R)) or DSM-IV Autistic Disorder (AD) or PDD-NOS (Magyar and Pandolfi 2007; Stella et al. 1999), while another (DiLalla and Rogers 1994) also included children with “non-pervasive developmental disorders” (e.g., Adjustment Disorder). Further, some studies required that the children met a CARS cut off of 30, while others included all children who met DSM criteria and let the CARS score vary freely (DiLalla and Rogers 1994; Magyar and Pandolfi 2007), while another (Stella et al. 1999) included children only when their DSM-III/DSM-IV diagnosis was consistent with their CARS diagnostic classification (i.e., only children with CARS scores of 30+). These sample differences likely resulted in differences in the range and variability of CARS scores included in each factor analysis and contributed to the variability in findings across studies (e.g., 3–5 factors retained).

**Table 1** Previous exploratory factor analyses of the childhood autism rating scale

	Sample size and diagnoses	Sample age	CARS methodology	Factor analysis methodology	Factor solution
DiLalla and Rogers (1994)	N=69 DSM-III(R): AD, PDD-NOS, “non-pervasive developmental disorder”	M=46.5 months Range: 24–73 months	20-min videotaped play interaction with a familiar adult	Principal components factor analysis	Three factors: social impairment, negative emotionality and distorted sensory response 69% of the variance accounted for
Stella, Mundy and Tuchman (1999)	N=90 DSM-III(R): AD, PDD-NOS	M=65.78 months Range: not reported	Direct observation and parent-report	Principal components factor analysis	Five factors: social communication, emotional reactivity, social orienting, cognitive and behavioral consistency, and odd sensory exploration 69% of the variance accounted for
Magyar and Pandolfi (2007)	N=164 DSM-IV: AD, PDD-NOS	M=43.27 months Range: 20–82 months	Direct observation, caretaker report and chart review	Principal axis factor analysis	Four factors: social-communication, social interaction, stereotypies and sensory abnormalities and emotional regulation 42% of the variance accounted for

AD autistic disorder, CARS childhood autism rating scale, DSM-III(R) diagnostic and statistical manual of mental disorders, third edition (revised), DSM-IV diagnostic and statistical manual of mental disorders, fourth edition, PDD-NOS pervasive developmental disorder, not otherwise specified

Studies also varied in the sample of behavior upon which CARS ratings were based, and in the age range of children included. Consistent with intended CARS procedures (Schopler et al. 1980), two studies based ratings on a combination of direct behavioral observations and parent/caregiver report (Magyar and Pandolfi 2007; Stella et al. 1999). In contrast, another study based CARS ratings solely on a 20-min videotape of the child interacting with an unfamiliar adult (DiLalla and Rogers 1994). While the mean age of children included in previous analyses was between 3 and 5 years, the ages of children included in each study ranged substantially, even up to 12 years (Stella et al. 1999). Given expected changes in symptom presentation as a child develops, the inclusion of a wide age range can be both a strength and a limitation. While it may result in findings that are applicable to a larger subset of children, it may also result in findings that are not well suited to specific age groups (e.g., toddlers). Beyond information about age, previous studies often included limited information about critical participant characteristics (e.g., cognitive and adaptive functioning levels), thereby limiting generalizability of results to new samples.

While each previous factor analysis has utilized a distinct methodology and yielded a unique result, findings all support a multi-factor solution that includes domains related to social communication/interaction, emotional reactivity and sensory sensitivities. Across previous analyses, factors related to social communication have included the highest number of items, and have accounted for the largest proportion of variance in CARS scores (38.5–52%; DiLalla and Rogers 1994; Magyar and Pandolfi 2007; Stella et al. 1999). As a result, each analysis results in factors that partially reflect DSM-IV(-TR) domains of ASD symptomatology (e.g., social interaction, communication), in addition to supporting the importance of understanding emotional reactivity and sensory sensitivities in children with ASD. Notably, sensory sensitivities are more clearly emphasized as a diagnostic symptom in the DSM-5 (APA 2013) than they were in previous editions of the DSM.

## The Present Study

Given increases in the understanding of the early behavioral profiles of ASD and the subsequent decrease in the age at which children can reliably be diagnosed with ASD, the present study seeks to extend previous investigations of the factor structure of the CARS to 2 year olds. Principal axis factor analysis was conducted, similar to Magyar and Pandolfi (2007), as a result of non-normality in the distribution of our data (Costello and Osborne 2005). In order to generalize findings to the majority of 2 year olds for whom the CARS may be completed, all children with DSM-IV-TR diagnoses of AD or PDD-NOS were retained, even if

their CARS score was below the cutoff of 25.5 for 2 year olds (Chlebowski et al. 2010). In addition, given the significant heterogeneity of the ASD population, the current study sought to thoroughly characterize the cognitive and adaptive functioning of the sample utilized. Further, given that the stability of factor analysis increases with increasing sample size, the current study utilized a larger sample than in previous analyses. Additionally, to address the likelihood of replication in additional samples, the current study included an internal split-half cross-validation, as is recommended in Thompson (2004) and Osborne and Fitzpatrick (2012). Given that larger samples yield more stable findings, the factor analysis conducted on the complete sample was utilized as the basis for interpretation (Thompson 2004).

Finally, to understand the resulting factors in reference to an established measure of ASD symptomatology, correlations between factor scores and ADOS Calibrated Severity Scores (Gotham et al. 2009; Hus et al. 2014) were calculated, and interpreted utilizing Cohen's (1988) conventions. In the present study, CARS scores were completed in part based on the observation of the ADOS (discussed further in the "Procedures" section, below). Therefore, the comparison of CARS factor scores and ADOS Calibrated Severity Scores should not be considered to be an independent validity assessment of CARS factors. Rather, these comparisons should serve as a basis for understanding the similarities and differences between these two measures of symptom severity, and in turn, their unique contributions to diagnostic procedures and treatment planning.

Based on previous analyses, it is hypothesized that a factor structure consistent with the core domains of DSM-IV-TR ASD (Social Communication, Social Interaction and Restricted, Repetitive Behaviors/Interests) will emerge, with the possibility of an additional factor reflecting emotional reactivity (see DiLalla and Rogers 1994; Magyar and Pandolfi 2007; Stella et al. 1999). Differences between the resulting factor structure and previously determined factor structures in older children may reflect important differences in symptom presentation between 2-year-old children and 4- to 5-year-old children with ASD. Increasing our understanding of the factors that emerge within the CARS total score for 2-year-old children will enhance this tool's utility in this age group, in part by better reflecting possible heterogeneity in symptom severity across domains by expanding the single total score into separate domain scores.

## Methods

### Participants

Participants include a subset of children participating in an ongoing study to evaluate the psychometric properties



of an autism-specific screening questionnaire, the Modified Checklist for Autism in Toddlers with follow-up (M-CHAT/F, Robins et al. 1999) and its revision (M-CHAT-R/F; Robins et al. 2009). Children in the current study ( $N=282$ ) were recruited for the study through their pediatrician ( $n=106$ ), early intervention provider or psychologist ( $n=123$ ), or caregiver self-referral ( $n=53$ ). Informed consent was obtained from parents of all children included in the study. This research was approved by the University of Connecticut Institutional Review Board (IRB) and the University of Washington IRB.

Following positive screening on the M-CHAT/F or the M-CHAT-R/F, children were provided a developmental and diagnostic evaluation. Children were included in the current study if they were evaluated between 16 and 32 months of age, had complete CARS, Autism Diagnostic Observation Schedule (ADOS) and Mullen Scales of Early Learning (Mullen) data, and received a DSM-IV-TR diagnosis of AD ( $n=160$ ) or PDD-NOS ( $n=122$ ). An additional descriptor of low mental age (Low MA) was given to 36 children who were functioning below the 12-month level across all domains except Fine Motor on the Mullen Scales of Early Learning (Mullen 1995). This descriptor was utilized to identify children who should be excluded from analyses including measures, like the ADOS (Lord et al. 2000), which are not well validated for use in this population.

See Table 2 for general sample characteristics. Children were evaluated at a mean age of 24.96 months ( $SD=3.91$ ; range=16.89–31.97). The sample was 77.3% male ( $n=218$ ) and 22.7% female ( $n=64$ ). This ratio (3.4:1) is largely consistent with the currently estimated gender ratio in the wider population of children with ASD (4.5:1; CDC 2016). The majority of children in the sample were White ( $n=206$ , 73.0%), as indicated by their caregivers (see

Table 2 for additional information). Children in the current sample had a broad range of cognitive and adaptive abilities as assessed by the Mullen and Vineland Adaptive Behavior Scales (VABS), and a broad range of symptom severity as assessed by the CARS and the ADOS CSS (see Table 3). ADOS CSS of children with Low MA were not included, given that the ADOS is not well validated for use with this population (Lord et al. 2000; Gotham et al. 2007). Mullen Early Learning Composite scores ranged from 41 to 120 ( $M=60.11$ ,  $SD=11.26$ ). VABS Adaptive Behavior Composite scores ranged from 50 to 99 ( $M=69.10$ ,  $SD=8.65$ ) (see Table 3).

**Procedure**

Children’s caregivers completed M-CHAT (Robins et al. 1999) or M-CHAT-R (Robins et al. 2009) screeners at their pediatrician’s office during their child’s 18 or 24-month well-child visit, or at an early intervention site or psychologist’s office. The M-CHAT(-R) is a brief, autism-specific, parent-report screening measure that consists of 23 (M-CHAT) or 20 (M-CHAT-R) yes or no questions. Completed questionnaires were sent to the University of Connecticut ( $n=252$ ) or University of Washington ( $n=30$ ) Early Detection laboratories to be scored. If a child screened positive, their caregiver was contacted via telephone to complete the relevant structured Follow-Up items. If a child continued to screen positive after the Follow-Up phone interview, and did not have severe sensory or motor deficits that would preclude evaluation with study instruments (e.g., blind and deaf),

**Table 2** General characteristics of sample ( $N=282$ )

	N	%
Gender		
Male	218	77.3
Female	64	22.7
Ethnicity		
White	206	73.0
Hispanic/Latino	24	8.5
Black or African American	16	5.7
Asian or Pacific Islander	13	4.6
Biracial	13	4.6
Unknown	10	3.5
DSM-IV-TR diagnosis		
Autistic disorder	160	56.7
PDD-NOS	122	43.3

*PDD-NOS* pervasive developmental disorder, not otherwise specified

**Table 3** Clinical characteristics of sample

Measure	<i>N</i>	<i>M</i> ( <i>SD</i> )	Range
CARS total score	282	32.42 (5.19)	20.0–48.5
Mullen scales of early learning ( $M=50$ , $SD=10$ )			
Visual reception	282	30.04 (9.86)	20–68
Fine motor	282	30.87 (10.44)	20–60
Receptive language	282	22.77 (6.86)	20–59
Expressive language	281	25.60 (8.19)	20–68
Vineland adaptive behavior scales ( $M=100$ , $SD=15$ )			
Communication	281	67.85 (9.35)	42–108
Daily living	281	73.48 (11.48)	25–119
Socialization	281	71.22 (9.13)	50–104
Motor skills	279	83.44 (10.97)	56–111
ADOS calibrated severity scores			
Social affect CSS	246	6.51 (2.14)	1–10
RRB CSS	246	6.34 (2.61)	1–10
Total CSS	246	6.31 (2.19)	1–10

*CARS* childhood autism rating scale, *ADOS* autism diagnostic observation schedules, *RRB* restricted and repetitive behavior, *CSS* calibrated severity score

he or she was invited to attend a free developmental and diagnostic evaluation conducted at the University of Connecticut or the University of Washington. Evaluations were conducted by a licensed clinical psychologist or a developmental pediatrician and a graduate student in clinical psychology. Evaluation procedures were standardized across clinicians, and included measures of cognitive skills, adaptive skills, language abilities and ASD symptoms. Measures were completed with the caregiver and child in the same room. At the conclusion of the evaluation, caregivers were provided with feedback regarding the assessment, which included any diagnoses the child might qualify for, as well as recommendations for intervention and resources.

A diagnosis of AD or PDD-NOS was assigned based on clinical judgment of experienced clinicians with expertise in ASD (licensed psychologists or developmental pediatricians) utilizing scores from all available measures including direct testing and parent interviews in accordance with the clinicians' best estimate diagnosis using DSM-IV-TR diagnostic criteria (APA 2000). DSM-IV-TR diagnostic criteria were utilized throughout this longitudinal project to maintain consistency despite recent changes in diagnostic criteria (DSM-5, APA 2013).

## Measures

The following measures were utilized in the ongoing study: M-CHAT/F, M-CHAT-R/F, ADOS, Toddler Autism Symptom Interview (TASI; Barton et al. 2012), Mullen, VABS, and the CARS. These measures are widely used in the field of ASD, and have excellent psychometric properties (with the exception of the TASI, which is currently being validated). Please note that several measures included in the current study have been revised since the initiation of this longitudinal project. Measures were kept consistent throughout the study (except where noted below) in order to facilitate comparisons between children and across time. The current study analyzes data from the measures described below.

### Autism Diagnostic Observation Schedule: Generic (ADOS)

The ADOS (Lord et al. 2000) is a semi-structured, standardized, play-based assessment of four domains of autism symptomatology: Reciprocal Social Interaction, Communication, Stereotyped Behaviors and Restricted Interests and Play. ADOS Modules 1 and 2 were utilized in the current study, as the ADOS, Second Edition, Toddler Module (ADOS-2; Lord et al. 2012) had not yet been released at the time of study initiation. In order to compare symptom severity across modules, (Gotham et al. 2009; Hus et al. 2014) developed the ADOS Calibrated Severity Score (CSS). The CSS is a measure of autism severity that takes

into account a child's age and language abilities, allowing for a measure of symptom severity that is less influenced by age or verbal abilities (Gotham et al. 2009; Hus et al. 2014). Procedures for calculating raw Social-Affect (SA) and raw Restricted Repetitive Behavior (RRB) scores were followed as per Gotham et al. 2007, and CSSs were calculated as per Gotham et al. 2009 and Hus et al. 2014. CSSs range from one to ten, with higher CSSs indicating greater severity. Total CSS, Social-Affect (SA) CSS and Restricted Repetitive Behavior (RRB) CSS were included in the current analyses.

### Childhood Autism Rating Scale (CARS)

The CARS (Schopler et al. 1980) is a 15 item, observation-based rating scale designed to accurately differentiate children with ASD from those with developmental delays without features of autism. Information regarding the psychometric properties of the CARS can be found above in the introduction. In the present study, CARS ratings were completed by a licensed clinical psychologist or developmental pediatrician based on direct observation of cognitive testing, the ADOS, and parent-report. A reformatted version, the CARS2-ST, was released in 2010 (Schopler et al. 2010), which includes the same items and scaling as the version utilized in the current project, and therefore, the investigations of the CARS remain applicable to current practice with the CARS2-ST.

### Vineland Adaptive Behavior Scales: Interview Edition (Versions I and II)

The VABS (Sparrow et al. 1984) is a structured, parent-report interview measure of adaptive functioning across four domains: Communication, Daily Living Skills, Socialization and Motor Skills. Scores are determined for each domain, which are combined to form a total score (the Adaptive Behavior Composite, ABC). In the current study, children's caregivers were administered the VABS (Sparrow et al. 1984) or an updated version which was released in 2005, the Vineland Adaptive Behavior Scales—Second Edition (VABS-II; Sparrow et al. 2005). As a result of the high degree of similarity between the two versions, VABS and VABS-II scores were analyzed collectively.

### Mullen Scales of Early Learning

The Mullen (Mullen 1995) assesses five domains of cognitive development: Visual Reception (problem solving abilities), Gross Motor, Fine Motor, Expressive Language and Receptive Language. The measure provides a summative "Early Learning Composite" (ELC) score, which is computed from the Visual Reception, Fine Motor, Expressive

Language and Receptive Language domains. In the current study, the Gross Motor scale was not administered.

## Results

The sample consisted of children diagnosed with ASD displaying a range of symptom severity. The mean CARS score was 32.42 ( $SD=5.19$ ), which falls within the “mild to moderate autism” range based on original CARS cutoff of 30 (Schopler et al. 1995), and on the upper end of the “PDD-NOS” range based on updated cutoff of 25.5 for 2 and 3-year-old children (Chlebowski et al. 2010). CARS scores ranged from 20.0 to 48.5, indicating that ASD symptom severity varied widely across children. Scores on other clinical measures can be found in Table 3.

### Internal Principal Axis Factor Analysis Cross-Validation

In order to provide preliminary information regarding the potential replicability of the factor analytic structure found utilizing the current sample, an internal cross-validation was conducted as per the procedures outlined in Thompson (2004) and Osborne and Fitzpatrick (2012). The sample was randomly split into two equal groups ( $n=141$  each). Separately, in each group, the 15 items of the CARS were factor analyzed using principal axis factor analysis with Promax (oblique) rotation (Costello and Osborne 2005), as the factors were expected to be correlated based on the findings of Magyar and Pandolfi (2007). In each group, the data were deemed suitable for factor analysis as the respective Kaiser–Meyer–Olkin values (0.834, 0.843) exceeded the recommended value of 0.6 (Kaiser 1960, 1974) and Bartlett’s Test of Sphericity (Bartlett 1954) reached statistical significance. Factors were extracted based on scree test criteria (Cattell 1966), as recommended by Costello and Osborne (2005), and consistency in findings across samples was assessed. Each sample supported a three-factor solution, and 14 out of 15 items yielded the highest item loadings on equivalent factors across the two samples. Only one item (Visual Response) did not load on equivalent factors across samples. Within each factor, squared differences between item factor loadings in the two samples were calculated, and 12 out of 14 items demonstrated squared differences  $<0.4$ . According to criteria outlined in Osborne and Fitzpatrick (2012), this indicates a good degree of consistency, and in turn, supports the potential for replicability of factor analysis findings in additional independent samples. As a result of these preliminary findings, following the recommendations of Thompson (2004), a principal axis factor analysis utilizing our complete sample was conducted.

### Whole Sample Principal Axis Factor Analysis

In the complete sample ( $N=282$ ), the 15 items of the CARS were factor analyzed using principal axis factor analysis with Promax (oblique) rotation, with extraction based on scree test criteria (as above). Data was deemed suitable for factor analysis based on an acceptable Kaiser–Meyer–Olkin value (0.862) and significant Bartlett’s Test of Sphericity (Bartlett 1954). Principal axis factor analysis resulted in three factors with eigenvalues above the “break point” in the scree plot that accounted for 51.45% of the common variance. Eigenvalues of retained factors all exceeded a value of 1.0. Items were retained in a factor if their factor pattern loading exceeded 0.40. Table 4 presents the item correlation matrix. See Table 5 for the pattern loadings by factor.

The first factor accounted for 32.64% of the variance and was labeled Social Communication as it consisted of six items relating to social interaction and communication with others, and one item related to intellectual functioning. The second factor, Emotional Reactivity, accounted for 9.87% of the variance and consisted of three items relating to emotional reactivity. The final factor accounted for 8.94% of the variance and was labeled Stereotyped Behaviors and Sensory Sensitivities as it consisted of four items relating to stereotypy and unusual sensory responsivity. One item (Visual Response) did not load on any factor (pattern loading  $<0.40$ ). All other items significantly loaded on only one factor, indicating good discriminability between factors. Correlations between factors were medium to large per Cohen (1988; see Table 6). Internal consistency was good for the Social Communication factor ( $\alpha=.86$ ), while it was lower for the Emotional Reactivity ( $\alpha=.50$ ) and Stereotyped Behaviors and Sensory Sensitivities ( $\alpha=.53$ ) factors.

As the last item of the CARS (General Impression) is meant to be a rating of overall ASD severity, the factor analysis was re-run without this item. The factor structure continued to hold and less variance was explained, thus the General Impression item was retained as part of the Social Communication factor.

### Mean Factor Scores

Mean factor scores were calculated for each factor within the sample such that factor scores could be conceptualized similarly to individual CARS items with scores ranging from 1 (behavior within normal limits) to 4 (severely abnormal behavior). The highest mean factor score was obtained on the Social Communication factor ( $M=2.48$ ,  $SD=0.46$ ), which fell in the “mildly to moderately abnormal” range. The Stereotyped Behaviors and Sensory Sensitivities factor score ( $M=1.94$ ,  $SD=0.43$ ) fell in the “mildly abnormal” range. As expected, the Emotional Reactivity factor score was the lowest,  $M=1.74$  ( $SD=0.48$ ).

**Table 4** Item correlation matrix

Item #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	1.00														
2	0.61	1.00													
3	0.35	0.34	1.00												
4	0.16	0.26	0.16	1.00											
5	0.37	0.45	0.21	0.36	1.00										
6	0.14	0.12	0.37	0.07	0.17	1.00									
7	0.40	0.32	0.30	0.22	0.34	0.13	1.00								
8	0.53	0.40	0.25	0.20	0.37	0.10	0.33	1.00							
9	0.09	0.17	0.15	0.24	0.24	0.06	0.08	0.18	1.00						
10	0.02	0.08	0.22	0.02	0.12	0.16	0.08	0.04	0.08	1.00					
11	0.43	0.58	0.17	0.19	0.33	0.04	0.22	0.42	0.19	-0.06	1.00				
12	0.56	0.53	0.21	0.21	0.32	0.12	0.29	0.43	0.13	0.03	0.49	1.00			
13	0.03	0.02	0.10	0.15	0.18	-0.12	0.09	0.01	0.17	0.03	0.09	-0.09	1.00		
14	0.35	0.32	0.11	0.13	0.25	0.03	0.19	0.28	0.07	-0.12	0.39	0.28	0.23	1.00	
15	0.70	0.67	0.46	0.25	0.52	0.22	0.45	0.51	0.24	0.12	0.55	0.51	0.10	0.35	1.00

**Relationship with Other Severity Measures**

Pearson correlations were used to examine the relationship between CARS factors and ADOS calibrated severity scores (CSS), specifically when CARS ratings were made in part based on ADOS observation. Given that the ADOS is not well-validated for use with children with a mental age below 12 months (Lord et al. 2000; Gotham et al. 2007), children with Low MA were removed from the ADOS analyses, resulting in sample size of 246. There was a medium to large (Cohen 1988), positive correlation between the CARS Social Communication Factor mean score and the ADOS Total CSS ( $r = .449, p < .001$ ) and the ADOS Social Affect CSS ( $r = .507, p < .001$ ). There was also a medium to large (Cohen 1988), positive correlation between the CARS Stereotyped Behaviors and Sensory Sensitivities Factor mean score and the ADOS RRB CSS,  $r = .411 (p < .001)$ . Weaker (small to medium per Cohen 1988) correlations were seen between the CARS Emotional Reactivity Factor and all ADOS CSSs. Relationships between all CARS factors and ADOS severity scores can be found in Table 7.

**Discussion**

**Whole Sample Principal Axis Factor Analysis**

The current study investigated the factor structure of the CARS in a large sample of 2-year-old children ( $M = 24.96$  months;  $SD = 3.91$ ; range = 16.89–31.97) with DSM-IV-TR (APA 2000) diagnoses of AD or PDD-NOS. An internal split-half cross-validation supported the potential for replication of factor analysis findings utilizing the current sample, and therefore, a factor analysis was conducted in the complete sample. The results of the whole-sample principal axis factor analysis indicated a conceptually meaningful three-factor solution that accounted for 51.45% of the variance in CARS scores. The first factor, Social Communication, accounted for 32.64% of the variance, the second factor, Emotional Reactivity, accounted for 9.87% of the variance, and the final factor, Stereotyped Behaviors and Sensory Sensitivities, accounted for 8.94% of the variance.

These findings are partially supportive of our hypothesis that factors would reflect a child’s emotional reactivity, in addition to the core symptom domains of DSM-IV-TR ASD (Social Communication, Social Interaction, and RRBs). Importantly, however, consistent with revised DSM-5 symptom domains, our analyses resulted in a single Social Communication factor, as opposed to hypothesized separate social communication and social interaction factors as per DSM-IV-TR. In sum, it appears that two factors identified in the current analyses are reflective of the DSM-5 symptom domains (social communication, stereotyped behaviors and



**Table 5** Factor pattern loadings from principal axis factor analysis with promax rotation of CARS items

CARS item (item no.)	Factor		
	Social communication	Emotional reactivity	Stereotyped behaviors and sensory sensitivities
Relating to people (1)	0.834		
Verbal communication (11)	0.783		
Nonverbal communication (12)	0.779		
Imitation (2)	0.753		
General impressions (15)	0.687		
Listening response (8)	0.601		
Level and consistency of intellectual response (14)	0.567		
Emotional response (3)		0.666	
Adaptation to change (6)		0.557	
Fear or nervousness (10)		0.436	
Activity level (13)			0.502
Body use (4)			0.455
Object use (5)			0.447
Taste, smell, touch response and use (9)			0.423

Only pattern loadings >0.4 are reported

**Table 6** CARS factor intercorrelations

	Factor		
	SC	ER	SB and SS
SC	–		
ER	0.531	–	
SB and SS	0.517	0.401	–

SC social communication, ER emotional reactivity, SB and SS stereotyped behaviors and sensory sensitivities

sensory sensitivities). As a result, the current study provides initial support for the continued relevance of the CARS in the diagnosis of ASD in 2 year olds using the DSM-5. Further, the current study’s finding of an additional Emotional Reactivity factor supports the growing body of literature highlighting the importance of understanding emotion

regulation and emotional distress in individuals with ASD (e.g., Mazefsky 2015).

In the current study, in the large majority of cases, items loaded onto factors in a conceptually meaningful way (see Table 5 for item loadings). Two possible exceptions were noted in the Social Communication domain, on which the following items loaded: “level and consistency of intellectual response” and “general impressions.” Unlike the other items that loaded on this factor, these items do not appear to directly reflect a child’s social or communication abilities. It is possible that clinicians see variability in a child’s intellectual profile as an identifying feature of ASD, and therefore, rate the “intellectual response” item similarly to those assessing the core features of the disorder (i.e., social and communication items). Similarly, given that social and communication deficits are often considered the core features of ASD, clinicians may make their “general impressions” based on a child’s abilities in these areas. If these hypotheses about clinician beliefs are accurate, they may explain why the “general impressions” and “level and consistency of intellectual responses” items loaded with items more directly reflective of a child’s social and communication functioning. Future studies may benefit from testing this hypothesis directly through qualitative or quantitative reports by clinicians regarding their use of these items.

In the current study, one item (Visual Response) did not load onto any factor (pattern loading <0.40). This is consistent with findings in the internal cross-validation where the item did not load onto equivalent factors in separate samples. The mean score on this item indicates that this is an area of mild to moderate symptom severity for 2 year olds, and therefore, understanding why this item did not load onto any factor may be important. The wording of this item is somewhat vague, possibly leading to differences in raters’ perception of the item. Some clinicians may rate this item based on the quality of a child’s eye contact, whereas others may base this rating on a child’s visual sensory behaviors (i.e., atypical visual sensory seeking). In support of this interpretation, the Visual Response item showed similar loading values for the Social Communication factor (0.31) and for the Stereotyped Behaviors and Sensory Sensitivities factor (0.20), possibly reflecting that some clinicians rate this item based on social behavior (i.e., eye contact), while others rate it based on a stereotyped behavior or sensory sensitivity (i.e., atypical visual sensory seeking). Clarification of the meaning of this item may improve its utility in understanding a child’s symptom severity, and in turn, in establishing treatment recommendations.

In the current study, internal consistency was good for the Social Communication factor ( $\alpha = .86$ ), while it was lower for the Emotional Reactivity ( $\alpha = .50$ ) and Stereotyped Behaviors and Sensory Sensitivities ( $\alpha = .53$ ) factors. This is consistent with the CARS emphasis on social interaction

**Table 7** Correlations between CARS factor scores and ADOS calibrated severity scores

	CARS factor scores		
	SC	ER	SB and SS
ADOS calibrated severity scores			
Social affect CSS	0.507***	0.324***	0.145*
RRB CSS	0.243***	0.140*	0.411***
Total CSS	0.449***	0.300***	0.275***

SC social communication, ER emotional reactivity, SB and SS stereotyped behaviors and sensory sensitivities, RRB restricted and repetitive behaviors, CSS calibrated severity score

\* $p \leq .05$

\*\*\* $p \leq .001$

and communication symptoms, and in turn, with the larger number of items loading on the Social Communication factor. Future revisions of the CARS may consider including additional items assessing emotional reactivity and stereotyped behaviors and sensory sensitivities to increase the internal consistency and reliability of these domains.

### Mean Factor Scores

In the current sample of 2-year-old children, Social Communication symptoms were the most severe (mild to moderately abnormal), followed by Stereotyped Behaviors and Sensory Sensitivities (mildly abnormal) and Emotional Reactivity (mildly abnormal). Differences between mean severity scores across domains support the use of individual factor scores, in addition to an overall severity score. By utilizing individual factor severity scores, we can better reflect heterogeneity in a child's symptom severity across domains. Differences in mean symptom severity between the Social Communication and Stereotyped Behaviors and Sensory Sensitivities domains may reflect the fact that 2-year-old children diagnosed with PDD-NOS in the current study may not have displayed significant repetitive behaviors, whereas all children displayed social difficulties. Some studies have suggested that RRBs may emerge later in the preschool years in some children with ASD (e.g., Barton et al. 2013).

This difference may also reflect the challenges of assessing stereotyped behaviors and sensory sensitivities in toddlers. Assessing stereotyped behaviors and sensory sensitivities in this age group is complicated by the fact that repetitive motor behaviors are often present in typically developing 2 year olds, and help children to master more complex motor skills (Harrop et al. 2014). Further, whereas repetitive motor behavior decreases as children master more complex skills during their toddler years, adherence to routines and interest in specific objects remains common through the preschool years, and, in turn, can be difficult

to distinguish from atypical behavior (Leekam et al. 2007; Wolff et al. 2014). As a result, it may be more difficult to determine whether a child's repetitive behaviors and restricted interests fall outside of the typical range, resulting in generally lower scores for all 2 year olds in this domain. Further, while clinicians utilized information from caregiver report in addition to direct observations during the 3-h evaluation, it is possible that ratings were lower in this domain because stereotyped behaviors were not observed during the time-limited evaluation period.

In the current sample of 2-year-old children, mean Emotional Reactivity and Stereotyped Behaviors and Sensory Sensitivities scores reflected a similar mean level of severity (mildly abnormal). This is notable given that while stereotyped behaviors are included as a core symptom domain in DSM-IV-TR and DSM-5, emotional reactivity is not. Emotion regulation difficulties are certainly not specific to ASD, and therefore, likely would not be a meaningful core symptom domain; however, understanding difficulties in emotion regulation in this group may prove helpful in creating meaningful subgroups in research studies (Mazesfky 2015), and in planning appropriate intervention targets. Given that the gold standard diagnostic measures, the ADOS and ADI-R, do not emphasize this domain of functioning, the CARS could provide useful unique information in this area. Subsequently, if clinician's find that a child's CARS Emotional Reactivity factor is elevated, they may consider the use of a more thorough assessment of emotion regulation (e.g., Infant-Toddler Social and Emotional Assessment, Carter and Briggs-Gowan 2006; Pervasive Developmental Disorders Behavior Inventory; Cohen et al. 2003).

### Relationship with Other Severity Measures

Comparisons between CARS factor scores and ADOS Calibrated Severity Scores were conducted in order to better understand CARS factor scores in relation to an established, and commonly used, measure of ASD symptom severity. Findings provide information regarding similarities and differences between these two measures, but should not be considered an independent "validation" of CARS factor scores, given that CARS ratings were completed in part based on ADOS observations. Factor correlations were medium to large between the CARS Social Communication factor and the ADOS Social Affect and Total CSS values. Similarly, medium to large correlations were found between the CARS Stereotyped Behaviors and Sensory Sensitivities factor and the ADOS RRB CSS value. This is to be expected given that the CARS and the ADOS assess overlapping symptoms in the core domains of ASD (Social Communication and RRBs). Relatively weaker, small to medium correlations were found between the CARS Emotional Reactivity domain and all ADOS CSS values, likely as a result of the

greater emphasis on emotional responsivity in the CARS than in the ADOS. Thus, CARS factor severity scores, and specifically, the Emotional Reactivity score, may serve as an important addition to diagnostic information obtained from the ADOS.

### Comparison to Previous Exploratory Factor Analyses

In regards to number of factors retained, the specific findings of the current study were most similar to those of DiLalla and Rogers (1994), who also found that a three-factor solution best fit the data. They identified three factors which were highly similar to those identified here (Social Impairment, Negative Emotionality and Distorted Sensory Response), with some differences in specific item loadings. Similar to the current study, DiLalla and Rogers (1994) included children with a wide range of ASD symptom severity and utilized highly trained raters. Similarities in findings between the two studies remain notable, however, considering significant differences in the type of information on which CARS ratings were based (see Table 1). DiLalla and Rogers (1994) based ratings on a 20-min videotaped interaction with a familiar adult, which appears to contrast significantly with the methods of the current study in which the CARS was rated based on a 3-h evaluation including both direct observation and caregiver report. It is possible, however, that similar results were found despite this apparent difference in methods because DiLalla and Rogers (1994) explicitly elicited each behavior to be rated on the CARS during their 20-min observation, thereby collecting all required information during a brief period.

Findings of the current study were also largely similar to those in Magyar and Pandolfi (2007), who identified four factors and utilized the most similar methodologies to the current study (see Table 1). Despite the difference in the number of factors identified (three in the current study vs. four in Magyar and Pandolfi 2007), individual item loadings for Magyar and Pandolfi's (2007) Social Communication, Stereotypies and Sensory Sensitivities, and Emotional Regulation factors were highly similar to those found on equivalent factors in the current study. The major difference in the two studies findings is that in Magyar and Pandolfi (2007), two separate factors reflective of social communication abilities emerged (i.e., Social Communication, Social Interaction), whereas in the current study, only one general Social Communication factor was identified. Notably, Magyar and Pandolfi's (2007) Social Interaction factor only included two items, and by certain conventions (e.g., Costello and Osborne 2005), factors with fewer than three items are considered weak and unstable. In addition, Magyar and Pandolfi (2007) did not perform any analyses addressing the potential replicability of their findings (e.g., an internal cross-half validation), and

therefore, it is possible that this fourth factor may not hold with replication.

The greatest difference in findings was found between the current study and Stella, Mundy and Tuchman (1999), who identified five factors. Importantly, however, as with Magyar and Pandolfi (2007), Stella and colleagues (1999) chose to retain factors with less than three items. As a result, two of the five factors identified may be weak and unstable, and in turn, difficult to replicate. If only factors which included three or more items were considered (Social Communication, Social Orienting, and Emotional Reactivity), critical differences would still exist between the findings of the current study and Stella et al. (1999). This may be the result of the inclusion of significantly older children in the latter study, a population in which the CARS may perform differently.

In contrast, studies with more similar findings to the current study (DiLalla and Rogers 1994; Magyar and Pandolfi 2007), utilized younger children and a narrower age range (age 4–6 years). Future work is needed to determine the extent of differences in appropriate CARS factors between children of different age groups. Specifically, future work should attempt to directly compare factors identified in children of different age groups, using otherwise highly similar methodologies (e.g., type of factor analysis conducted, inclusion criteria, basis of ratings, qualifications of raters). In doing so, we can begin to establish appropriate age groupings for which certain CARS factors can be utilized. At present, given the absence of a single study including children of different ages, it is recommended that researchers and clinicians utilize factors determined by studies with a sample closest in age to the child of interest.

### Limitations and Future Directions

There are several limitations to the current study. First, children included in the current study all received DSM-IV-TR (APA 2000) diagnoses of AD or PDD-NOS, as opposed to a DSM-5 diagnoses of ASD. Barton et al. (2013) found that when applying DSM-5 criteria to toddlers with DSM-IV-TR diagnoses of AD or PDD-NOS, 29% of children no longer met diagnostic criteria. Therefore, it is unclear to what extent the findings of the current study apply to children diagnosed with ASD using DSM-5. Importantly, however, diagnostic agreement between the CARS and DSM-5 has been found to be similar to the diagnostic agreement with DSM-IV-TR (Mayes et al. 2014; Dawkins et al. 2016). Additionally, the factors identified in the current study appear to overlap with the two core symptom domains emphasized in DSM-5 (social communication, stereotyped behaviors and sensory sensitivities), with an additional factor reflective of emotional reactivity. In order to address this possible limitation in the generalizability of the factors identified

in the current study, future work is required to confirm the appropriateness of the current factor structure in a sample of 2-year-old children who meet DSM-5 criteria for ASD (i.e., a confirmatory factor analysis). Our group is currently collecting data utilizing DSM-5 ASD criteria, as well as current assessment measures such as the ADOS-2 Toddler Module (Lord et al. 2012), that will allow for a confirmatory factor analysis once a sufficient sample size is achieved.

Secondly, this study utilized a sample of children within a narrow age range (16–32 months). This can be considered both a strength and a limitation of the current project. Including a narrow age range allowed for the determination of factors specific to two-year-old children with ASD, an increasingly important group given the trend toward earlier diagnosis. As a result of the narrow age range, however, a direct comparison between factors identified in 2-year-old vs. 4-year-old children cannot be made, and in turn, whether current findings can be extended to older children with ASD cannot be determined. As noted above, this is an important direction for future work.

Thirdly, given that CARS ratings were based in part on ADOS observations, an independent validation of CARS factors could not be conducted as a component of the current study. As has been done with the CARS total score, future work should consider comparing CARS factor scores to other well-established diagnostic measures (e.g., ADI-R, ADOS) and clinical judgment, when these measures are completed by independent raters based on separately collected information. Following confirmatory factor analyses, and independent validations of the current factors, future work may also consider investigating the relationship between factor scores and other domains of functioning, including a child's cognitive abilities.

## Conclusions

Given recent advances in understanding early behavioral profiles of ASD, and earlier detection and intervention, it is critical that diagnostic measures, such as the CARS, be appropriate for use with 2-year-old children. The current study extended previous research by investigating the factor structure of the CARS in this population. The results of the current study indicate a three-factor solution that accounts for 51.45% of the variance in CARS scores and identifies the following factors: Social Communication, Stereotyped Behaviors and Sensory Sensitivities, and Emotional Reactivity. Preliminary internal cross-validation analyses support the possibility of replication of these findings in additional independent samples of 2 year olds with ASD. Given that symptom severity is often not uniform across domains, utilizing these factors will allow for a more nuanced understanding of a child's ASD symptoms, and in turn, will lead to more tailored recommendations for intervention.

**Acknowledgments** The authors thank the families who participated in the current study, the physicians who assisted by offering the screening study to their patients, and the research teams at the University of Connecticut and the University of Washington, especially Wendy Stone and Geraldine Dawson. Geraldine Dawson is now Professor in the Departments of Psychiatry and Behavioral Sciences, Pediatrics, and Psychology and Neuroscience and the Director of the Duke Center for Autism and Brain Development at Duke University, North Carolina. Wendy Stone is Professor of Psychology at The University of Washington. The authors also acknowledge the following funding sources: Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant R01HD039961, U.S. Department of Education Student-Initiated Research Grant, Maternal and Child Health Bureau Grant R40MC00270, the University of Connecticut's Research Foundation Faculty Grant, the National Alliance of Autism Research, and a National Institute of Mental Health Predoctoral Fellowship F31MH12550.

**Author Contributions** EM participated in conception of the study design, coordination and data collection, performed the statistical analyses and drafted sections of the manuscript; KB participated in conception of the study design, coordination and data collection, performed statistical analyses and drafted sections of the manuscript; MB participated in conception of the study design, data collection, interpretation of the data and the revising of the manuscript; DF participated in conception of the study design, data collection, interpretation of the data, and the revising of the manuscript. All authors read and approved the final manuscript.

**Funding** This study was funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant R01HD039961, U.S. Department of Education Student-Initiated Research Grant, Maternal and Child Health Bureau Grant R40MC00270, the University of Connecticut's Research Foundation Faculty Grant, the National Alliance of Autism Research, and a National Institute of Mental Health Predoctoral Fellowship F31MH12550.

## Compliance with Ethical standards

**Conflict of Interest** Deborah Fein, and Marianne Barton are co-owners of M-CHAT, LLC, which receives royalties from companies that incorporate the M-CHAT into commercial products. Data reported in the current paper is from the freely available paper version of the M-CHAT. Authors Emily Moulton and Kathryn Bradbury declare that they have no conflict of interest.

**Ethical Approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed Consent** Informed consent was obtained from all individual participants included in the study.

## References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th edn.). Washington: American Psychiatric Publishing.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th edn.). Arlington: American Psychiatric Publishing.



- Bartlett, M. S. (1954). A note on the multiplying factors for various chi square approximations. *Journal of the Royal Statistical Society, 16* (Series B), 296–298.
- Barton, M. L., Boorstein, H., Dumont-Mathieu, T., Herlihy, L. E., & Fein, D. (2012). Toddler ASD symptom interview (TASI). Self-published.
- Barton, Marianne, L., Robins, Diana, L., Jashar, D., Brennan, L., & Fein, D. (2013). Sensitivity and specificity of proposed DSM-5 criteria for autism spectrum disorder in toddlers. *Journal of Autism and Developmental Disorders, 43*(5), 1184–1195.
- Breibord, J., & Croudace, T. J. (2013). Reliability generalization for childhood autism rating scale. *Journal of Autism and Developmental Disorders, 43*, 2855–2865. doi:10.1007/s10803-013-1832-9.
- Carter, A., & Briggs-Gowan, M. (2006). *Manual for the infant-toddler social and emotional assessment (ITSEA), version 2*. San Antonio: Psychological Corporation, Harcourt Press.
- Cattell, R. B. (1966). The *Scree Test* for the number of factors. *Multivariate Behavioral Research, 1*(2), 245–276.
- Center (CDC). (2016). Prevalence and characteristics of autism spectrum disorder among children aged 8 years: Autism and developmental disabilities monitoring network, 11 Sites, United States, 2012. *Morbidity and Mortality Weekly Reports (MMWR), 65*(SS-3), 1–23.
- Charman, T., Jones, C. R. G., Pickles, A., Simonoff, E., Baird, G., & Happe', F. (2011). Defining the cognitive phenotype of autism. *Brain Research, 1380*(1943), 10–21. doi:10.1016/j.brainres.2010.10.075.
- Chawarska, K., Klin, A., Paul, R., Macari, S., & Volkmar, F. (2009). A prospective study of toddlers with ASD: Short-term diagnostic and cognitive outcomes. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 50*(10), 1235–1245. doi:10.1111/j.1469-7610.2009.02101.x.
- Chlebowski, C., Green, J. A., Barton, M. L., & Fein, D. (2010). Using the childhood autism rating scale to diagnose autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*(7), 787–799. doi:10.1007/s10803-009-0926-x.
- Cohen, I. L., Schmidt-Lackner, S., Romanczyk, R., & Sudhalter, V. (2003). The PDD behavior inventory: A rating scale for assessing response to intervention in children with pervasive developmental disorder. *Journal of Autism and Developmental Disorders, 33*(1), 31–45.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edn.). Hillsdale: Lawrence Earlbaum Associates.
- Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*(7), 1–9.
- Dawkins, T., Meyer, A. T., & Van Bourgondien, M. E. (2016). The relationship between the childhood autism rating scale: Second edition and clinical diagnosis utilizing the DSM-IV-TR and the DSM-5. *Journal of Autism and Developmental Disorders, 66*(7). doi:10.1007/s10803-016-2860-z.
- DiLalla, D. L., & Rogers, S. J. (1994). Domains of the childhood autism rating scale: Relevance for diagnosis and treatment. *Journal of Autism and Developmental Disorders, 24*(2), 115–128. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8040157>.
- Eaves, L. C., & Ho, H. H. (2004). The very early identification of autism: Outcome to age 4 1/2–5. *Journal of Autism and Developmental Disorders, 34*(4), 367–378. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15449513>.
- Garfin, D. G., & McCallon, D. (1988). Validity and reliability of the childhood autism rating scale with autistic adolescents. *Journal of Autism and Developmental Disorders, 18*(3), 367–378.
- Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39*(5), 693–705. doi:10.1007/s10803-008-0674-3.
- 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. doi:10.1007/s10803-006-0280-1.
- Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Early diagnosis of autism spectrum disorder: Stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 54*(5), 582–590. doi:10.1111/jcpp.12008.
- Harrop, C., McConachie, H., Emsley, R., Leadbitter, K., & Green, J. (2014). Restricted and repetitive behaviors in autism spectrum disorders and typical development: Cross-sectional and longitudinal comparisons. *Journal of Autism and Developmental Disorders, 44*(5), 1207–1219. doi:10.1007/s10803-013-1986-5.
- Hus, V., Gotham, K., & Lord, C. (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders, 44*, 2400–2412. doi:10.1007/s10803-012-1719-1.
- Johnson, C., & Myers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*(5), 1183–1215. doi:10.1542/peds.2007-2361.
- Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20*, 141–151.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31–36.
- Kerns, C. M., & Kendall, P. C. (2012). The presentation and classification of anxiety in autism spectrum disorder. *Clinical Psychology: Science and Practice, 19*(4), 323–347. doi:10.1111/cpsp.12009.
- Kim, S. H., Macari, S., Koller, J., & Chawarska, K. (2015). Examining the phenotypic heterogeneity of early autism spectrum disorder: Subtypes and short-term outcomes. *Journal of Child Psychology and Psychiatry, 57*(1), 93–102. doi:10.1111/jcpp.12448.
- Kleinman, J. M., Robins, D. L., Ventola, P. E., Pandey, J., Boorstein, H. C., Esser, E. L., ... Fein, D. (2008). The modified checklist for autism in toddlers: A follow-up study investigating the early detection of autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(5), 827–839. doi:10.1007/s10803-007-0450-9.
- Klin, A., Lang, J., Cicchetti, D. V., & Volkmar, F. R. (2000). Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV autism field trial. *Journal of Autism and Developmental Disorders, 30*(2), 163–167. doi:10.1023/A:1005415823867.
- Leekam, S., Tandos, J., McConachie, H., Meins, E., Parkinson, K., Wright, C., ... Le Couteur, A. (2007). Repetitive behaviours in typically developing 2-year-olds. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 48*(11), 1131–1138. doi:10.1111/j.1469-7610.2007.01778.x.
- Levy, S. E., Mandell, D. S., & Schultz, R. T. (2009). Autism. *Lancet, 374*(9701), 1627–1638. doi:10.1016/S0140-6736(09)61376-3.
- Lloyd, M., MacDonald, M., & Lord, C. (2013). Motor skills of toddlers with autism spectrum disorders. *Autism, 17*(2), 133–146. doi:10.1177/1362361311402230.
- Lord, C. (1995). Follow-up of two-year-olds referred for possible Autism. *Journal of Child Psychology and Psychiatry, 36*(8), 1365–1382.
- Lord, C., Luyster, R., Gotham, K., & Guthrie, W. (2012). *Autism diagnostic observation schedule, 2nd edition (ADOS-2) Manual (Part II): Toddler module*. Torrence: Western Psychological Services.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H. J., Leventhal, B. L., DiLavore, P. C., ... Rutter, M. (2000). The Autism diagnostic schedule-generic: A standard measures of social and

- communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223.
- Macdonald, M., Lord, C., & Ulrich, D. (2013). The relationship of motor skills and adaptive behavior skills in young children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 7, 1383–1390.
- Magyar, C. I., & Pandolfi, V. (2007). Factor structure evaluation of the childhood autism rating scale. *Journal of Autism and Developmental Disorders*, 37(9), 1787–1794. doi:10.1007/s10803-006-0313-9.
- Mayes, S. D., Calhoun, S. L., Murray, M. J., Pearl, A., Black, A., & Tierney, C. D. (2014). Research in autism spectrum disorders final DSM-5 under-identifies mild autism spectrum disorder: Agreement between the DSM-5, CARS, CASD, and clinical diagnoses. *Research in Autism Spectrum Disorders*, 8, 68–73.
- Mazefsky, C. A. (2015). Emotion regulation and emotional distress in autism spectrum disorder: Foundations and considerations for future research. *Journal of Autism and Developmental Disorders*, 45, 3405–3408.
- McDougle, C. J., Stigler, K. A., & Posey, D. J. (2003). Treatment of aggression in children and adolescents with autism and conduct disorder. *The Journal of Clinical Psychiatry*, 64(suppl 4), 16–25.
- Moulton, E., Barton, M., Robins, D. L., Abrams, D. N., & Fein, D. (2016). Early characteristics of children with ASD who demonstrate optimal progress between age two and four. *Journal of Autism and Developmental Disorders*. doi:10.1007/s10803-016-2745-1.
- Mullen, E. (1995). *Mullen scales of early learning*. Circle Pines: American Guidance Services.
- Osborne, J. W., & Fitzpatrick, D. C. (2012). Replication analysis in exploratory factor analysis: What it is and why it makes your analysis better. *Practical Assessment, Research and Evaluation*, 17(15), 1–8.
- Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., ... Iosif, A.-M. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*. doi:10.1111/jcpp.12421.
- Perry, A., Condillac, R. A., Freeman, N. L., Dunn-Geier, J., & Belair, J. (2005). Multi-site study of the childhood autism rating scale (CARS) in five clinical groups of young children. *Journal of Autism and Developmental Disorders*, 35(5), 625–634. doi:10.1007/s10803-005-0006-9.
- Pilowsky, T., Yirmiya, N., Shulman, C., & Dover, R. (1998). The autism diagnostic interview-revised and the childhood autism rating scale: Differences between diagnostic systems and comparison between genders. *Journal of Autism and Developmental Disorders*, 28(2), 143–151.
- Rellini, E., Tortolani, D., Trillo, S., Carbone, S., & Montecchi, F. (2004). Childhood autism rating scale (CARS) and autism behavior checklist (ABC) correspondence and conflicts with DSM-IV criteria in diagnosis of autism. *Journal of Autism and Developmental Disorders*, 34(6), 703–709.
- Reszka, S. S., Boyd, B. A., Mabee, M., Hume, K. A., & Odom, S. L. (2014). Brief report: Concurrent validity of autism symptom severity measures. *Journal of Autism and Developmental Disorders*, 44, 466–470. doi:10.1007/s10803-013-1879-7.
- Robins, D.L., Fein, D., & Barton, M. (1999). *The modified checklist for autism in toddlers (M-CHAT)*. Self-published.
- Robins, D.L., Fein, D., & Barton, M. (2009). *The modified checklist for autism in toddlers*. Revised, with follow-up (M-CHAT-R/F). Self-published.
- Rutter M., Le Couteur A., & Lord, C. (2003). *Autism diagnostic interview revised (ADI-R)*. Los Angeles: Western Psychological Services.
- Saemundsen, E., Magnusson, P., Smari, J., & Sigurdardottir, S. (2003). Autism diagnostic interview-revised and the childhood autism rating scale: Convergence and discrepancy in diagnosing autism. *Journal of Autism and Developmental Disorders*, 33(3), 319–328. doi:10.1023/A:1024410702242.
- Schopler, E., Reichler, R., & Rothen Renner, B. (1995). *Childhood autism rating scale (CARS)*. Los Angeles: Western Psychological Services.
- Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood autism rating scale (CARS). *Journal of Autism and Developmental Disorders*, 10(1), 91–103.
- Schopler, E., Van Bourgondien, M. E., Wellman, G. J., Love, S. R. (2010). *The childhood autism rating scale*, (2nd edn.). Los Angeles: Western Psychological Services.
- Sparrow, S., Balla, D., & Cicchetti, D. (1984). *The Vineland adaptive behavior scales*. Circle Pines: American Guidance Services.
- Sparrow, S., Cicchetti, D., & Balla, D. (2005). *Vineland-II adaptive behavior scales: Survey forms manual*. Circle Pines: AGS Publishing.
- Steiner, A. M., Goldsmith, T. R., Snow, A. V., & Chawarska, K. (2012). Practitioner's Guide to assessment of autism spectrum disorders in infants and toddlers. *Journal of Autism and Developmental Disorders*, 42, 1183–1196. doi:10.1007/s10803-011-1376-9.
- Stella, J., Mundy, P., & Tuchman, R. (1999). Social and nonsocial factors in the childhood autism rating scale. *Journal of Autism and Developmental Disorders*, 29(4), 307–317. doi:10.1023/A:1022111419409.
- Sutera, S., Pandey, J., Esser, E. L., Rosenthal, M. A., Wilson, L. B., Barton, M., ... Fein, D. (2007). Predictors of optimal outcome in toddlers diagnosed with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 37(1), 98–107. doi:10.1007/s10803-006-0340-6.
- Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications* (1st edn.). Washington: American Psychological Association.
- Turner, L. M., & Stone, W. L. (2007). Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 48(8), 793–802. doi:10.1111/j.1469-7610.2007.01744.x.
- Ventola, P. E., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J., ... Fein, D. (2006). Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders*, 36, 839–847. doi:10.1007/s10803-006-0128-8.
- Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 45(1), 135–170. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14959806>.
- Wolff, J. J., Botteron, K. N., Dager, S. R., Elison, J. T., Estes, A. M., Gu, H., ... Piven, J. (2014). Longitudinal patterns of repetitive behavior in toddlers with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 8, 945–953. doi:10.1111/jcpp.12207.
- Zwaigenbaum, L., Bauman, M. L., Fein, D., Pierce, K., Buie, T., Davis, P. A., ... Wagner, S. (2015). Early screening of autism spectrum disorder: Recommendations for practice and research. *Pediatrics*, 136, S41–S59.