Using the Childhood Autism Rating Scale to Diagnose Autism Spectrum Disorders

Colby Chlebowski, **James A. Green**, **Marianne L. Barton**, and **Deborah Fein**
Department of Psychology, University of Connecticut, 406 Babbidge Road, U-1020 Storrs, CT 06269-1020, USA

Colby Chlebowski: colby.chlebowski@uconn.edu

Abstract

This study investigated the childhood autism rating scale (CARS) as a tool for ASD diagnoses for 2-year-old ($n = 376$) and 4-year-old ($n = 230$) children referred for possible autism. The cut-off score to distinguish autistic disorder from PDD-NOS was 32 in the 2-year-old sample (consistent with Lord in J Child Psychol Psychiatry Allied Discipl, 36, 1365–1382, 1995), and 30 in the 4-year-old sample, with good sensitivity and specificity at both ages. The cut-off score to distinguish ASD from non-ASD at both ages was 25.5, with good sensitivity and specificity. Results confirm the utility of the CARS in distinguishing autistic disorder from PDD-NOS, and distinguishing ASD from other developmental disorders and typical development and suggest that an ASD cutoff around 25, which is in common clinical use, is valid.

Keywords

Introduction

Autism spectrum disorders (ASD) are neurodevelopmental disorders characterized by impaired social interactions, deficits in verbal and nonverbal communication, and repetitive behaviors or unusual or severely limited interests (American Psychiatric Association 2000). The conceptualization of autism as a spectrum disorder suggests that the disorder exists on a continuum of impairment, with autistic disorder representing the most severe presentation of the disorder. Pervasive developmental disorder, not otherwise specified (PDD-NOS) is often thought to represent the less severe end of a spectrum of autism severity and is sometimes loosely interchanged with the label of high functioning autism to denote a milder version of autism (Volkmar et al. 1994).

Despite ongoing theoretical debate regarding the categorical boundaries of PDD-NOS, no reliable diagnostic criteria have emerged (Matson and Boisjoli 2007). A diagnosis of PDD-NOS is commonly used as a diagnosis of exclusion and is often seen as a "catch all" diagnostic category used when criteria for other PDD diagnoses are not met (Filipek et al. 1999; Tidmarsh and Volkmar 2003). Clinically, PDD-NOS diagnoses are generally used for less severe cases that do not meet full criteria for autistic disorder or cases in which the required profile of diagnostic criteria is not present (e.g., an individual who does not have at least two symptoms in the social domain). PDD-NOS is also commonly used to describe

Correspondence to: Colby Chlebowski, `colby.chlebowski@uconn.edu`.

atypical symptoms of autism, cases in which onset is not prior to 30 months of age, and individuals who present with autism symptoms and comorbid disorders, such as ADHD (Buitelaar et al. 1999; Perry 1998). PDD-NOS is one of the most commonly diagnosed spectrum disorders; epidemiological data reported by Chakrabarti and Fombonne (2001) indicate that PDD-NOS diagnoses are at least twice as common as diagnoses of autistic disorder.

Although it is a common clinical diagnosis, the use of PDD-NOS as a sub-threshold category without clear diagnostic cutoffs makes reliable use of the diagnosis difficult. Data from the DSM-IV autism/PDD field trial suggests that the lack of clarity in diagnostic criteria for PDD-NOS influences diagnostic accuracy and reduces agreement among raters. Volkmar et al. (1994) reported strong agreement ($\kappa = 0.95$) in the differentiation of autism from non-spectrum diagnoses; however, agreement fell ($\kappa = 0.65$) when assessing the reliability of distinguishing between autistic disorder and PDD-NOS. Since specific diagnoses on the autism spectrum (e.g., autistic disorder and PDD-NOS) are less reliable than an ASD diagnosis, especially in young children, diagnostic measures that utilize an ASD cutoff may be more useful and reliable. Although clinical best practice in the diagnosis of ASD calls for the use of standardized measures as well as clinical judgment, the lack of an ASD cutoff on several widely used diagnostic instruments results in ASD diagnoses being made primarily from clinical judgment with minimal support from diagnostic instruments.

One widely used rating scale for the detection and diagnosis of autism is the childhood autism rating scale (CARS; Schopler et al. 1980, 1988). The CARS consists of 14 domains assessing behaviors associated with autism, with a 15th domain rating general impressions of autism. Each domain is scored on a scale ranging from one to four; higher scores are associated with a higher level of impairment. Total scores can range from a low of 15 to a high of 60; scores below 30 indicate that the individual is in the non-autistic range, scores between 30 and 36.5 indicate mild to moderate autism, and scores from 37 to 60 indicate severe autism (Schopler et al. 1988). The psychometrics of the CARS have been well documented (Schopler et al. 1988; Perry and Freeman 1996; Nordin et al. 1998; Tachimori et al. 2003).

The CARS demonstrates strong agreement with DSM-IV criteria for autistic disorder. In a sample of 274 preschool children, Perry et al. (2005) found an agreement rate of 88% between classifications made by the CARS and DSM-IV diagnostic criteria. A study of 54 children aged 18 months to 11 years with diagnoses of autistic disorder found complete agreement between the CARS and DSM-IV criteria (Rellini et al. 2004). Ventola et al. (2006) reported a high rate of chance corrected agreement ($\kappa = 0.691$, $p < 0.001$) between the CARS and clinical judgment based on DSM-IV TR criteria in a sample of toddlers referred for possible autism. The authors reported a sensitivity of 0.89 when the CARS was used to diagnose an ASD, as compared to clinical judgment, and 0.96 when used to diagnose autistic disorder (Ventola et al. 2006). Eaves and Milner (1993) reported a sensitivity of 0.98 when using the CARS for diagnoses of autistic disorder, with 47 of 48 individuals diagnosed with autistic disorder receiving a CARS score at or above the autism cutoff of 30.

Although it is highly sensitive, the CARS appears to over-diagnosis young children as having autism. Lord (1995) found that the CARS consistently classified non-autistic intellectually disabled children as having autism in a sample of 2-year-olds referred for possible autism. Lord (1995) reported that a CARS cutoff score of 30 correctly classified 61.5% of the non-autistic children and 93.7% of the children with autism; however, increasing the CARS autism cutoff to 32 improved classification and accurately classified

84.6% of the non-autistic children, while still correctly classifying 93.7% of the children with autism.

In addition to a relative lack of empirical testing of the cut-off of 30 for autistic disorder in toddlers and preschool children, another limitation of the CARS is the lack of an empirically based *ASD* cutoff. Although significant group differences on CARS total scores have been reported among clinical groups (Perry et al. 2005) the CARS is not designed to distinguish PDD-NOS from autistic disorder, or the ASD spectrum from non-spectrum (Perry et al. 2005). The lack of an ASD cutoff on the instrument reduces the diagnostic agreement among the CARS, other autism diagnostic instruments, and clinical judgment (Chlebowski et al. 2008).

When determining optimal cutoff scores for ASD on the CARS, levels of sensitivity and specificity will vary depending on how the instrument is being used (e.g., for screening or diagnosis). When screening for autism, the goal is to identify as many children with an ASD as possible and sensitivity should be high so that the measure misses few cases and does not falsely reassure parents that their children are not at risk (Charman and Baron-Cohen 2006). High sensitivity results in lower specificity since screening is erring on the side of producing more "false positives" than "false negatives." However, when using an instrument for diagnosis, specificity needs to increase in order to avoid inaccurately diagnosing children, unnecessarily worrying parents, and providing unnecessary referrals for costly intervention services. Filipek et al. (1999) recommend that instruments used in the diagnosis of ASD should have moderate sensitivity and good specificity for autism.

The only reported study assessing the use of a cutoff for pervasive developmental disorders (PDD) on the CARS used the Tokyo version of the CARS (CARS-TV) in a Japanese sample of 430 individuals ranging in age from 25 to 294 months. Tachimori et al. (2003) reported that a cutoff score of 25.5/26 on the CARS-TV distinguished individuals with PDD (including diagnoses of autistic disorder, PDD-NOS, Asperger's disorder, and childhood disintegrative disorder) from those with mental retardation without a history of PDD with sensitivity of 0.86 and specificity of 0.83. This cutoff had a high positive predictive value (0.97) but a low negative predictive value (0.50).

The current study was conducted to investigate the use of the CARS in a clinically referred sample of young children with ASD. There were three main objectives; the first was to replicate studies assessing the ideal CARS cutoff for a diagnosis of autistic disorder in samples of toddler and preschool aged children. The second objective was to calculate sensitivity, specificity, and positive and negative predictive values to determine the optimal CARS cutoff score for an ASD diagnosis. The third objective was to assess how the use of a CARS ASD cutoff influences rates of diagnostic agreement between the CARS and the ADOS, and between the CARS and clinical judgment based on DSM-IV criteria.

Method

Participants

Participants were 606 children (482 male and 124 female) who failed the modified checklist for autism in toddlers (M-CHAT; Robins et al. 2001) and a follow up telephone interview and received a developmental evaluation at the University of Connecticut. The children were seen for an initial evaluation at approximately age two and a follow up evaluation at approximately age four. For this study, the sample was divided into two groups according to age at evaluation; the two groups will herein be referred to as the 2-year-old sample and the 4-year-old sample. The 2-year-old sample was comprised of 376 children (296 male and 80 female) and ranged in age at time of evaluation from 21 to 30 months ($M = 26$, SD = 4.61).

The 4-year-old sample was comprised of 230 children (186 male and 44 female); age at reevaluation ranged from 42 to 66 months ($M = 54$, SD $= 11.78$). These samples were not independent; a subset of the children ($n = 173$) were seen for evaluations at both age two and age four and data from both evaluations are included in this study.

For the majority of the analyses, the children in each of the samples were divided into four groups based on their diagnosis at evaluation or reevaluation. The autistic disorder group consisted of children who met DSM-IV criteria for a diagnosis of autistic disorder, as per clinical best estimate, based on the autism diagnostic observation schedule (ADOS), autism diagnostic interview, revised (ADI-R), parent interview, and direct observation of the child (Klin et al. 2000). The PDD-NOS group consisted of children with diagnoses of PDD-NOS per clinical best estimate, based on ADOS, parent interview, and direct observation of the child. The non-ASD diagnosis (non-ASD) group consisted of children with diagnoses of intellectual disability, global developmental delay, developmental language disorder, or other DSM-IV diagnoses. The no diagnosis group consisted of children who did not meet criteria for any DSM-IV diagnoses, as well as children who were judged to be typically developing by the clinicians in the study (see Table 1).

Procedure

Participants were part of a large screening study and all children were screened using the M-CHAT at a pediatrician's or early intervention office between 16 and 30 months of age. All children in this study failed the M-CHAT as well as a follow up telephone interview and were offered a free evaluation at the University of Connecticut. Data for the current study were collected during the initial developmental evaluation and/or during a reevaluation occurring approximately 2 years later.

The evaluations took place at the Psychological Services Clinic at the University of Connecticut and were conducted by a licensed psychologist or developmental pediatrician and a doctoral student. All children seen for an evaluation in this study were administered the CARS, the ADOS, the ADI-R, the mullen scales of early learning (MSEL), and received diagnoses based on clinical judgment using DSM-IV-TR criteria.

Instruments

The modified checklist for autism in toddlers (M-CHAT; Robins et al. 2001) is a parent report checklist designed to screen for ASD in children 16–30 months of age. A positive screen is indicated by failing three of the 23 items, or two of six critical items. Critical items were determined by discriminant function analysis.

The childhood autism rating scale (CARS; Schopler et al. 1980, 1988) is a behavioral rating scale used for assessing the presence and severity of symptoms of autism spectrum disorders. The CARS was completed by the evaluators after the developmental evaluation was completed, and information from clinical observation, test measures, and parent report was used to assign CARS ratings. Both the licensed clinician and doctoral student completed the CARS; reliability statistics were calculated in a randomly selected subset of the sample ($n = 100$). The agreement between raters for CARS total scores was very large ($r(98) = 0.94$), according to criteria outlined by Cicchetti (1994). A kappa analysis used to assess the level of agreement between the two raters in regards to the overall CARS classification (e.g., autism or non-autism) produced a kappa of 0.90 ($p < 0.001$), indicating excellent agreement between raters (Cicchetti and Sparrow 1981). Inter-rater reliability analyses were also conducted independently for the 2-year-old and 4-year-old samples; for both samples the correlation coefficients were very large (Cicchetti 1994) and the kappa analyses indicated

excellent agreement (Cicchetti and Sparrow 1981). The CARS ratings made by the licensed clinician were used for the analyses in this study.

The autism diagnostic observation schedule (ADOS; Lord et al. 2000) is a semi-structured standardized assessment of communication, social interaction and play behaviors. The instrument consists of "presses," or planned social interactions, that are presented by a trained evaluator in order to encourage the child to initiate and respond to social interactions in a naturalistic setting. The ADOS has four modules corresponding to varying expressive language levels ranging from pre-verbal/single words to fluent speech. The ADOS algorithm provides diagnostic cutoffs for autistic disorder, ASD, and non-ASD.

The mullen scales of early learning (MSEL; Mullen 1995) is a measure of cognitive development for children up to 68 months of age that includes items that measure skills on five scales: gross motor, visual reception (nonverbal problem solving skills), fine motor, receptive language, and expressive language. The gross motor scale was not administered in this study. The MSEL was administered to all children in the study, though some children were untestable and their scores are therefore not included in the analyses. Scores from the MSEL were available for 333 2-year-olds (89% of the 2-year-old sample) and 182 4-year-olds (79% of the 4-year-old sample).

Clinical judgment by experienced clinicians is considered to be the "gold standard" for autism diagnosis (Spitzer and Siegel 1990; Klin et al. 2000). When diagnosing children in this study, clinicians used the DSM-IV-TR criteria for pervasive developmental disorders (APA 2000) as the basis for their clinical judgments and a diagnosis of an autism spectrum disorder was given if the licensed clinician determined that the child met the necessary diagnostic criteria.

Results

Analyses of Sample Characteristics

Two-Year-Old Sample—Of the 376 children who were evaluated at age two, 142 were in the autistic disorder group, 101 in the PDD-NOS group, 95 in the non-ASD group, and 38 were in the no diagnosis group. In the 2-year-old sample, there was a significant difference in age among the four diagnostic groups ($F(3, 372) = 6$, $p < 0.001$), with the no diagnosis group significantly younger than all other diagnostic groups. There were no significant differences among the diagnostic groups in gender ($\chi^2 (3, N = 376) = 3.19$, $p = 0.363$) or ethnicity ($\chi^2 (3, N = 376) = 13.7$, $p = 0.746$).

In 333 of the 376 2-year-old children, developmental level was measured by the MSEL. There were significant differences in the developmental quotient (DQ) among the four diagnostic groups ($F(3, 329) = 86.1$, $p < 0.001$); the Levene's test for the equality of variances was significant, indicating that the error variance of MSEL ELC scores were not assumed to be similar across diagnostic groups. Post hoc analyses (Dunnett's T3) revealed that mean DQ score of the no diagnosis group ($M = 95.6$, SD = 12.9) was significantly higher than all other groups. The DQ score of the non-ASD group ($M = 69.5$, SD = 17.2) did not differ significantly from the score of the PDD-NOS group ($M = 64$, SD = 14.6), though both scores were significantly higher than the DQ score of the autistic disorder group ($M = 58.7$, SD = 10.9) (see Table 1). However it should be noted that the developmental level of these three groups, while statistically different, are all in the very low range and do not represent clinically significant differences.

In order to assess the potential variability across domains on the MSEL, *t*-scores for the MSEL domains (visual reception, fine motor, receptive language and expressive language)

were compared across diagnostic groups. In the 2-year-old sample there were significant differences in the MSEL domain scores across diagnostic groups. This information can be found in Table 2.

Four-Year-Old Sample—Of the 230 children who were evaluated at age four, 104 were in the autistic disorder group, 44 in the PDD-NOS group, 34 in the non-ASD group, and 48 in the no diagnosis group. There were no significant differences in age ($F(3, 226) = 2.48$, $p = 0.062$) or ethnicity ($\chi^2 (3, N = 230) = 8.73$, $p = 0.891$) among the diagnostic groups in the 4-year-old sample. There were a higher proportion of males to females in the autistic disorder and non-ASD groups ($\chi^2 (3, N = 230) = 14.3$, $p < 0.01$).

Developmental level was measured in 182 of the 230 4-year-old children with the MSEL. DQ scores were significantly different across diagnostic groups ($F(3, 178) = 54.7$, $p < 0.001$). The DQ score of the no diagnosis group ($M = 98.9$, SD = 12.4) was significantly higher than the score of the PDD-NOS group ($M = 84.8$, SD = 18.2), which was significantly higher than the score of the non-ASD group ($M = 74.1$, SD = 19.5), which in turn was significantly higher than the DQ score of the autistic disorder group ($M = 59.4$, SD = 16.1) (see Table 1).

T-scores for the MSEL domains (visual reception, fine motor, receptive language and expressive language) were again compared across diagnostic groups. In the 4-year-old sample there were significant differences in the MSEL domain scores across diagnostic groups; see Table 2.

Internal Consistency

Cronbach's alpha was computed as a measure of internal consistency reliability for the CARS. Cronbach's alpha was 0.91 for the entire sample ($N = 606$), 0.90 for the 2-year-old sample ($n = 376$) and 0.93 for the 4-year-old sample ($n = 230$), indicating excellent levels of internal consistency (Cicchetti 1994).

CARS Scores and Cognitive Level

Pearson's correlation coefficients were calculated to examine the relationship between CARS scores and developmental quotient (DQ), as measured by the early learning composite (ELC) from the MSEL. In both the 2-year-old ($r(331) = -0.58$) and 4-year-old ($r(180) = -0.72$) samples there were strong negative correlations between DQ scores and CARS total scores, with large and very large effect sizes respectively (Cicchetti 1994). The large correlations between CARS scores and MSEL ELC scores may have been driven by between group differences and in order to assess for that influence, correlation analyses were run separately within groups (see Table 3 for correlations for the 2-year-old and 4-year-old samples). Between group differences did not appear to influence the correlations in the 2-year-old sample, but may have contributed to the lack of correlation in the No Diagnosis group in the 4-year-old sample due to restriction of range.

Comparison of CARS Scores by Age and Diagnostic Group

Mean CARS total scores were calculated for each diagnostic group. In order to assess whether differences in autism severity varied across diagnostic groups, a one-way ANOVA was conducted for each age group. In the 2-year-old sample, the analysis produced a main effect of diagnostic group ($F(3, 372) = 286.89$, $p < 0.001$), indicating that mean CARS scores differed significantly by diagnostic group. Post hoc comparisons (Tukey's LSD) indicated that the autistic disorder sample had the highest mean CARS score ($M = 35.1$, SD = 4.2), which was significantly higher than the mean CARS score of the PDD-NOS group ($M = 29$, SD = 4), which was significantly higher than the mean CARS score of the non-

ASD group (M = 22.5, SD = 3.2), which in turn was significantly higher than the mean CARS score of the no diagnosis group (M = 19.7, SD = 3.4).

Similar results were found in the 4-year-old sample. There was a main effect of diagnostic group (F(3, 226) = 216.37, p <0.001) and post hoc comparisons (Tukey's LSD) indicated that the autistic disorder sample had the highest mean CARS score (M = 34.2, SD = 5), which was significantly higher that the mean CARS score of the PDD-NOS group (M = 25.9, SD = 3.4), which was significantly higher than the mean CARS score of the non-ASD group (M = 21.7, SD = 3), which was significantly higher than the mean CARS score of the no diagnosis group (M = 17.8, SD = 2.2).

CARS Cutoff for Autistic Disorder Diagnosis

In order to assess the ideal CARS cutoff score for a diagnosis of autistic disorder in samples of toddler and pre-school children, sensitivity, specificity, positive predictive value, and negative predictive value were calculated for cutoffs distinguishing an autistic disorder diagnosis from a PDD-NOS diagnosis. Positive predictive value refers to the proportion of children who were correctly classified as receiving a diagnosis of autistic disorder by the CARS cutoff and negative predictive value refers to the proportion of children who are correctly classified as not having autistic disorder (which includes children with PDD-NOS).

Using a CARS cutoff of 30 in a sample of 2-year-old children with autistic disorder or PDD-NOS diagnoses (n = 243) produced high sensitivity (0.93) but low specificity (0.49). A cutoff of 30 correctly classified 181 of 243 children and accurately classified 93% of the children with autistic disorder; however, it incorrectly diagnosed 52 children with PDD-NOS as having autism and only accurately classified 49% of children with PDD-NOS diagnoses. This result is consistent with the findings of Lord (1995) who found that a CARS cutoff of 30 produced many false positives and over-diagnosed 2-year-olds with autistic disorder. Using a cutoff of 32, as proposed by Lord (1995) reduced sensitivity (0.79) to a level defined as fair (Cicchetti et al. 1995) but dramatically increased specificity (0.81) to a level defined as good and a level acceptable for diagnosis (Cicchetti et al. 1995). A cutoff of 32 correctly classified 79% of the children with autism and 81% of PDD-NOS children (Fig. 1; Table 4). The proposed cutoff score of 32 for an autistic disorder diagnosis in the 2 year old sample was also examined separately for the children aged 24–30 months (n = 172) and those 21–23 months of age (n = 71). The sensitivity of a cutoff score of 32 was slightly lower in the sample of children under 24 months of age (0.73) as compared to the sensitivity in the children aged 24 months and above (0.81) though specificity did not differ between the two age groups.

In a sample of 4-year-old children with autistic disorder or PDD-NOS diagnoses (n = 148), the CARS cutoff of 30 produced sensitivity (0.86) and specificity (0.80) at level defined as good (Cicchetti et al. 1995) and accurately classified 122 children (86% of the autistic disorder sample and 80% of the PDD-NOS sample). Lowering the cutoff to 29 reduced specificity (0.73) without an increase in sensitivity; raising the cutoff to 31 lowered sensitivity to 0.76, raised specificity to 0.89, and reduced the percentage of autism children accurately classified by the CARS to 76% (see Fig. 1; Table 5).

CARS Cutoff for Autistic Spectrum Disorder Diagnoses

Sensitivity, specificity, positive predictive values, and negative predictive values were calculated for potential cutoff scores to distinguish an ASD diagnosis (e.g., diagnoses of autistic disorder or PDD-NOS) from children with non-ASD diagnoses or no diagnoses. In the sample of 376 2-year-old children, a cutoff score of 25 produced a sensitivity of 0.93 and a specificity of 0.85 and accurately classified 339 of 376 children. A cutoff of 26 increased

specificity (0.91) and accurately classified four additional children; however, sensitivity was reduced and four children with ASD diagnoses were classified as false negatives. Using a midpoint cutoff of 25.5 produced a sensitivity of 0.92, which is defined as excellent by clinical criteria, and a specificity of 0.89, which is defined as good (Cicchetti et al. 1995). This cutoff correctly classified 92% of the ASD sample and 89% of the non-ASD sample (see Fig. 2; Table 6). Cutoffs for an ASD diagnosis were again examined separately for the children aged 24–30 months ($n = 256$) and those 21–23 months of age ($n = 120$) and there were no significant differences between the sensitivity and specificity of the ASD cutoffs between the two groups.

Cutoffs for distinguishing ASD from non-ASD diagnoses were also assessed in the 4-year-old sample ($N = 230$). A cutoff score of 25 accurately classified the largest number of children (201 out of 230, 84% of the ASD children and 93% of the non-ASD children) with a sensitivity of 0.84 and a specificity of 0.93. Using a midpoint score of 25.5 decreased sensitivity to 0.82 and increased specificity to 0.95, resulting in an increase in the accurate classification of non-ASD children (see Fig. 2; Table 7).

Influence of the CARS Cutoff on Diagnostic Agreement

To assess the influence of a CARS ASD cutoff on rates of agreement among autism diagnostic instruments and clinical judgment, kappa analyses were conducted to assess agreement for ASD diagnoses based on DSM-IV criteria, the ADOS, and the CARS using the traditional CARS cutoff of 30 and the proposed ASD cutoff score of 25.5. Kappa analyses were conducted with sub-samples of the original 2-year-old and 4-year old samples ($n = 354$ and $n = 190$, respectively); children were excluded if they did not have sufficient ADOS data available. Levels of clinical significance are defined by Cicchetti and Sparrow's (1981) criteria.

Using the CARS cutoff of 30 for the children in the 2-year-old sample, there was 76% agreement and 268 cases of agreement between ASD diagnoses made by the CARS and diagnoses based on DSM-IV criteria. In the sample, 155 children received ASD diagnoses from both the CARS and clinical judgment based on DSM-IV criteria, 5 children received diagnoses from the CARS but did not receive a diagnosis based on DSM-IV criteria, 81 children met DSM-IV criteria for ASD diagnoses but did not meet criteria on the CARS, and 113 children did not receive a diagnosis from the CARS or clinical judgment. Kappa analyses revealed fair agreement between the CARS and DSM-IV criteria ($\kappa = 0.57$, $p < 0.001$) using a cutoff score of 30 on the CARS.

There was 76% agreement between ASD diagnoses made by the CARS using the cutoff score of 30 and ASD diagnoses made by the ADOS. Of the 354 children who received both the CARS and the ADOS, 155 received ASD diagnoses on both measures, 5 children met criteria for ASD on the CARS but did not meet on the ADOS, 79 children met criteria on the ADOS but did not receive diagnoses of ASD based on the CARS, and 115 were classified as non-ASD by both instruments. Kappa analyses revealed good agreement between the CARS and the ADOS ($\kappa = 0.60$, $p < 0.001$).

Using an ASD cutoff score of 25.5 on the CARS produced the highest level of agreement among diagnostic instruments and clinical judgment for the 2-year-old sample. Agreement between ASD diagnoses made by the CARS and those based on DSM-IV criteria increased to 88% agreement; 208 children received ASD diagnoses from both the CARS and clinical judgment based on DSM-IV criteria, 13 children received diagnoses from the CARS but did not receive a diagnosis based on DSM-IV criteria, 28 children received a diagnosis based on DSM-IV criteria but did not meet criteria for ASD on the CARS, and 105 children did not receive a diagnosis from the CARS or clinical judgment. Chance corrected agreement

increased to excellent between the CARS and diagnoses based on DSM-IV criteria (κ = 0.75, p <0.001).

Agreement between the ADOS and the CARS increased to 86% agreement and 203 children received ASD diagnoses on both measures, 18 children met criteria for ASD on the CARS but not on the ADOS, 31 children met criteria on the ADOS but did not receive diagnoses of ASD on the CARS, and 102 were classified as non-ASD by both instruments. Kappa analyses were good (κ = 0.70, p <0.001).

In the 4-year-old sample, there were 145 cases of agreement with 76% agreement between ASD diagnoses made by the CARS using an autism cutoff score of 30 and those based on DSM-IV criteria. 83 children received ASD diagnoses from both the CARS and clinical judgment based on DSM-IV criteria, 45 children received a diagnosis based on DSM-IV criteria but did not meet criteria for ASD on the CARS, and 62 children did not receive a diagnosis from the CARS or clinical judgment.

ASD diagnoses made using a cutoff score of 30 on the CARS and the ADOS had 82% agreement, 82 met ASD criteria on both instruments, 1 child met criteria for ASD on the CARS but did not meet criteria on the ADOS, 34 children received ASD diagnoses on the ADOS but did not receive diagnoses of ASD on the CARS, and 73 were classified as non-ASD by both instruments. Kappa analyses revealed fair agreement between the CARS and DSM-IV criteria diagnoses (κ = 0.55, p <0.001). Kappa analyses revealed good agreement between the CARS and the ADOS (κ = 0.64, p <0.001).

Using an ASD cutoff score of 25.5 on the CARS increased agreement to 86% between the CARS and DSM-IV ASD diagnoses and resulted in 163 cases of agreement. 104 children received ASD diagnoses from both the CARS and clinical judgment based on DSM-IV criteria, 3 children received diagnoses from the CARS but did not receive a diagnosis based on DSM-IV criteria, 24 children received a diagnosis based on DSM-IV criteria but did not meet criteria for ASD on the CARS, and 59 children did not receive a diagnosis from the CARS or clinical judgment. Kappa analyses of chance corrected agreement increased to good between the CARS and diagnoses based on DSM-IV criteria (κ = 0.74, p <0.001).

Agreement between the ADOS and the CARS increased to 87% agreement with the use of an ASD cutoff of 25.5 in the 4-year-old sample; there were 165 cases of agreement, 99 received ASD diagnoses on both measures, 8 children met criteria for ASD on the CARS but not on the ADOS, 17 children met criteria on the ADOS but did not receive diagnoses of ASD on the CARS, and 66 were classified as non-ASD by both instruments. Kappa analyses were good (κ = 0.73, p <0.001).

Discussion

The purpose of the current study was to investigate the use of the CARS in large samples of toddlers and preschool aged children referred for an evaluation for ASD. The study aimed to identify appropriate CARS cutoffs for diagnoses of autistic disorder and PDD-NOS and assess the influence of an ASD cutoff on CARS diagnostic agreement with the ADOS and clinical judgment. All of the children in the current study were evaluated using the CARS along with validated autism diagnostic measures and received diagnoses from clinical judgment based on DSM-IV TR criteria. The strong internal consistency and inter-rater reliability of the CARS reported in this study is consistent with statistics reported by the instrument's authors and previous studies (Schopler et al. 1988; Saemundsen et al. 2003; Tachimori et al. 2003) and supports the use of the CARS with young children.

Results of the current study support previous findings indicating that the CARS total score differs significantly by diagnostic group, with children diagnosed with autistic disorder having significantly higher total CARS scores than children with PDD-NOS, who in turn had significantly higher scores than those with non-ASD diagnoses. The clinically significant differences reported in CARS total scores among diagnostic groups are consistent with the findings of Tachimori et al. (2003) and Perry et al. (2005) and support the use of the CARS as a reliable measure of autism severity. In this study, the distribution of CARS total scores across the autism spectrum did not vary by age, suggesting that the CARS represents a comparable measure of autism severity for both toddler and preschool aged children.

In the 2-year-old sample, the CARS cutoff of 30 for Autistic Disorder had high sensitivity; however, the specificity was not high enough to reliably diagnose autism with appropriate specificity and incorrectly classified approximately half of the children with diagnoses of PDD-NOS as having autistic disorder. These results are consistent with previous findings of Lord (1995) and provide support for raising the cutoff score for autism from 30 to 32 in 2-year-old clinical samples. In the 4-year-old sample, the CARS cutoff of 30 for autism diagnoses produced ideal sensitivity and specificity and these findings support the continued use of the cutoff score of 30 suggested by Schopler et al. (1988) for diagnosing autism in 4-year-old children.

The results regarding the ASD cutoff in both the 2-year-old and 4-year-old sample are consistent with the cutoff of 25.5/26 proposed by Tachimori et al. (2003). In the 2-year-old sample, a cutoff of 25.5 (e.g., a score of 25.5 or higher) produced a sensitivity of 0.92 and a specificity of 0.89. Utilizing the 25.5 cutoff in the 4-year-old sample produced adequate sensitivity (0.82) and high specificity (0.95) suggesting that an ASD cutoff of 25.5 works effectively in a preschool aged sample as well. These findings indicate that any CARS score over 25 (e.g., 25.5 and above) is consistent with ASD and supports the recommendation of utilizing a score of 25.5 as the CARS cutoff for an ASD diagnosis.

Research indicates that diagnoses of autistic disorder are valid and stable when made at age two (Lord 1995; Stone et al. 1999; Kleinman et al 2008), and a growing literature suggests that diagnoses of autistic disorder made even before age two are generally stable over time (Adrien et al. 1992; Baron-Cohen et al. 1992). Early diagnoses of an ASD, such as PDD-NOS, have been found to be less stable; however, the majority of children with a PDD-NOS diagnosis at age two remain on the spectrum at later follow up (Lord 1995; Stone et al. 1999; Chlebowski et al. 2009). Despite the growing evidence of diagnostic stability, providing reliable diagnoses of ASD can be difficult in toddlers. Research suggesting that validated measures, such as the ADI-R, do not work reliably in 2-year-old children highlights this difficulty (Lord et al. 1993; Lord 1995; Ventola et al. 2006). The utilization of the modified CARS cutoff score of 32 for autism and the proposed ASD cutoff of 25.5 in samples of referred toddlers can add to the accuracy and validity of early ASD diagnoses.

The difficulty in differentiating autistic disorder from other pervasive developmental disorders that has been consistently reported in the literature is not surprising considering that the diagnostic category of PDD includes a wide spectrum of diagnoses (de Bildt et al. 2004) and the fact that, due to variation in clinical features and severity of presentation, individuals with PDD-NOS have been notoriously difficult to categorize as a homogenous group (Buitelaar et al. 1999, Luteijn et al. 2000). Despite this difficulty, the specificity of the cutoffs to distinguish autistic disorder from PDD-NOS in both the 2-year-old and 4-year-old samples in the current study is high enough to assist with accurate differentiation, which indicates that the CARS cutoff scores will serve as a useful aide in accurately differentiating autistic disorder from PDD-NOS for toddlers and preschool aged children in clinical settings.

When using the CARS as a diagnostic instrument in a clinically referred sample, a cutoff for ASD improves diagnostic agreement among the CARS, a reliable autism diagnostic measure (e.g., ADOS), and clinical judgment for both 2-year-old and 4-year-old samples. The increase in diagnostic agreement also improves overall diagnostic accuracy because, as Risi et al. (2006) note, consistency across diagnostic instruments and correspondence between the diagnostic instrument and clinical judgment simplifies the diagnostic process and increases the accuracy of the diagnosis.

Limitations

There are several limitations to the current study. The sample in this study was a clinically referred sample of children who were at greater risk of receiving an ASD diagnosis based on a positive screen on an autism specific screener. Even the children who were placed in the non-ASD and no diagnosis groups initially screened positive on the autism screener, suggesting that they exhibited some of the behaviors consistent with an ASD diagnosis and therefore, there was no normal control group for the study. Additionally, the majority of the sample was Caucasian (74%) and male (80%), which limits the generalizability of these findings to female samples or ethnic minorities.

Additionally, the CARS has been validated for use with children aged 24 months and older and the use of the CARS in younger samples has not been adequately studied. In the 2-year-old sample in this study, participants' ages ranged from 21 to 30 months and 120 of the 376 children (32% of the 2-year-old sample) were under 24 months of age, which may have influenced the results from the CARS in the 2-year-old sample; however, as noted previously the results in the 2-year-old sample were largely consistent between children 24–30 months of age and those under 24 months of age.

It is important to acknowledge is that the CARS is not a standardized measure and CARS score are not made independently of clinical judgment. The use of clinical judgment is an important component of scoring the CARS and therefore influences the outcome scores on the measure. In this study, clinical diagnoses were not made independently from CARS scores; the clinician who completed the CARS also made the diagnosis based on DSM-IV criteria, which likely inflated the relationship between CARS scores and clinical diagnoses. However, the excellent inter-rater reliability between independent raters suggests that the CARS score reflects a consensus of clinical judgment and not simply a reflection of the judgment of an individual clinician.

Perhaps the most significant caveat of the CARS, and a limitation to the results in this study, is the fact that CARS ratings are only as good as the behavior sample upon which they are based. In this study, CARS rating were based on information from a parent interview regarding developmental history, results from autism diagnostic instruments, and clinical observation over a 3-h evaluation, which allowed for well informed ratings incorporating several sources of information. Additionally, CARS ratings were made by licensed professionals with extensive experience with autism spectrum disorders. Although the CARS is promoted for use in a variety of setting with professionals with varying level of training, the CARS ratings in this study may not be commensurate with ratings made by less experienced clinicians (e.g., professional with little experience with autism) or in settings that do not allow for extensive behavioral observations (e.g., when used during well-child visits). That it, one cannot assume that the CARS will operate in this same way unless all CARS ratings are based on similar samples of behavior used by raters with a comparable level of training.

Despite these limitations, the results presented here support the use of the CARS ASD cutoff in both 2-year-old and 4-year-old children referred for ASD. The use of an ASD cutoff is

encouraged in order to improve diagnostic accuracy and agreement with other diagnostic instruments.

Acknowledgments

References

Adrien J, Perrot A, Sauvage D, Leddet I, Larmande C, Barthelemy C. Early symptoms in autism from family home movies. Evaluation and comparison between 1st and 2nd year of life using I.B.S.E. scale. Acta Paedopsychiatrica. 1992; 55(2):71–75. [PubMed: 1374996]

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4. Washington, DC: American Psychiatric Association; 2000. (Text revision)

Baron-Cohen S, Allen J, Gillberg C. Can autism be detected at 18 months? The needle, the haystack, and the CHAT. The British Journal of Psychiatry. 1992; 161(6):839–843. [PubMed: 1483172]

Buitelaar J, Van der Gaag R, Klin A, Volkmar F. Exploring the boundaries of pervasive developmental disorder not otherwise specified: Analyses of data from the DSM-IV autistic disorder field trial. Journal of Autism and Developmental Disorders. 1999; 29:33–43. [PubMed: 10097993]

Chakrabarti S, Fombonne E. Pervasive developmental disorders in preschool children. Journal of the American Medical Association. 2001; 285:3093–3099. [PubMed: 11427137]

Charman, T.; Baron-Cohen, S. Screening for autism spectrum disorders in populations: Progress, challenges, and questions for future research and practice. In: Charman, T.; Stone, W., editors. Social and communication development in autism spectrum disorders. New York: Guilford Press; 2006. p. 63-82.

Chlebowski, C.; Barton, M.; Hodgson, S.; Fein, D. Does the CARS cutoff score influence agreement among diagnostic instruments for autism spectrum disorders and clinical judgment in preschool aged children?; Poster presented at the meeting of the American Academy of Clinical Neuropsychology; Boston, MA. 2008.

Chlebowski, C.; Barton, M.; Hodgson, S.; Fein, D. Diagnostic stability of autism spectrum disorder diagnoses made before age 2. Poster presented at the International Meeting for Autism Research; Chicago, Illinois. 2009.

Cicchetti D. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 1994; 6:284–290.

Cicchetti D, Sparrow S. Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. American Journal of Mental Deficiency. 1981; 86:127–137. [PubMed: 7315877]

Cicchetti DV, Volkmar F, Klin A, Showalter D. Diagnosing autism using ICD-10 criteria: A comparison of neural networks and standard multivariate procedures. Child Neuropsychology. 1995; 1:26–37.

de Bildt A, Sytema S, Ketelaars C, Kraijer D, Mulder E, Volkmar F, et al. Interrelationship between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. Journal of Autism and Developmental Disorders. 2004; 34:129–137. [PubMed: 15162932]

Eaves R, Milner B. The criterion-related validity of the childhood autism rating scale and the autism behavior checklist. Journal of Abnormal Child Psychology. 1993; 21:481–491. [PubMed: 8294649]

Filipek P, Accardo P, Baranek G, Cook E, Dawson G, Gordon B, et al. The screening and diagnosis of autistic spectrum disorders. Journal of Autism and Developmental Disorders. 1999; 29:437–482.

Kleinman J, Ventola P, Pandey J, Verbalis A, Barton M, Hodgson S, et al. Diagnostic stability in very young children with autism spectrum disorders. Journal of Autism and Developmental Disorders. 2008; 38:606–615. [PubMed: 17924183]

Klin A, Lang J, Cicchetti D, Volkmar F. Brief report: Interrater reliability of clinical diagnosis and DSM IV criteria for autistic disorder: Results of the DSM IV Field Trial. Journal of Autism and Developmental Disorders. 2000; 30:163–167. [PubMed: 10832781]

Lord C. Follow-up of two-year-olds referred for possible autism. Journal of Child Psychology and Psychiatry and Allied Disciplines. 1995; 36:1365–1382.

Lord C, Risi S, Lambrecht L, Cook EHJ, Leventhal BL, DiLavore PC, et al. The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders. 2000; 30:205–223. [PubMed: 11055457]

Lord C, Storoschuk S, Rutter M, Pickles A. Using the ADI-R to diagnose autism in preschool children. Infant Mental Health Journal. 1993; 14:234–252.

Luteijn E, Serra M, Jackson S, Steenhuis M, Althaus M, Volkmar F, et al. How unspecified are disorders of children with a pervasive developmental disorder not otherwise specified? A study of social problems in children with PDD-NOS and ADHD. European Child and Adolescent Psychiatry. 2000; 9:168–179. [PubMed: 11095039]

Matson J, Boisjoli J. Differential diagnosis of PDD-NOS in children. Research in Autism Spectrum Disorders. 2007; 1:75–84.

Mullen, EM. Mullen scales of early learning: AGS edition. Circle Pines, MN: American Guidance Service; 1995.

Nordin V, Gillberg C, Nydén A. The Swedish version of the childhood autism rating scale in a clinical setting. Journal of Autism and Developmental Disorders. 1998; 28:69–75. [PubMed: 9546304]

Perry R. Misdiagnosed ADD/ADHD: Rediagnosed PDD. Journal of the American Academy of Child & Adolescent Psychiatry. 1998; 37:113–114. [PubMed: 9444907]

Perry A, Condillac R, Freeman N, Dunn-Geier J, Belair J. Multi-site study of the childhood autism rating scale (CARS) in five clinical groups of young children. Journal of Autism and Developmental Disorders. 2005; 35:625–634. [PubMed: 16172810]

Perry, A.; Freeman, N. Psychometric properties of the Childhood Autism Rating Scale (CARS). Poster presented at the Ontario Association on Developmental Disabilities Research Special Interest Group Research Day; Toronto, ON. 1996.

Rellini E, Tortolani D, Trillo S, Carbone S, Montecchi F. Childhood autism rating scale (CARS) correspondence and conflicts with DSM-IV criteria in diagnosis of autism. Journal of Autism and Developmental Disorders. 2004; 34:703–708. [PubMed: 15679189]

Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. Journal of the American Academy of Child & Adolescent Psychiatry. 2006; 45:1094–1103. [PubMed: 16926617]

Robins D, Fein D, Barton M, Green J. The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. Journal of Autism and Developmental Disorders. 2001; 31:131–144. [PubMed: 11450812]

Saemundsen E, Magnusson P, Smari J, Sigurdarddottir S. Autism diagnostic interview—revised and the childhood autism rating scale: Convergence and discrepancy in diagnosing autism. Journal of Autism and Developmental Disorders. 2003; 33:319–328. [PubMed: 12908834]

Schopler E, Reichler R, DeVellis R. Toward objective classification of childhood autism: Childhood autism rating scale (CARS). Journal of Autism and Developmental Disorders. 1980; 10:91–103. [PubMed: 6927682]

Schopler, E.; Reichler, R.; Rochen Renner, B. The childhood autism rating scale. Western Psychological Services; 1988.

Spitzer R, Siegel B. The DSM-III-R field trial of pervasive developmental disorders. Journal of the American Academy of Child & Adolescent Psychiatry. 1990; 29:855–862. [PubMed: 2273011]

Stone W, Lee F, Ashford L, Brisie J, Hepburn S, Coonrod E. Can autism be diagnosed accurately in children under three? Journal of Child Psychiatry Psychology and Allied Disciplines. 1999; 40:219–226.

Tachimori H, Osada H, Kurita H. Childhood autism rating scale-Tokyo version for screening pervasive developmental disorders. Psychiatry and Clinical Neuroscience. 2003; 57:113–118.

Tidmarsh L, Volkmar F. Diagnosis and epidemiology of autism spectrum disorders. The Canadian Journal of Psychiatry/La Revue canadienne de psychiatrie. 2003; 48:517–525.

Ventola P, Kleinman J, Pandey J, Barton M, Allen S, Green J, et al. Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. Journal of Autism and Developmental Disorders. 2006; 36:839–847. [PubMed: 16897398]

Volkmar F, Klin A, Siegel B, Szatmari P. Field trial for autistic disorder in DSM-IV. American Journal of Psychiatry. 1994; 151:1361–1367. [PubMed: 8067493]
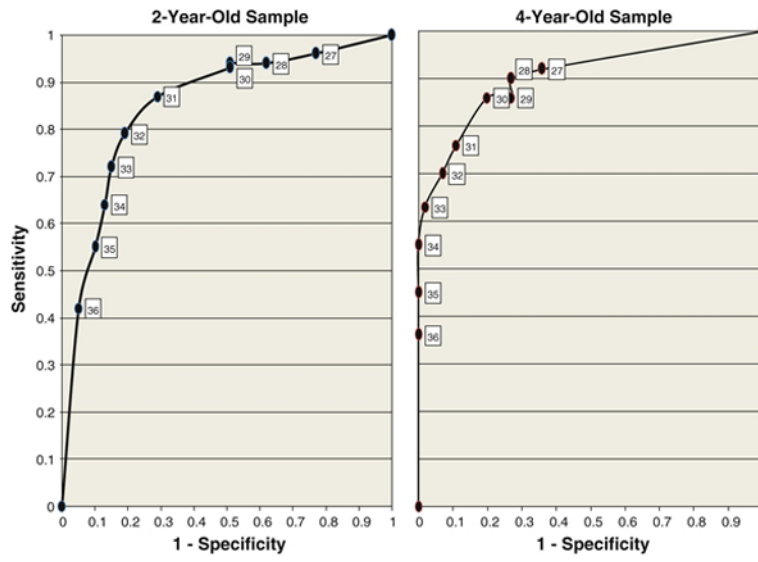
Fig. 1.
ROC curves for cutoffs for an autistic disorder diagnosis in both age samples
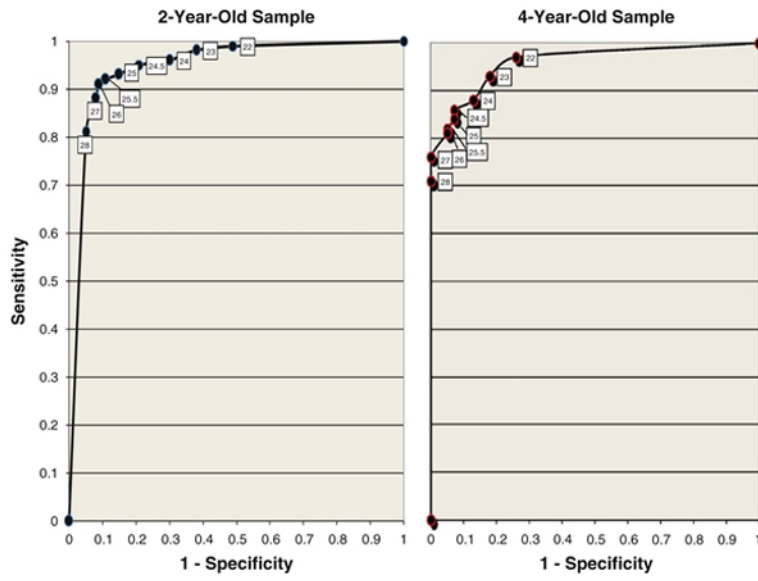
Fig. 2.
ROC curves for cutoffs for an ASD diagnosis in both age samples

Table 1

Sample characteristics

Diagnostic group	Sample characteristics	2-year-old sample $n = 376$	4-year-old sample $n = 230$
Autistic disorder	Sample size	$n = 142$	$n = 104$
	Mean age (SD)	26.8 (4.7)	53 (10.3)
	Gender	78% male	86% male
	Mean CARS score (SD)	35.1 (4.2)	34.2 (5)
	MSEL ELC (SD)[a]	58.7 (10.9)	59.4 (16.1)
	Mean ADOS cutoff score (SD)[b]	16.57 (4.12)	15.46 (4.34)
PDD-NOS	Sample size	$n = 101$	$n = 44$
	Mean age (SD)	25.9 (4.4)	52.7 (9.1)
	Gender	78% male	75% male
	Mean CARS score (SD)	29 (4)	25.9 (3.4)
	MSEL ELC (SD)[a]	63.2 (11)	84.8 (18.2)
	Mean ADOS cutoff score (SD)[b]	12.97 (4.66)	9.71 (3.42)
Non-ASD diagnosis	Sample size	$n = 95$	$n = 34$
	Mean age (SD)	26.6 (4.5)	53.4 (6.6)
	Gender	84% male	88% male
	Mean CARS score (SD)	22.5 (3.2)	21.7 (3)
	MSEL ELC (SD)[a]	68.4 (14.7)	74.1 (19.5)
	Mean ADOS cutoff score (SD)[b]	4.57 (3.79)	3.36 (2.33)
No diagnosis	Sample size	$n = 38$	$n = 48$
	Mean age (SD)	23.4 (4.1)	58.1 (17.9)
	Gender	71% male	65% male
	Mean CARS score (SD)	19.7 (3.4)	17.8 (2.2)
	MSEL ELC (SD)[a]	95.6 (12.9)	98.9 (12.4)
	Mean ADOS cutoff score (SD)[b]	4.11 (3.75)	1.78 (2.11)

[a]MSEL ELC results from a sub-set of children who received the mullen scales of early learning. (2-year-old sample: autistic disorder $n = 120$; PDD-NOS $n = 94$; Non ASD $n = 84$; no diagnosis $n = 35$); (4-year-old sample: autistic disorder $n = 82$; PDD-NOS $n = 35$; Non ASD $n = 28$; No diagnosis $n = 37$)

[b]ADOS results from a sub-set of children who received the autism diagnostic observation schedule. (2-year-old sample: autistic disorder $n = 130$; PDD-NOS $n = 97$; Non ASD $n = 91$; No diagnosis $n = 36$; (4-year-old sample: autistic disorder $n = 91$; PDD-NOS $n = 38$; Non ASD $n = 25$; No diagnosis $n = 36$)

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 2

T-scores for MSEL domains by diagnostic group in both age samples

	Autistic disorder	PDD-NOS	Non-ASD diagnosis	No diagnosis	F	p
	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)		
2-year-old sample						
Visual reception	29.08 (9.79)	32.30 (9.98)	35.06 (11.48)	49.06 (7.89)	36.25	<0.001
Fine motor	29.84 (11.68)	32.63 (10.84)	34.42 (12.58)	45.94 (8.53)	18.35	<0.001
Receptive language	22.27 (6.35)	25.04 (8.53)	30.30 (11.98)	47.26 (11.78)	70.65	<0.001
Expressive language	25.25 (8.64)	27.35 (7.61)	29.37 (8.11)	48.54 (9.80)	72.74	<0.001
4-year-old sample						
Visual reception	27.63 (13.89)	48.29 (15.70)	36.32 (15.72)	52.68 (10.45)	32.25	<0.001
Fine motor	26.34 (10.68)	38.69 (14.60)	33.93 (13.41)	47.81 (9.82)	30.19	<0.001
Receptive language	25.28 (10.27)	40.66 (11.84)	32.89 (10.68)	50.54 7.88)	56.97	<0.001
Expressive language	24.88 (9.03)	38.83 (9.85)	33.36 (10.30)	45.19 (8.57)	46.59	<0.001

Table 3

Correlations between CARS scores and MSEL ELC scores by diagnostic group in both age samples

	Two-year-old sample		Four-year-old sample	
	n	*r*	*n*	*r*
Total sample	333	$r(331) = -0.58$	192	$r(190) = -0.72$
Autistic disorder	120	$r(118) = -0.53$	82	$r(80) = -0.56$
PDD-NOS	94	$r(92) = -0.32$	35	$r(33) = -0.54$
NON-ASD	84	$r(82) = -0.26$	28	$r(26) = -0.50$
No diagnosis	35	$r(33) = -0.57$	37	$r(35) = 0.26$

Table 4

Sensitivity, specificity, positive predictive values and negative predictive values of CARS cutoffs for diagnosis of autistic disorder in a 2-year-old sample ($N = 243$)

CARS cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value	False positive errors[a]	False negative errors[a]
27	0.96	0.23	0.64	0.82	78	5
27.5	0.96	0.30	0.66	0.83	71	6
28	0.94	0.38	0.68	0.83	63	8
28.5	0.94	0.44	0.70	0.85	57	8
29	0.94	0.49	0.72	0.86	52	8
29.5	0.94	0.51	0.73	0.85	49	9
30	0.93	0.49	0.72	0.83	52	10
30.5	0.90	0.67	0.80	0.83	33	14
31	0.87	0.71	0.81	0.79	29	19
31.5	0.84	0.76	0.83	0.77	24	23
32	0.79	0.81	0.85	0.73	19	30
32.5	0.76	0.83	0.86	0.71	17	34
33	0.72	0.85	0.87	0.68	15	40
33.5	0.68	0.86	0.87	0.66	14	45
34	0.64	0.87	0.88	0.63	13	51
34.5	0.61	0.89	0.88	0.62	11	55
35	0.55	0.90	0.89	0.59	10	64

[a] False positive and false negative errors are reported for each CARS cutoff using diagnosis based on clinical judgment as the gold standard for diagnosis

Table 5

Sensitivity, specificity, positive predictive values and negative predictive values of CARS cutoffs for diagnosis of autistic disorder in a 4-year-old sample (*N* = 148)

CARS cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value	False positive errors[a]	False negative errors[a]
27	0.92	0.64	0.86	0.78	16	8
27.5	0.91	0.68	0.87	0.77	14	9
28	0.90	0.73	0.89	0.76	12	10
28.5	0.88	0.75	0.89	0.72	11	13
29	0.86	0.73	0.88	0.68	12	15
29.5	0.85	0.77	0.90	0.68	10	16
30	0.86	0.80	0.91	0.71	9	14
30.5	0.81	0.86	0.93	0.66	6	20
31	0.76	0.89	0.94	0.61	5	25
31.5	0.73	0.91	0.95	0.59	4	28
32	0.70	0.93	0.96	0.57	3	31
32.5	0.68	0.98	0.99	0.56	1	34
33	0.63	0.98	0.99	0.53	1	38
33.5	0.59	1.00	1.00	0.51	0	43
34	0.55	1.00	1.00	0.48	0	47
34.5	0.50	1.00	1.00	0.46	0	52
35	0.45	1.00	1.00	0.44	0	57

[a]False positive and false negative errors are reported for each CARS cutoff using diagnosis based on clinical judgment as the gold standard for diagnosis

Table 6

Sensitivity, specificity, positive predictive values and negative predictive values of CARS cutoffs for diagnosis of ASD in a 2-year-old sample ($N = 376$)

CARS cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value	False positive errors[a]	False negative errors[a]
21	0.99	0.42	0.76	0.98	77	1
21.5	0.99	0.48	0.78	0.97	69	2
22	0.99	0.51	0.79	0.97	65	2
22.5	0.98	0.57	0.80	0.94	57	5
23	0.98	0.62	0.83	0.93	50	6
23.5	0.97	0.66	0.84	0.93	45	7
24	0.96	0.70	0.85	0.90	40	10
24.5	0.95	0.79	0.89	0.89	28	13
25	0.93	0.85	0.92	0.87	20	17
25.5	0.92	0.89	0.94	0.85	14	19
26	0.91	0.91	0.95	0.85	12	21
26.5	0.90	0.91	0.95	0.84	12	24
27	0.88	0.92	0.95	0.81	11	28
27.5	0.85	0.94	0.96	0.78	8	36
28	0.81	0.95	0.97	0.73	6	46
28.5	0.79	0.95	0.97	0.71	6	52
29	0.77	0.95	0.97	0.69	6	57
29.5	0.75	0.95	0.97	0.68	6	61

[a]False positive and false negative errors are reported for each CARS cutoff using diagnosis based on clinical judgment as the gold standard for diagnosis

Table 7

Sensitivity, specificity, positive predictive values and negative predictive values of CARS cutoffs for diagnosis of ASD in a 4-year-old sample (*N* = 230)

CARS cut-off	Sensitivity	Specificity	Positive predictive value	Negative predictive value	False positive errors[a]	False negative errors[a]
21	0.99	0.70	0.85	0.98	25	1
21.5	0.98	0.72	0.86	0.95	23	3
22	0.97	0.74	0.87	0.92	21	5
22.5	0.95	0.77	0.88	0.89	19	8
23	0.93	0.82	0.90	0.86	15	11
23.5	0.89	0.83	0.90	0.81	14	16
24	0.88	0.87	0.92	0.80	11	18
24.5	0.86	0.93	0.96	0.80	6	20
25	0.84	0.93	0.95	0.77	6	23
25.5	0.82	0.95	0.97	0.74	4	27
26	0.81	0.95	0.97	0.74	4	28
26.5	0.78	0.98	0.98	0.71	2	32
27	0.76	1.00	1.00	0.69	0	36
27.5	0.74	1.00	1.00	0.68	0	39
28	0.71	1.00	1.00	0.66	0	42
28.5	0.69	1.00	1.00	0.64	0	46
29	0.68	1.00	1.00	0.63	0	48
29.5	0.66	1.00	1.00	0.62	0	50

[a]False positive and false negative errors are reported for each CARS cutoff using diagnosis based on clinical judgment as the gold standard for diagnosis