J Autism Dev Disord (2008) 38:440–448
DOI 10.1007/s10803-007-0407-z

ORIGINAL PAPER

Does Parent Report of Behavior Differ Across ADOS-G Classifications: Analysis of Scores from the CBCL and GARS

Darryn M. Sikora · Trevor A. Hall · Sigan L. Hartley · Aimee E. Gerrard-Morris · Sarah Cagle

Received: 12 June 2006 / Accepted: 4 June 2007 / Published online: 6 July 2007

Abstract Behavior checklists are often utilized to screen for Autism Spectrum Disorders (ASDs) when comprehensive evaluations are unfeasible. The usefulness of two behavioral checklists, the Gilliam Autism Rating Scale (GARS) and Child Behavior Checklist (CBCL), in identifying ASDs was investigated among 109 children with Autism, 32 children with ASD, and 51 Non-Spectrum children based on Autism Diagnostic Observation Schedule-Generic classifications. The GARS did not distinguish children with ASDs from those without. The Withdrawn and Pervasive Developmental Problems subscales of the CBCL were higher among children with Autism than among Non-Spectrum children. These CBCL subscales also had better sensitivity and specificity in identifying children with Autism than the GARS. Results suggest that the CBCL is a useful behavioral checklist for screening ASDs.

Keywords Pervasive developmental disorders · Autism assessment · Behavior checklists · Clinical utility

Introduction

Autism spectrum disorders (ASDs) are a group of developmental disorders defined by impairments in the areas of communication and socialization, as well as patterns of restricted or repetitive behaviors [American Psychiatric Association (APA), 2000]. Standardized observations, structured parent interview, and evaluations by specialized professionals are instrumental in the reliable diagnosis of ASDs (South et al., 2002). However, developmental concerns such as ASDs often first present at primary care or educational settings that are unable to provide such thorough evaluations. Brief screening tools such as parent-completed behavior checklists are often utilized in these settings to determine referrals for further evaluation. The present study examined the diagnostic utility of two behavior checklists, the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2000) and Gilliam Autism Rating Scale (GARS; Gilliam, 1995), for identifying ASDs.

The current "gold standard" for diagnosing ASDs involves the use of standardized direct observations, namely the Autism Diagnostic Observation Schedule-Generic (ADOS-G; Lord et al., 2000) in concert with parent interview and clinical judgment. The ADOS-G is a semi-structured, standardized, play-based assessment of interaction, communication, play skills, and repetitive and stereotyped behaviors. Activities are designed to provide planned opportunities to elicit autistic behaviors that are then coded and entered into a diagnostic algorithm. The resulting cutoff scores establish three classification categories, Autism, ASD, or Non-Spectrum. The authors report good to excellent reliability of the items, domains, and classification categories and satisfactory validity in distinguishing children with autism from those without (Lord et al., 2000).

The high cost, time, and specialized training required for standardized direct observations such as the ADOS-G, however, limits its use in primary care and educational settings. Parent-rated behavior checklists are often the only standardized instruments used to screen for ASDs in these settings. Behavior checklists have the advantage of requiring minimal time commitment and cost, and can be administered

D. M. Sikora (✉) · T. A. Hall · S. L. Hartley · A. E. Gerrard-Morris · S. Cagle
Department of Pediatrics, Child Development and Rehabilitation Center, Oregon Health and Science University, CDRC/OHSU, P.O. Box 574, Portland, OR 97207-0574, USA
e-mail: sikorad@ohsu.edu

by non-specialists. Moreover, scores on checklists can be compared to normative data, potentially providing more objective information about a child and limiting the number of "mistakes" that might be made by non-specialists.

There are numerous behavior checklists commercially available that evaluate behaviors consistent with ASDs. Validity and reliability data are reported in most behavior checklist manuals. However, independent evaluation of these checklists is imperative to ensure that measures are statistically sound across samples (e.g., no shrinkage of coefficients) and can be reliably used in the screening process or as part of a comprehensive evaluation. It is also imperative to understand the sensitivity and specificity of behavior checklists in correctly identifying children with ASDs. This information is critical to increasing accurate and early diagnosis of ASDs. Concern has been raised that if a child is categorized as negative for an ASD on a behavior checklist in a primary care or educational setting, and thus fails to receive a referral to specialty services for further evaluation, the long-term consequences for that child are potentially devastating (South et al., 2002).

Some behavior checklists are specific to ASDs, while others provide information about ASD symptoms as part of a broader review of behavioral and emotional problems. Many diagnostic clinics use both ASD specific checklists and checklists that measure a broad range of behavioral and emotional problems in a single evaluation, thus increasing cost, clinician time, and response burden for families. If a checklist that measures a broad range of behavioral and emotional problems proved to be as useful as an ASD specific checklist in distinguishing children with ASDs from those without ASDs, both time and money potentially could be saved. That same checklist could serve as an adequate screening instrument for a variety of behavioral and emotional problems that are often found in children with ASDs.

The purpose of the present study was to determine whether a checklist that measures a broad range of behavioral and emotional problems, the CBCL, is at least as clinically useful as an ASD specific behavior checklist, the GARS, in identifying children with ASDs. The GARS is widely accepted as an accurate measure of a child's behavior in schools, clinics, and on research projects (South et al., 2002). It is an informant-rated checklist of behaviors that correspond to the Diagnostic and Statistical Manual, Fourth Edition (DSM-IV, APA, 1994) core deficits of Autistic Disorder. There are three behavioral subscales and a scale addressing early developmental history. The total score or Autism Quotient (AQ) is a standard score (*M* 100, SD 15) meant to measure the "likelihood that a child has autism." Higher scores indicate a greater likelihood of autism, with a score of 90 or above categorized as "probably autistic." The GARS manual reports satisfactory inter-item and test–retest reliability and adequate concurrent validity with the Autism Behavior Checklist (ABC; Krug, Arick, & Almond, 1993). The CBCL is the most widely used single measure of child behavior (Daugherty & Shapiro, 1994). It is an informant-rated checklist of a wide variety of behavioral problems and includes a Pervasive Developmental Problems scale. Scale scores are based on T-scores, with scores ≥70 generally considered to be within the Clinically Significant range. The CBCL manual reports satisfactory reliability and validity.

Only two studies have been published regarding the clinical utility of the GARS in identifying ASDs. South et al. (2002) examined GARS data in a sample of 119 children aged 3–10 years with a DSM-IV diagnosis of Autistic Disorder made from recognized autism experts using combinations of the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003), ADOS-G, and Vineland Scales of Adaptive Behavior, Survey Form (Sparrow, Balla, & Cicchetti, 1984). The mean GARS AQ score of 90.10 (SD 13.92), was significantly below that of the reference mean reported in the GARS manual (*M* 100, SD 15), at a $p < 0.05$ level. Moreover, 63 (52%) of children fell below the specified AQ criterion of 90, and were considered to have a low probability of autism. This translates into a GARS sensitivity of 48%. Lecavalier (2005) reported similar results in his sample of 284 students with special education eligibility of ASD. The mean GARS AQ score of 85.8 (SD 15.8) was also below the referenced mean, and resulted in 62.2% of the sample scoring below the specified AQ criterion of 90 and considered to have a below average, low or very low probability of autism. Neither study provided data on the specificity of the GARS in a clinical setting.

Although many studies have been published regarding the clinical utility of the CBCL, only one study could be found related to the identification of children with ASD (Duarte, Bordin, de Oliveira, & Bird, 2003). Duarte et al. (2003) compared scores of a now outdated version of the CBCL (the CBCL/4-18) among 36 children with autism and related conditions, 31 children with other psychiatric disorders, and a control group of 34 school-aged children. In addition to the specified CBCL Scale scores, they evaluated a non-standardized factor, "Autistic/Bizarre," derived from a factor analysis on a sample of 204 preschool-aged boys. They found that the Thought Problems scale differentiated children with autism from children with other psychiatric disorders with a sensitivity of 82.9% and specificity of 71.0%. The Autistic/Bizarre scale differentiated children with autism from children with other psychiatric disorders with a sensitivity of 88.6% and specificity of 80.0%. A combination of the Autistic/Bizarre scale and the Aggression scale increased sensitivity to 91.4% and specificity to 96.7%. Duarte et al. concluded that the CBCL could be used to identify children with autism when more comprehensive evaluations were not possible.

Our clinical impressions of the GARS and CBCL during evaluations of children referred to a tertiary Autism Clinic, aligned with these findings. Over time, we noticed that scores on the GARS did not correspond to results from other standardized instruments or with professional judgment. More specifically, whether a child obtained an AQ above 90 on the GARS did not seem to predict whether the child would receive an ASD diagnosis. On the other hand, we noticed that clinically significant elevations on the DSM-oriented scale of Pervasive Developmental Problems and the syndrome scale of Withdrawn on the CBCL were in much better agreement with the clinical diagnosis than was the GARS AQ. Based on results from the studies described above and our own clinical observations, it was hypothesized that the CBCL would be at least as good if not better than the GARS at distinguishing children with and without ASDs based on their ADOS-G classifications.

Methods

Participants

This study examined GARS and CBCL scores for 147 children (38 girls and 109 boys) aged 36–71 months (*M* 53.54, SD 10.59) who participated in an evaluation through the Autism Program at the Child Development and Rehabilitation Center (CDRC) at Oregon Health and Science University (OHSU) between August 2003 and June 2005. Ethnicity data was obtained from 45.6% of the sample. For this subset, 77.6% of participants were Caucasian, 6.0% were African American, 9.0% were Asian American/Pacific Islander, 6.0% were Hispanic, and 1.5% were of other ethnicity. This ethnicity breakdown corresponds to the demographics of children seen at the CDRC more generally. Caregivers did not report ethnicity for 54.4% of participants on an informant-reported questionnaire. On the Mullen Scales of Early Learning-AGS Edition (MSEL; Mullen, 1995), participants had an average Developmental Composite score of 61.57 (SD 20.61). Participants were divided into three groups based on ADOS-G classifications. Seventy-nine children received an ADOS-G classification of Autism, 18 children received an ADOS-G classification of ASD, and 50 children received an ADOS-G classification of Non-Spectrum.

Measures

Autism Diagnostic Observation Scale-Generic

The ADOS-G is a semi-structured, standardized, play-based assessment measure designed to elicit autistic behaviors that are then coded and entered into a diagnostic algorithm. The ADOS-G is divided into four separate modules: each module is aimed at a specific level of expressive language ability. The use of different modules reduces possible biasing effects of differences in language skills (Lord et al., 2000). Sixty-four participants were administered module 1, 31 participants were administered module 2, and five participants were administered module 3. The module used with 47 participants was not recorded.

Scoring of the ADOS-G occurs immediately after its administration. Each item is scored on a 0–3 scale (0 = no evidence of abnormal behavior to 3 = markedly abnormal behavior) scores (Lord et al., 2000). The ADOS-G algorithms contain those items with the highest inter-rater reliabilities that discriminated among Autism, ASD, and non-spectrum individuals in the standardization sample. Each module has a different algorithm. Items used in the algorithms are divided into four areas: Communication, Social Interaction, Play/Creativity, and Restricted/Repetitive Behaviors or Interests. Cutoff scores in the domains of Communication, Social Interaction, and Combined (Communication + Social Interaction), allow an individual to be placed in a(n) Autism, ASD, or Non-spectrum category.

The authors report good to excellent reliability of the items, domains, and classification categories (Lord et al., 2000). Validity studies were conducted by carrying out several analyses. Correlation matrices were generated for all items on each module for all domains. Inter-correlations that were above 0.70 for two or more items within a module and overlapped in conceptualization were removed from the algorithm (Lord et al., 2000). A fixed-effects analysis of variance (ANOVA) was then carried out to compare samples of autism and non-spectrum individuals. Items that did not show significant differences were excluded from the algorithm. Further analyses were conducted to compare the three classification groups (i.e., Autism, ASD, and non-spectrum) for each of the items that had been retained in the algorithm.

Gilliam Autism Rating Scale

The GARS is a behavior checklist developed for use in individuals ages 3–22 years. The questionnaire consists of 56 items, each describing a different behavior often observed in individuals with autism, and is divided into four subscales: Social Interaction, Communication, Stereotyped Behaviors, and Developmental Disturbances. Those completing the form are asked to rate the frequency of each behavior based on a four-point scale (0 = Never Observed, 1 = Seldom Observed, 2 = Sometimes Observed, and 3 = Frequently Observed). The scores for each scale are then summed and converted to standard scores (*M* 10, SD 3), based on a normative sample of 1,092 individuals previously diagnosed with ASDs. These four standard scores are then combined into a summary score, the AQ (M 100, SD = 15), which is then used to predict the probability that a child has an ASD diagnosis. The AQ is broken down into seven different predictive categories, ranging from a "Very Low" to a "Very High" probability of autism. An AQ of 90 or above suggests that the child is "probably autistic." The manual for the GARS reports adequate reliability and validity.

Child Behavior Checklist

The CBCL has been developed to measure a wide variety of behavior concerns, in both internalizing and externalizing areas. Several versions of the CBCL are available: The CBCL for ages 1.5–5 years was used in this study. The questionnaire consists of 100 items; each item describes a specific behavior. Those completing the form are asked to rate the frequency of each behavior on a three-point Likert scale (0 = Not True, 1 = Somewhat or Sometimes True, and 2 = Very True or Often True). Scores are then summed and converted to T-scores (M 50, SD 10) on seven different syndrome scales (Emotionally Reactive, Anxious/Depressed, Somatic Complaints, Withdrawn, Sleep Problems, Attention Problems, and Aggressive Behavior), as well as five different DSM-oriented scales (Affective Problems, Anxiety Problems, Pervasive Developmental Problems, Attention Deficit/Hyperactivity Problems and Oppositional Defiant Problems). These scores combine to yield an Internalizing Problems, Externalizing Problems, and Total Problems composite score. Only the syndrome and DSM-Oriented scale scores were examined in the present study. A T-score of 70 and above is generally considered "clinically significant." The manual for the CBCL reports adequate reliability and validity for scale scores.

Procedures

The Autism Program was established within the CDRC at OHSU in August 2003. Primary care physicians across the state and region referred patients to the CDRC for the purpose of having an inter-disciplinary team evaluate behaviors that are suspected of being consistent with ASDs. The majority of children were accompanied to the clinic by their parents. Occasionally foster parents or caseworkers accompanied the child.

Since 2003, the clinic team has developed a state-of-the-art diagnostic protocol, to include the most reliable and accurate instruments available. Participants in the present study were administered the MSEL-AGS Edition, the ADOS-G, and the Autism Screening Instrument for Educational Planning, second edition (Krug et al., 1993) by a random combination of team members. Caregivers were given a comprehensive, semi-structured diagnostic interview of DSM-IV-TR criteria for ASDs, the Vineland Adaptive Behavior Scales, Survey Form and were asked to complete the GARS and the CBCL. The GARS and the CBCL were given to caregivers in random order so as to eliminate potential test order bias. Caregiver completed the forms on the day of the appointment. The majority of informants who completed the forms were female. Upon completion, the forms were collected and scored under the supervision of a licensed psychologist. The ADOS-G was administered and scored immediately after administration by two, licensed clinicians that had reached clinical reliability on the instrument, and before results from the behavior checklists were reviewed.

Data were entered into a de-identified clinical database, which was approved by the OHSU's Institutional Review Board for research use. The GARS AQ standard score and T-scores from the DSM-oriented and syndrome scales of the CBCL were analyzed. In an effort to reduce bias, ADOS-G classifications rather than exiting diagnoses were utilized as grouping variables. ADOS-G scores inform clinical judgment and there was overlap (82.99%) between these two diagnostic methods, however, only the former was used in this study.

Results

Table 1 displays the subject characteristics of children with ADOS-G classifications of Autism, ASD, and Non-Spectrum. Chi-square and one-way analyses of variance (ANOVAs) were conducted to identify potential differences in subject characteristics among the three groups. Chi-square analyses revealed no significant group difference with respect to sex ($\chi^2 = 3.95$, $p = 0.14$) or ethnicity ($\chi^2 = 0.84$, $p = 0.84$). A one-way ANOVA indicated a significant difference in age among the groups [F (2, 147) = 7.25, $p < 0.01$]. Post hoc analyses using the Tukey Honestly Significant test revealed that the Autism and ASD group were younger than the Non-Spectrum group ($p < 0.01$). There was also a significant group difference in the Mullen Early Learning Developmental Composite score [F (2, 110) = 12.60, $p < 0.01$]. Post hoc analyses using Tukey Honestly Significant Test revealed that the Autism group had a lower Developmental Composite score than the ASD group ($p < 0.01$) and the Non-Spectrum group ($p < 0.01$). There was not a significant difference in the Developmental Composite score between the ASD and Non-Spectrum group ($p = 0.98$).

Pearson correlations were conducted to evaluate the association between the GARS AQ and CBCL scale scores. There was a significant positive correlation between the GARS AQ and all CBCL scale scores: Affective Problems ($r = 0.52$, $p < 0.01$), Anxiety Problems

Table 1 Subject characteristics of the autism, autism spectrum disorder, and non-spectrum group

	Autism	ASD	Non-spectrum
Sex			
Males	54	13	42
Females	25	5	8
Age*			
Mean	50.66	55.06	57.54
SD	10.30	11.46	9.45
Ethnicity			
Caucasian	23	5	24
African American	3	0	1
Hispanic	2	0	2
Asian/Pacific Islander	4	0	2
Other	1	0	0
Missing	33	5	29
Mullen early learning composite*			
Mean	53.98	71.38	72.42
SD	13.51	25.15	23.98

* One-way analysis of variance indicated presence of a significant difference among groups

($r = 0.46$, $p < 0.01$), Pervasive Developmental Problems ($r = 0.66$, $p < 0.01$), Attention Deficit/Hyperactivity Problems ($r = 0.48$, $p < 0.01$), Oppositional Defiant Problems ($r = 0.36$, $p < 0.01$), Emotionally Reactive ($r = 0.47$, $p < 0.01$), Anxious/Depressed ($r = 0.37$, $p < 0.01$), Somatic Complaints ($r = 0.38$, $p < 0.01$), Withdrawn ($r = 0.58$, $p < 0.01$), Sleep Problems ($r = 0.34$, $p < 0.01$), Attention Problems ($r = 0.50$, $p < 0.01$), and Aggressive Behavior ($r = 0.46$, $p < 0.01$). The CBCL Withdrawn and Pervasive Developmental Problems scale scores were most strongly correlated with the GARS AQ.

Table 2 displays the means and standard deviations of the GARS AQ and CBCL scale scores for the Autism, ASD, and Non-Spectrum group. A multivariate analysis of variance (MANOVA) was used to examine potential differences between the three groups on the GARS and CBCL. Significant multivariate effects were obtained between groups [*Pillai's Trace* = 0.47, F (32, 252) = 2.45, and $p < 0.01$]. As shown in Table 2, there was not a significant difference among the groups for the GARS AQ. In contrast, there were significant differences, with small to moderate effect sizes, among ADOS-G classification groups for the following CBCL scales: Pervasive Developmental Problems, Anxious/Depressed, Withdrawn, and Aggressive Behavior. Post hoc analyses using the Bonferroni technique revealed significant differences between the Autism and the Non-Spectrum group on these scales ($p < 0.05$). There were no significant differences between the ASD and Non-Spectrum group or between the ASD and Autism group. Among the CBCL scales that revealed significant differences, the Withdrawn Scale accounted for the most variance among groups (partial eta squared = 0.11).

Pearson product moment and Biserial correlations were conducted to identify associations between subject characteristics and relevant CBCL and GARS scores. There was not a significant correlation between the GARS and age ($r = 0.02$, $p = 0.86$), sex ($r = 0.06$, $p = 0.49$), or Mullen Early Learning Developmental Composite score ($r = -0.15$,

Table 2 Multivariate analysis of variance for GARS and CBCL scale scores for the Autism, ASD, and non-spectrum groups

	Autism	ASD	Non-spectrum	F	*p*-value	Partial ETA2
GARS AQ	90.56 (14.19)	89.33 (17.29)	87.10 (16.52)	0.77	0.46	0.01
CBCL						
Emotionally reactive	61.92 (9.70)	64.20 (10.15)	66.04 (11.63)	2.39	0.10	0.03
Anxious/depressed	56.86 (7.23)	60.00 (7.46)	61.48 (12.47)	3.89	0.03	0.05
Somatic complaints	59.17 (6.77)	60.33 (7.71)	60.02 (7.26)	0.26	0.77	0.00
Withdrawn	73.33 (10.79)	66.93 (8.22)	65.96 (9.84)	8.71	0.00	0.11
Sleep problems	58.83 (9.63)	59.67 (8.63)	62.00 (11.85)	1.43	0.24	0.02
Attention problems	63.35 (9.54)	62.27 (8.57)	65.52 (9.85)	1.06	0.35	0.02
Aggressive behavior	62.14 (11.74)	59.60 (6.71)	67.82 (14.73)	4.11	0.02	0.06
Affective problems	62.08 (11.33)	64.80 (8.41)	65.06 (11.59)	1.22	0.30	0.02
Anxiety problems	58.58 (9.55)	61.20 (8.49)	62.44 (10.62)	2.12	0.12	0.03
Pervasive developmental problems	75.04 (9.22)	73.20 (6.56)	70.08 (10.30)	4.25	0.02	0.06
Attention deficit/hyperactivity problems	61.73 (8.61)	60.00 (8.15)	62.28 (8.92)	0.4	0.67	0.01
Oppositional defiant problems	59.53 (11.37)	58.73 (7.14)	64.00 (10.45)	3.04	0.05	0.04

$p = 0.13$). There was not a significant correlation between the CBCL Pervasive Developmental Problems and age ($r = -0.13$, $p = 0.11$), sex ($r = -0.03$, $p = 0.72$), or Mullen Early Learning Developmental Composite score ($r = -0.09$, $p = 0.36$). There was also not a significant correlation between CBCL Aggression and age ($r = -0.70$, $p = 0.41$), sex ($r = -0.05$, $p = 0.55$), or Mullen Early Learning Developmental Composite score ($r = -0.07$, $p = 0.44$). Similarly, there was not a significant correlation between CBCL Anxious/Depressed and age ($r = -0.02$, $p = 0.83$), sex ($r = -0.01$, $p = 0.93$), or Mullen Early Learning Developmental Composite score ($r = 0.17$, $p = 0.08$). There was a significant negative correlation between the CBCL Withdrawn and age ($r = -0.21$, $p = 0.01$) and Mullen Early Learning Developmental Composite score ($r = -0.22$, $p = 0.02$), but not sex ($r = -0.01$, $p = 0.91$).

Multiple analyses of covariance (MANCOVAs) were conducted to determine whether the GARS and CBCL scales were able to distinguish among ADOS-G classification groups when subject characteristics were controlled. Analysis revealed a significant multivariate effect between ADOS-G classification groups [*Pillai's Trace* = 0.26, F (10, 154) = 2.27, and $p = 0.02$]. There was not a significant group effect for then GARS [F (2, 309) = 1.27, $p = 0.28$]. There was a significant group effect for CBCL Withdrawn [F (2, 587) = 5.87, $p = 0.01$] and Pervasive Developmental Problems [F (2, 307) = 3.76, $p = 0.03$]. There was not a significant group effect for CBCL Aggression [F (2, 425) = 2.67, $p = 0.07$] or Anxious/Depressed [F (2, 112) = 1.51, $p = 0.23$]. Because the CBCL Aggression and Anxious/Depressed scales were no longer able to distinguish among the Autism, ASD, and Non-Spectrum groups when controlling for subject characteristics and ADOS module, these scales were not further analyzed.

The sensitivity and specificity of identifying children with ADOS-G classifications of Autism (i.e., autism positive) versus Non-Spectrum (i.e., autism negative) is presented in Table 3. The established cutoff of ≥90 was used to distinguish children with autism (i.e., autism positive) from those without (i.e., autism negative) for the GARS. A cutoff of ≥70 was used to distinguish children with autism (i.e., autism positive) from those without (i.e., autism negative) on the CBCL Withdrawn and Pervasive Developmental Problems scales. The CBCL Withdrawn and Pervasive Developmental Problems scales had better sensitivity then the GARS. The CBCL Withdrawn scale also had better specificity than the GARS. Table 4 displays the sensitivity and specificity rates by cognitive functioning [high functioning (Mullen Early Learning Developmental Composite score >70) versus low functioning (Mullen Early Learning Developmental Composite score ≤70)] and sex. In all analyses, the CBCL Withdrawn and Pervasive Developmental Problems scales had better sensitivity than the GARS AQ in identifying children with autism.

Discussion

Concerns of ASD often first present at primary care and educational settings. These settings often do not have the time, financial resources, or clinicians with specialized training for comprehensive ASD evaluations and often utilize behavioral checklists to guide referrals for further evaluation. The purpose of the present study was to determine the clinical utility of two behavior checklists, the GARS and CBCL, in identifying ASDs in young children. In order to accomplish this aim, CBCL and GARS scores were assessed among 147 children aged 3–5 years who had received an ADOS-G classification of Autism, ASD, or Non-Spectrum from a tertiary Autism Clinic.

Findings from the present study indicate that two CBCL scales, Withdrawn and Pervasive Developmental Problems, have better discriminate validity in distinguishing children with autism from children without autism than the GARS AQ. Both the CBCL Withdrawn and Pervasive Developmental problems scales significantly differentiated children with an ADOS-G classification of Autism from children with an ADOS-G classification of Non-Spectrum, after controlling for subject characteristics. In contrast, the

Table 3 Sensitivity and specificity of the GARS and CBCL withdrawn (Wd) and pervasive developmental problems (PDP) scales in identifying children with ADOS-G classification of autism versus non-spectrum

Behavioral checklist		ADOS-G classification		Sensitivity	Specificity
		Autism positive	Autism negative		
GARS	Autism positive	42	23	42/79 = 53.16%	27/50 = 54.00%
	Autism negative	37	27		
CBCL Wd	Autism positive	51	19	51/79 = 64.56%	31/50 = 62.00%
	Autism negative	28	31		
CBCL PDP	Autism positive	63	29	63/79 = 79.75%	21/50 = 42.00%
	Autism negative	16	21		

Table 4 Sensitivity and specificity of the GARS and CBCL Withdrawn (Wd) and pervasive developmental problems (PDP) scales in identifying ADOS-G classifications of autism versus non-spectrum by sex and cognitive functioning

	Behavioral checklist		ADOS-G classification		Sensitivity	Specificity
			Autism positive	Autism negative		
Girls	GARS	Autism positive	12	3	12/20 = 60.00%	3/11 = 50.00%
		Autism negative	8	3		
	CBCL Wd	Autism positive	16	4	16/20 = 80.00%	2/6 = 33.33%
		Autism negative	4	2		
	CBCL PDP	Autism positive	13	2	13/20 = 65.00%	4/6 = 66.67%
		Autism negative	7	4		
Boys	GARS	Autism positive	24	13	24/44 = 54.55%	14/27 = 51.85%
		Autism negative	20	14		
	CBCL Wd	Autism positive	33	9	33/44 = 75.00%	18/27 = 66.66%
		Autism negative	11	18		
	CBCL PDP	Autism positive	36	15	36/44 = 81.82%	12/27 = 44.44%
		Autism negative	8	12		
High functioning	GARS	Autism positive	1	7	1/6 = 16.67 %	5/12 = 41.67%
		Autism negative	5	5		
	CBCL Wd	Autism positive	2	5	2/6 = 33.33%	7/12 = 58.33%
		Autism negative	4	7		
	CBCL PDP	Autism positive	3	8	3/6 = 50.00%	4/12 = 33.33%
		Autism negative	3	4		
Low functioning	GARS	Autism positive	35	16	35/58 = 60.34%	22/38 = 57.89%
		Autism negative	23	22		
	CBCL Wd	Autism positive	44	6	44/58 = 75.86%	15/21 = 71.43%
		Autism negative	14	15		
	CBCL PDP	Autism positive	49	11	49/58 = 84.48%	10/21 = 47.62%
		Autism negative	9	10		

High cognitive functioning = Mullen early learning composite score >70; low cognitive functioning = Mullen early learning composite score >70

GARS AQ did not significantly differentiate these groups. This finding is in line with previous research that also suggests that the established cutoff for the GARS AQ has poor diagnostic utility in identifying children with ASDs (Lecavalier, 2005; South et al., 2002).

The CBCL Withdrawn and Pervasive Developmental Problems scales also had better sensitivity in distinguishing children with ADOS-G classifications of Autism from children with classifications of Non-Spectrum than did the GARS AQ when the entire sample was considered. Moreover, these two CBCL scales had better sensitivity when analyses were conducted separately by sex and cognitive functioning, indicating the CBCL has better diagnostic utility than the GARS for both boys and girls, and high- and low-functioning children. Past studies have similarly shown that the GARS has poor sensitivity as a ASD screening tool (Lecavalier, 2005; South et al., 2002). The sensitivity of the GARS in the present study was somewhat higher that that reported in these past studies (Lecavalier, 2005; South et al., 2002), which may be due to differences in sampling as well as criteria for establishing ASDs. Factor analyses by Lecavalier (2005) suggest that the GARS poor sensitivity may be due to an over-emphasis on repetitive and stereotyped behaviors and an under-emphasis on social and communication impairments. Further research is needed to investigate this hypothesis.

The poor sensitivity of the GARS AQ suggests that this checklist may be more likely to lead to false negatives (i.e., identifying a child who has autism as not having autism) than the CBCL Withdrawn and Pervasive Developmental Problems scales. Researchers have voiced concern for the potentially serious ramifications of failing to accurately identify children with autism in terms of delaying referrals to specialized clinics and subsequently delaying appropriate intervention and services (South et al., 2002). Thus, the increased sensitivity of the CBCL Withdrawn and Pervasive Developmental Problems scales, while perhaps small in terms of percentage points, translates into meaningful differences in terms of ensuring that children with ASDs are identified early on and are enrolled in intervention services.

Specificity was higher for the CBCL Withdrawn scale than the Pervasive Developmental Problems scale or GARS

AQ. Stated differently, the CBCL Withdrawn scale resulted in the lowest rate of false positives (i.e., incorrectly identifying a child who does not have autism as having autism). This was true both for the group-wide analysis as well as when specificity was examined separately for boys and high and low-cognitive functioning children. For girls, the CBCL Pervasive Developmental Problems scale had the highest sensitivity. False positives can also negatively impact children and families by hindering the differential diagnosis process and resulting in unneeded referrals to specialty clinics and thus unneeded financial, time, and emotional burdens for families. Moreover, accurate diagnosis and identification of problems is required before appropriate interventions can be designed. It should be noted that both the CBCL Withdrawn and Pervasive Developmental Problems scales and GARS AQ had fairly poor specificity, underscoring the need for comprehensive evaluations that include direct observations, parent interview, and expert clinical opinion to ultimately determine the appropriateness of ASD diagnosis.

Together, findings from the present study suggest that the CBCL Withdrawn and Pervasive Developmental Problems scales are at least as good as, if not better than the GARS in identifying ASDs in young children when comprehensive evaluations are not possible. Duarte et al. (2003) also found the CBCL to be an adequate screening tool for identifying ASDs in older children, aged 4–18 years. The sensitivity and specificity of the CBCL for older children reported by Duarte et al. was higher than that found in the present study. This may reflect differences in the presentation of autism in younger ages and/or the challenges of assessing social, communication, and behavior impairments in young children more generally (e.g., Stone, Hoffman, Lewis, & Ousley, 1994; Vig & Jedrysek, 1999).

Unlike the GARS, the CBCL has the added benefit of supplying information on a broad array of behavior problems. This information is important for differential diagnosis of ASDs versus other behavioral and emotional problems. Moreover, children with ASDs have a high rate of behavioral and emotional problems (Ozonoff, Goodlin-Jones, & Solomon, 2005). The CBCL can help identify comorbid behavioral and emotional problems, which may need further evaluation. The CBCL can also offer a profile of personal strengths (e.g., positive affect) and weaknesses (e.g., struggles with inattention) to help guide the design of individualized interventions to address the specific needs and abilities of children. The efficiency of the CBCL in screening for ASDs as well as other behavioral problems is thus appealing for primary care and educational settings where efficiency, cost, and response burden must be taken into consideration. Efficiency of screening tools may also be vital for ensuring that data collected are of high quality. Clinical researchers have observed that increased response burden on parents may lead to degradation in responding due to fatigue and the discomfort that may occur when the parents must answer a number of seemingly irrelevant questions (Kelly, O'Malley, Kallen, & Ford, 2005).

It is important to note that recently a revised version of the GARS, the GARS-2 (Gilliam, 2006) was published. According to the manual, the GARS-2 has been improved through establishing new norms and guidelines for interpreting the AQ, newly named Autism Index. This revised version however did not alter the extent to which the checklist addresses the three defining areas of impairment of autism: Social, communication, and repetitive or stereotyped play, which has been identified as a potential source of the poor diagnostic utility of the GARS (Lecavalier, 2005). Future research will have to determine whether these new norms and interpretation guidelines have improved the diagnostic utility of the GARS-2 as an ASDs screening tool.

The present study offered an initial evaluation of the diagnostic utility of two behavioral checklists in identifying children with ASDs. Further validation of the CBCL Withdrawn and Pervasive Developmental Problems scales would come from comparisons with scores from the ADI-R and expert clinical judgment. Moreover, the diagnostic utility of these two CBCL scales should be evaluated in other samples, including children with other developmental and psychological diagnoses. A focused analysis of the CBCL Withdrawn and Pervasive Developmental Problems scales is also warranted in order to ensure that the critical areas of social and communication skills are given adequate attention.

The present study utilized a clinical sample of children who were referred for evaluation to an autism clinic. The use of a clinical sample increased ecological validity of findings, but prevented a true experimental design, resulting in an unequal number of subjects in the various diagnostic groups as well as unequal variance. However, post hoc corrections were used in the multivariate analysis to account for this limitation. Based on the subset of the children for whom ethnicity was reported (45.6%), the present sample is representative of children referred to a children's hospital in Oregon more generally. The higher prevalence of males as compared to females also parallels rates of autism more generally (APA, 1994). Children who received an ADOS-G classification of Autism had lower cognitive functioning than children with classifications of ASD or Non-Spectrum. This is not surprising given the higher rate of mental retardation in children with autism as compared to their peers (Edelson, 2006). However children who received an ADOS-G classification of Autism were also younger than children classified as Non-Spectrum. While this age difference was controlled for in analyses, it may mean that findings from the present study are more reflective of children aged 3 or 4 as opposed to 5.

The use of a clinical sample also meant that if caregivers were unwilling or unable to complete the behavioral checklists, information was not gathered. We cannot rule out the possibility that caregivers who did not complete the behavioral checklists differed from those who completed measures. Finally, it is important to highlight that no behavior checklist should be used as the only standardized measure of autistic symptoms in diagnosing ASDs. In summary, results from the present study suggest that the CBCL, and the Withdrawn and Pervasive Developmental Problems scales in particular, can be used to screen for ASDs in settings where a more thorough evaluation cannot occur.

Acknowledgments The authors would like to thank the clinicians who comprise the CDRC Autism Team for their dedication to children with autism and their families and their participation in data collection for this study.

References

Achenbach, T., & Rescorla, L. (2000). *Child behavior checklist*. Burlington, VT: ASEBA.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: American Psychiatric Association Press.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders, 4th edition, text revision, DSM-IV-TR*. Washington, DC: APA.

Daugherty, T. K. & Shapiro, S. E. (1994). Behavior checklists and rating forms. In: H. Thomas, N. J. King, & W. Yule (Eds.), *International handbook of phobic and anxiety disorders in children and adolescents* (pp. 331–347). New York, NY: Plenum Press.

Duarte, C. S., Bordin, I. A. S., de Oliveria, A., & Bird, H. (2003). The CBCL and the identification of children with autism and related conditions in Brazil: Pilot findings. *Journal of Autism and Developmental Disorders, 33*(6), 703–707.

Edelson, M. E. (2006). Are the majority of children with autism mentally retarded? A systematic evaluation of the data. *Focus on Autism and Other Developmental Disabilities, 21*, 66–68.

Gilliam J. E. (1995). *Gilliam autism rating scale*. TX: Pro-ed.

Gilliam J. E. (2006). *Gilliam autism rating scale* (2nd ed.). TX: Pro-ed.

Kelly, P. A., O'Malley, K. J., Kallen, M. A., & Ford, M. E. (2005). Integrating validity theory with use of measurement instruments in clinical settings. *Health Services Research, 40*(5), 1605–1619 Part II.

Krug, D. A., Arick, J., & Almond, P. (1993). *The autism screening instrument for educational planning, 2nd edition, ASIEP-2*. Austin, TX: Pro-Ed.

Lecavalier, L. (2005). An evaluation of the Gilliam autism rating scale. *Journal of autism and developmental disorders, online first*. Retrieved from http://www.springerlink.com Accessed on 10 November 2005.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., Pickles, A., & Rutter, M. (2000) The autism diagnostic observation schedule-generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205–223.

Mullen, E. (1995). *Mullen scales of early learning: AGS edition*. Circle Pines, MN: American Guidance Systems.

Ozonoff, S., Goodlin-Jones, B. L., & Solomon, M. (2005). Evidence-based assessment of autism spectrum disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*(3), 523–540.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *ADI-R: The autism diagnostic interview-revised*. Los Angelos, CA: Western Psychological Services.

South, M., Williams, B. J., Mcmahon, W. M., Owley, T., Filipek, P. A., Shernoff, E., Corsello, C., Lainhart, J., Landa, R., & Ozonoff, S. (2002). Utility of the Gilliam autism rating scale in research and clinic populations. *Journal of Autism and Developmental Disorders, 32*(6), 593–599.

Sparrow, S., Balla, D., & Cicchetti, D. (1984). *Vineland adaptive behavior scales—Interview edition*. Circle Pines, MN: AGS.

Stone, W. L., Hoffman, E. L., Lewis, S. E., & Ousley, O. Y. (1994). Early recognition of autism. Parental reports vs clinical observation. *Archives of Pediatrics & Adolescent Medicine, 148*(2), 174–179.

Vig, S. & Jedrysek, E. (1999). Autistic features in young children with significant cognitive impairment: Autism or Mental Retardation? *Journal of Autism and Developmental Disorders, 29*(3), 235–248.