ORIGINAL PAPER

Comparative Analysis of Three Screening Instruments for Autism Spectrum Disorder in Toddlers at High Risk

Iris J. Oosterling · Sophie H. Swinkels · Rutger Jan van der Gaag · Janne C. Visser · Claudine Dietz · Jan K. Buitelaar

Abstract Several instruments have been developed to screen for autism spectrum disorders (ASD) in high-risk populations. However, few studies compare different instruments in one sample. Data were gathered from the *Early Screening of Autistic Traits Questionnaire*, *Social Communication Questionnaire*, *Communication and Symbolic Behavior Scales-Developmental Profile*, *Infant-Toddler Checklist* and key items of the *Checklist for Autism in Toddlers* in 238 children (mean age = 29.6 months, SD = 6.4) at risk for ASD. Discriminative properties are compared in the whole sample and in two age groups separately (8–24 months and 25–44 months). No instrument or individual item shows satisfying power in discriminating ASD from non-ASD, but pros and cons of instruments and items are discussed and directions for future research are proposed.

I. J. Oosterling (✉) · S. H. Swinkels · R. J. van der Gaag · J. C. Visser · J. K. Buitelaar
Karakter Child and Adolescent Psychiatry University Center, Reinier Postlaan 12, 6525 GC Nijmegen, The Netherlands
e-mail: i.oosterling@karakter.com

S. H. Swinkels · R. J. van der Gaag · J. K. Buitelaar
Department of Psychiatry, University Medical Centre St Radboud, Nijmegen, The Netherlands

C. Dietz
Department of Child and Adolescent Psychiatry, Rudolf Magnus Institute for Neurosciences, University Medical Centre Utrecht, Utrecht, The Netherlands

J. K. Buitelaar
Nijmegen Centre for Evidence-Based Practice (NCEBP), Nijmegen, The Netherlands

Introduction

In the last 15 years great efforts have been put into developing methods and instruments for earlier detection of autism spectrum disorder (ASD). Research projects show that earlier identification of children with ASD is indeed feasible (Charman and Baird 2002). Two models for early detection of ASD prevail in the field. The *first* model includes a systematic population screening (first-level screening), in which autism-specific screeners are applied to all children at certain ages (e.g. 18 and 24 months of age), e.g. by primary care providers in conjunction with routine developmental surveillance. This population screening is advocated by the American Academy of Pediatrics (Johnson et al. 2007). The *second* model includes a two-stage screening approach, in which a specific screening instrument for ASD is only applied to children showing a deviant developmental path at a routine developmental surveillance (second-level screening). Such an approach is recommended in the Practice Parameters endorsed by the American Academy of Neurology and Child Neurology Society (Filipek et al. 2000).
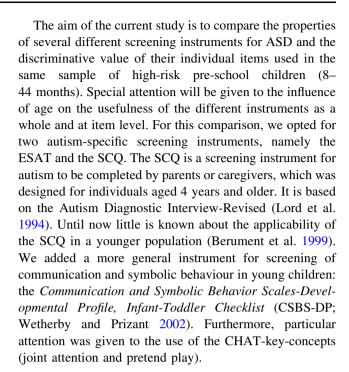
Two screening instruments have been evaluated in large unselected population samples. These first-level screening instruments are the *Checklist for Autism in Toddlers* (CHAT; Baron-Cohen et al. 1992; Baron-Cohen et al. 2000) and the *Early Screening of Autistic Traits Questionnaire* (ESAT; Dietz et al. 2006; Swinkels et al. 2006). The CHAT was developed in order to prospectively

identify autism at 18 months of age in a general population sample (Baron-Cohen et al. 1992). This checklist is based on the assumption that early impairments of joint attention skills are precursors of problems in developing a theory-of-mind functioning that is hypothesized to be a core deficit in autism later in life (Charman and Baron-Cohen 2006). The CHAT assesses 'simple' pretend play and joint attention behaviours using parental report and health practitioner observation through direct testing. The ESAT was developed to prospectively identify autism as early as at 14 months of age in a general population (Dietz et al. 2006; Swinkels et al. 2006). Using an empirical bottom-up approach, potential screening items were selected from the literature and tested in a pilot study. This resulted in the development of a population-based pre-screening instrument, the 4-item ESAT, and a longer 14-item version of the ESAT for use in populations at high-risk because either screened positive on the 4-item ESAT or determined by other means to be at high risk.

Several other autism-specific screening instruments have been developed and further studied in recent years. Examples of these screening instruments are the *Modified-CHAT* (M-CHAT; Robins et al. 2001), the *Social Communication Questionnaire* (SCQ; Berument et al. 1999; Rutter et al. 2003), the *Screening Test for Autism in Toddlers* (STAT; Stone et al. 2004), and the *Pervasive Developmental Disorders Screening Test-II* (PDDST-II; Siegel 2004). A common characteristic of most of these screening instruments is the inclusion of items on all three areas of impairment in ASD. The instruments vary, however, (a) in terms of coverage of other symptom areas, (b) in terms of the age at which they are to be administered, (c) as to whether they are to be used as a parent questionnaire or for direct observation by a professional (Bryson et al. 2003), and (d) as to whether they were originally intended and/or further studied as screens to be used in a general population (first-level screening), or in high-risk groups (second-level screening). For an overview of first- and second-level screening instruments, see Johnson et al. (2007, p. 1200–1201).

So far, little research has been completed on comparing the properties of different screening instruments at an early age within the one and the same sample. In addition, empirical evidence with regard to the use of different items for children at different ages is limited. Studies with the CHAT showed that items on pretend play and joint attention are important in screening children aged 18 months (Baron-Cohen et al. 1992, 2000), whereas findings of the ESAT studies revealed that at 14 months of age items related to: (a) direct smiling (smile directed to others), (b) reacting when spoken to, and (c) interest in other people, are most predictive for ASD (Dietz et al. 2006; Swinkels et al. 2006).

The aim of the current study is to compare the properties of several different screening instruments for ASD and the discriminative value of their individual items used in the same sample of high-risk pre-school children (8–44 months). Special attention will be given to the influence of age on the usefulness of the different instruments as a whole and at item level. For this comparison, we opted for two autism-specific screening instruments, namely the ESAT and the SCQ. The SCQ is a screening instrument for autism to be completed by parents or caregivers, which was designed for individuals aged 4 years and older. It is based on the Autism Diagnostic Interview-Revised (Lord et al. 1994). Until now little is known about the applicability of the SCQ in a younger population (Berument et al. 1999). We added a more general instrument for screening of communication and symbolic behaviour in young children: the *Communication and Symbolic Behavior Scales-Developmental Profile, Infant-Toddler Checklist* (CSBS-DP; Wetherby and Prizant 2002). Furthermore, particular attention was given to the use of the CHAT-key-concepts (joint attention and pretend play).

Method

Participants

The study sample included 238 children who were considered to be at risk for ASD because of either screen positive results on the ESAT 14-items ($n = 208$) or, when screen negative on the ESAT, because of sufficient clinical concern ($n = 30$). We had organised educational lectures to professionals in the region on recognising early signs of autism and on the use of screening questions. The inclusion into the study was as follows. Primary care workers, wishing to refer a child for assessment of ASD, were first required to complete the ESAT (with the assistance of the parents). Children who had a positive screen with the ESAT were always invited for further assessment. If a child screened negative with the ESAT, the referring professional had to provide additional information showing the child to be at high risk (either based on their own observations or based on parental comments). Children included in this study were referred to the child psychiatry outpatient unit in Nijmegen for further evaluation between October 2003 and April 2007. Forty-six children were 24 months or younger at screening (76% ASD), with the majority of them being between 18 and 24 months old. One Hundred and ninety-two children were between 25 and 44 months old at screening (65% ASD). No difference in IQ on a group level was found between the younger and older participants ($t(232) = -1.417$, $P > .05$). Seventy-eight percent was male. Ninety percent had a Dutch Caucasian

Table 1 Description of participants' primary diagnoses, age in months at screening and IQ scores

Diagnostic category	N		Age in months at screening				IQ[a]	
	8–24 months	25–44 months	Mean	SD	Min	Max	Mean	SD
Austism spectrum disorder (ASD)								
Autism	28	75	28.0	6.3	8	41	55	16
PDD-NOS	6	46	31.5	5.5	18	40	80	18
Asperger's syndrome	0	4	37.5	4.2	33	43	127	17
Rett's syndrome	1	0	12.0		12	12	49	
Psychiatric disorder-other								
Language disorder	3	20	29.1	6.4	12	38	83	9
ADHD	3	10	29.0	6.9	12	36	82	16
Mental retardation (without ASD)	0	12	33.6	4.5	27	39	58	9
ODD/behaviour disorder NOS	0	8	32.3	2.7	28	37	96	21
Mood disorder	0	4	32.5	5.0	26	38	112	14
Anxiety disorder	0	2	34.0	2.8	32	36	95	
Reactive attachment disorder	2	0	19.5	0.7	19	20	75	16
DC: 0–3 (e.g. regulation disorder or developmental phase problem)	1	11	33.0	6.2	19	44	100	17
No problem	2	0	17.0	2.8	15	19	99	29
Total	46	192	29.6	6.4	8	44	71	24

[a] Missing value = 4

background, while 10% came from non-western ethnic minorities. Parental education level was more or less normally distributed. For the purpose of this article only children whose parents had completed all questionnaires were included in the analyses. Description of the primary diagnoses, age at screening and IQ scores are summarised in Table 1. On average, diagnosis was established 3.5 months (SD = 1.6) after screening.

Diagnostic Protocol

The assessments included a standardized parent–child play observation (*Emotional Availability Scales*; Biringen et al. 2000), a clinical psychiatric examination, a standardized behaviour observation using the *Autism Diagnostic Observation Schedule-Generic* (ADOS-G; Lord et al. 2000) and a structured and standardized parent interview, the *Autism Diagnostic Interview-Revised* (ADI-R; Lord et al. 1994; Le Couteur et al. 2003). All assessments were performed and administered by certified child psychologists or psychiatrists. The child's cognitive abilities were measured with the *Mullen Scales of Early Learning* (MSEL; Mullen 1995) in 62% of the cases or the *Psychoeducational Profile-Revised* (PEP-R; Schopler et al. 1990) in 38% of the cases. In this study, IQs based on the PEP-R were calculated as follows: (developmental age in months/chronological age in months) × 100. Language abilities were examined using the Reynell test for language comprehension (van Eldik et al. 1995) and a Dutch test for language production (Schlichting et al. 1995) or a Dutch

pre-verbal speech test (NNST; Zink and Lembrechts 2000), measuring pre-verbal skills like imitation, social babbling, and use of simple gestures.

Measures

The ESAT was part of the referral procedures. Parents filled out two additional questionnaires during the clinical assessments: the SCQ (Berument et al. 1999; Rutter et al. 2003) prior to the ADI(-R) interview and the CSBS-DP Infant-Toddler Checklist (Wetherby and Prizant 2002). These questionnaires only served research purposes and were not used in diagnostic evaluations.

The *ESAT*[1] (Dietz et al. 2006; Swinkels et al. 2006) consists of 14 easy-to-administer items measuring early social-communication skills, play and restricted and repetitive behaviour (e.g. eye-contact, facial expressions, interest in others, varied play and sensory interest), to be answered with *yes* or *no*. Children failing three or more items are considered at risk for ASD.

The *SCQ*, originally named the Autism Screening Questionnaire (ASQ), is a 40-item parent questionnaire designed and validated for use with individuals aged 4 and older (Berument et al. 1999; Rutter et al. 2003). The items are based on the ADI-R (Lord et al. 1994). Each item is checked as *yes* or *no*, and assigned a point rating of 1, indicating presence of abnormal behaviour or absence of normal behaviour, or 0 indicating typical behaviour. Item-1

---

[1] The complete list of items can be obtained from the first author.

is not included in the scoring, but determines if the child has enough language to score items on abnormalities in language. If the child is nonverbal, items 2–7 are left out. The cut-off for ASD is established at 15, but for younger children a cut-off of 11 has also been suggested (Allen et al. 2006; Corsello et al. 2007).

The *CSBS-DP* is a standardized more general screening tool with three components designed for screening and evaluation of communication and symbolic abilities of infants and toddlers (Wetherby and Prizant 2002). We only used the parent-questionnaire component (Infant-Toddler Checklist), to be referred to further as "CSBS-DP", that measures skills from three composites: (a) *Social* (emotion, eye gaze and communication), (b) *Speech* (sounds and words) and (c) *Symbolic* (understanding and object use) and asks about developmental milestones. Nineteen of the 24 items have the answer options: *not yet* (0 points), *sometimes* (1 point) and *often* (2 points). The remaining questions are on *how many* questions (e.g. about how many words or phrases does your child understand without gestures?) with 0 points if *none*, and 1–4 points for items containing number choices. The higher the cumulative score the lesser the chance of being at risk for ASD. Norms are available by 1 month intervals, from 6 up to and including 24 months.

The *CHAT* (Baron-Cohen et al. 1992) was not administered in its original form. Only three questions representing the main concepts of the CHAT included in the SCQ (item 22: *protodeclarative pointing*) and the CSBS-DP (item 4: *gaze following* and item 24: *pretend play*) were taken into consideration. These items will be referred to as the "CHAT-key-items". If a child failed on all three items, he/she was considered to be at high-risk for autism. Children who failed on protodeclarative pointing but were not included in the high-risk group were predicted to be at medium risk for autism.

Data Analysis

For each screening instrument group differences in mean sum scores between the youngest (8–24 months at screening) and oldest age group (25–44 months at screening) and between children with different diagnoses (Autism, ASD-other and non-ASD) were established using univariate analyses of variance with post-hoc analyses (with standard Bonferonni correction) and *t*-tests. Level of significance was defined as $P < .05$.

To assess and compare the discriminative power of the screening instruments in distinguishing ASD-subjects from non-ASD-subjects in the total and separate age groups, different indices of diagnostic accuracy (outcome measures) were calculated as shown in Table 2. Also, Receiver-Operator-Characteristic (ROC) Area-Under-the-Curve (AUC) analyses were run, in order to investigate the ability of the instruments to predict the presence of ASD diagnoses. A ROC curve can be drawn by plotting the sensitivity (Se) against 1-the specificity (1-Sp) for every potential cut-off score of a test. The discriminative potential of a test increases as the curve comes closer to the upper left corner of the diagram. If the curve touches the ultimate upper left corner, sensitivity and specificity are 100%. The AUC is used as a measure for the discriminative potential of a diagnostic test, or in our case, a screening instrument. Only AUCs of .80 or higher indicate a reasonable to good concordance between the scores of the screen and the golden standard diagnosis.

In the item analyses, SCQ-items that are not applicable to nonverbal children were treated as missing values. Replacing all missing values by a score based on the number of positive-for-autism responses divided by the number responded to (a method suggested by Eaves et al. 2006b) did not result in different outcomes neither for the verbal nor for the nonverbal children at two different cut-off scores (11 and 15). Hence, in further analyses missing values were disregarded. With reference to the CSBS-DP, for children older than 24 months of age the American 24 months norms were used, as norms are not available for children older than 24 months. In establishing the AUCs, CSBS-DP-total-scores were coded reversely.

To establish and compare the usefulness of individual screening items, the same indices of diagnostic accuracy were calculated for each item in all age groups (total age group, 8–24 months, 25–44 months) as was done for the whole instruments. In addition, we calculated Phi-values, a

Table 2 Calculation of test properties

Screen result	True clinical diagnosis		
	Diagnosis present (ASD)	Diagnosis not present (non-ASD)	Total
Screen positive	A	B	A + B
Screen negative	C	D	C + D
Total	A + C	B + D	N(A + B + C + D)

Sensitivity (Se) = A/(A + C), Specificity (Sp) = D/(B + D), Positive Predictive Value (PPV) = A/(A + B), Negative Predictive Value (NPV) = D/(C + D)

measure of association of two variables calculated from 2 × 2 tables (Siegel and Castellan 1988, p. 232). Phi-values represent Chi-squared values corrected for the number of observations, with values varying between −1 and +1 that can be interpreted with the same rule of thumb that is used for correlation coefficients (−1.0 to −0.7 strong negative association, −0.7 to −0.3 weak negative association, −0.3 to +0.3 little or no association, +0.3 to +0.7 weak positive association, +0.7 to +1.0 strong positive association). As the CSBS-DP items have three or more answering options, they had to be dichotomized in order to make calculating the Phi-values and the indices of diagnostic accuracy possible. For items with three answering options the measures were calculated with 'not yet and sometimes' versus 'often' and with 'not yet' versus 'sometimes and often'. Likewise, for the *how many* questions measures were calculated for different combinations of answering options.

Results

Differences in Mean Sum Scores between Diagnostic Categories and Age Groups

Table 3 shows mean sum scores per screening instrument for the three diagnostic groups and different age groups. As expected, children with the core syndrome (Autism) had the highest mean scores for the ESAT, SCQ and CHAT-key-items and the lowest mean score for the CSBS-DP, whereas non-ASD children had the lowest mean scores on the ESAT, SCQ and CHAT-key-items and the highest mean score on the CSBS-DP.

For the ESAT, mean sum scores did not differ between age groups or between diagnostic groups. For the SCQ no age effect was found, though mean sum scores did differ between diagnostic groups ($F(2,232) = 12.18$, $P < .001$). For the CSBS-DP and the CHAT-key-items diagnostic group effects were found (CSBS-DP: $F(2,232) = 25.69$, $P < .001$; CHAT-key-items: $F(2,232) = 19.02$, $P < .001$) as well as age effects (CSBS-DP: $F(1,232) = 26.06$, $P < .001$; CHAT-key-items: $F(1,232) = 12.13$, $P < .01$). Age effects, as displayed in Table 3, represent younger children's lower mean scores on the CSBS-DP and higher mean scores on the CHAT-key-items in comparison with older children. Differences in mean scores among diagnostic groups are specified in the notes of Table 3. It should be noted that mean scores were found to differ between the autism and ASD-other group or between the autism and non-ASD group only. Yet no differences were found between the ASD-other group and the non-ASD group. In addition, differences in mean sum scores of verbal versus nonverbal children on the SCQ—not in table—were significant ($t(236) = 2.87$, $P < .01$), with a mean sum score of 14.82 (SD = 5.79;

Table 3 Mean scores (± SD) on the ESAT, SCQ, CSBS-DP and CHAT-key-items per age group for three different diagnostic categories

	Austism		ASD-other		Non-ASD	
	n = 103	(SD)	n = 57	(SD)	n = 78	(SD)
EAST						
Age group 8–24 months	6.3	(3.3)	4.9	(2.6)	4.9	(1.8)
Age group 25–44 months	6.0	(2.7)	5.9	(3.0)	5.2	(2.4)
Total age group 8–44 months	6.1	(2.9)	5.7	(3.0)	5.2	(2.3)
SCQ						
Age group 8–24 months	18.9[b]	(5.9)	15.1	(5.6)	13.5	(5.9)
Age group 25–44 months	18.2[a,b]	(5.3)	14.6	(5.9)	13.5	(5.4)
Total age group 8–44 months	18.4[a,b]	(5.4)	14.7	(5.9)	13.5	(5.4)
CSBS-DP						
Age group 8–24 months	23.6[b]	(11.5)	30.0	(12.0)	36.0	(13.5)
Age group 25–44 months	31.0[a,b]	(9.9)	42.0	(7.5)	43.2	(7.3)
Total age group 8–44 months	29.0[a,b]	(10.8)	40.5	(9.0)	42.2	(8.7)
CHAT-key-items						
Age group 8–24 months	1.9	(1.2)	1.1	(1.3)	0.9	(0.9)
Age group 25–44 months	1.4[a,b]	(1.1)	0.4	(0.8)	0.3	(0.5)
Total age group 8–44 months	1.5[a,b]	(1.1)	0.5	(0.9)	0.4	(0.6)

Age groups are based on age at screening with the ESAT. On average, children were 2.6 months older (SD = 1.7) at the time of the SCQ and CSBS-DP filling out

[a] Significant mean difference between Autism and ASD-other, $P < .05$ or better

[b] Significant mean difference between Autism and Non-ASD, $P < .05$ or better

$n = 123$) for verbal children and a *higher* mean sum score of 16.99 (SD = 5.89; $n = 115$) for nonverbal children.

Analyses of Whole Instruments

The various indices of diagnostic accuracy of the different screening instruments are summarized in Table 4 for the total age group and for two different age groups separately. The clinical significance of the various indices of diagnostic accuracy was evaluated by Cicchetti et al. (1995) and established as: <0.70 = poor; 0.70–0.79 = fair; 0.80–0.89 = good; 0.90–1.00 = excellent. Applying these criteria to the results in Table 4, not a single screening instrument, at the whole age range, or for the younger and older subgroups, demonstrated acceptable diagnostic accuracy for all four indices (Se, Sp, NPV, PPV). In fact, the most that occurred is that only two of the indices meet the 0.70 minimum. In addition, whereas the AUCs of all instruments turned out to be poor to fair only (with values between 0.58 and 0.74), none of the existing screening instruments seemed to have satisfactory discriminative power in differentiating between ASD and non-ASD in a

high-risk population at a very young age. Also, the use of PPVs is limited as the base-rate of ASD in the total sample is high (0.67 in the total age group). However, separate test properties for different measures showed certain strengths. With respect to the *total age group* and the *oldest age group* the sensitivity of the ESAT and the SCQ using a cut-off of 11 was high, ranging from 0.83 to 0.89. The PPV and specificity in these groups were especially high for the CHAT-key-item (using both the high-risk criteria alone and in combination with the medium-risk criteria), with outcome measures ranging from 0.87 to 1.00. With respect to the *youngest age group*, the sensitivity of the ESAT and SCQ with a cut-off of 11 was also high (0.86 and 0.89, respectively), whereas the sensitivity of the CSBS-DP in this age group appeared to be very high as well: 0.91. As in the total and oldest age group, the PPV of the CHAT-key-items in the youngest age group had high scores, using the high-risk criteria as well as the high- and medium-risk criteria together (0.93 and 0.88, respectively). But the specificity in this young age group was substantially lower than in the oldest age group when the high- and medium risk criteria are used in combination (0.73). The specificity of the CHAT-key-items using the high-risk criteria alone was

Table 4 Outcome measures of the three screening instruments and the CHAT-key-items for the total group and for two age groups separately

Screening instrument	N	PPV	NPV	Se	Sp	AUC	95% Confidence interval (AUC)
Total age group: 8–44 months							
ESAT 14-items (cut-off = 3)	238	0.68	0.37	**0.88**	0.14	0.58	0.50–0.65
SCQ (cut-off = 11)	238	0.71	0.47	**0.84**	0.28	0.67	0.60–0.74
SCQ (cut-off = 15)	238	0.79	0.48	0.66	0.64	0.67	0.60–0.74
CSBS-DP	238	0.78	0.50	0.71	0.59	0.73	0.66–0.80
CHAT key-items (H)	238	**0.97**	0.37	0.18	**0.99**	0.67	0.60–0.74
CHAT key-items (H + M)	238	**0.88**	0.45	0.48	**0.87**	0.67	0.60–0.74
Age: 8–24 months							
ESAT 14-items (cut-off = 3)	46	0.75	0.17	**0.86**	0.09	0.61	0.44–0.77
SCQ (cut-off = 11)	46	0.79	0.43	**0.89**	0.27	0.71	0.54–0.88
SCQ (cut-off = 15)	46	**0.84**	0.40	0.74	0.55	0.71	0.54–0.88
CSBS-DP	46	**0.84**	0.63	**0.91**	0.45	0.74	0.57–0.90
CHAT key-items (H)	46	**0.93**	0.32	0.40	**0.91**	0.73	0.57–0.90
CHAT key-items (H + M)	46	**0.88**	0.40	0.66	0.73	0.73	0.57–0.90
Age: 25–44 months							
ESAT 14-items (cut-off = 3)	192	0.66	0.42	**0.89**	0.15	0.57	0.49–0.65
SCQ (cut-off = 11)	192	0.68	0.48	**0.83**	0.28	0.66	0.58–0.74
SCQ (cut-off = 15)	192	0.77	0.49	0.63	0.66	0.66	0.58–0.74
CSBS-DP	192	0.76	0.49	0.66	0.61	0.71	0.64–0.79
CHAT key-items (H)	192	**1.00**	0.38	0.12	**1.00**	0.66	0.59–0.74
CHAT key-items (H + M)	192	**0.88**	0.45	0.42	**0.90**	0.66	0.59–0.74

Age groups are based on age at screening with the ESAT. The 'sensitivity' and 'specificity' are related to the percentage of children about whom there is already some concern about ASD; since we have no information about the screen negative children who were not referred, and whether they are true or false screen negative, true sensitivity or specificity cannot be known. *PPV* Positive Predictive Value; *NPV* Negative Predictive Value; *Se* Sensitivity; *Sp* Specificity; *AUC* Area-Under-the-Curve; *H* using the High risk criteria; *H + M* Using the High and Medium risk criteria; Indices of diagnostic accuracy with good to excellent values are printed in bold (Cicchetti et al. 1995)

0.91. In addition, for this youngest age group the PPV of the SCQ using a cut-off of 15 and of the CSBS-DP were notably high, namely 0.84 each.

Analyses of Single Items

Table 5 includes Phi-values and indices of diagnostic accuracy (PPV, NPV, Sensitivity and Specificity) for all individual items in the whole age group. The same measures were calculated for the two age groups separately, but are not presented in the table.

In sum, in all age groups a considerable number of associations between item classification and clinical diagnosis, as expressed by Phi-values are significant but weak (with a maximum of 0.35). In addition, the indices of diagnostic accuracy demonstrated, that also at the level of individual screening items, neither in the total age group nor in the two age groups separately, any of the items reached the 0.70 minimum for all four indices (Se, Sp, NPV, PPV; Cicchetti et al. 1995). However, various items did show specific strengths. In general, specificities of items appeared stronger than sensitivities. Overall, NPVs were poor while the PPVs showed higher values, but are, yet again, of limited value, as the base-rate of ASD is high.

As indicated by the relatively strongest Phi-value-based associations, items on *joint attention skills*, like 'Attracting attention' (CSBS-DP 5, 6, & 14), 'Showing', 'Giving', and 'Directing attention' (ESAT 9, SCQ 28, CSBS-DP 8, 9, & 10) and like 'Following attention' (CSBS-DP 4) performed relatively well. Items indicating *reciprocal social interaction* like 'Eye gaze' (SCQ 26), 'Checking' (CSBS-DP 2), 'Directing smile to others' (ESAT 12, CSBS-DP 3), 'Interest in children or adults' (ESAT 10, SCQ36), and 'Offering comfort' (SCQ 31) as well as items about *use of gestures*, like 'Nodding to mean "Yes"'(SCQ 24), 'Head shaking to mean "No"'(SCQ 25), 'Pointing' (SCQ 22), and 'Waving bye-bye' (CSBS-DP 11) stood out as relatively good discriminating items. Furthermore, items like 'Reacting when spoken to' (ESAT 14) and 'Imitation' (SCQ 21) and items indicating understanding and use of words or sounds in *verbal communication* (SCQ 2 and 20, CSBS-DP 15, 16, 17, 18, & 20) did relatively well. Finally, some items on *play* (ESAT 2, SCQ 40, CSBS-DP 24) and *use of objects* (ESAT 1, CSBS-DB 22 & 23) and some on *restricted, repetitive and stereotyped behaviour* (SCQ 7, 8, 11, 12, & 15) showed relatively good discriminating value.

In the item analyses of all instruments, the oldest age group (25–44 months) was virtually similar to the total age group. For the youngest age group (8–24 months), more 'mature' *joint attention skills* like 'Showing and directing attention' have obviously less discriminative value than 'earlier' joint attention skills like 'Following attention' (CSBS-DP 4) and 'Using words/sounds to get attention'

(CSBS-DP 14). Whereas *'gesture-items'* that were emphasized for the whole age-group performed relatively well in the youngest age group too, items that refer to *reciprocal social interaction* that discriminate specifically well in the youngest group are 'Interest in children or adults' (ESAT 10, SCQ 36) and 'Checking' (CSBS-DP 2). Furthermore, 'Imaginative play'(CSBS-DP 24), 'Repetitive use of objects' (SCQ 12) and 'Hand and finger mannerisms' (SCQ 15) stood out as relatively good discriminating items in the very young children.

With regard to the CHAT-key-items, 'Following pointing' showed excellent specificity in all age groups, but sensitivity was very poor. 'Pointing to express interest' had excellent specificity in the total and oldest age group, fair specificity in the youngest age group, but poor sensitivity in all age groups. 'Imaginative play' was an item with good specificity and poor sensitivity in the oldest age group, but excellent sensitivity and poor specificity in the youngest age group.

Calculations on outcome measures using the 'best' SCQ-items, with positive and significant Phi-values only, and as summarized in Table 6, showed that using a selection of SCQ-items in general counts for improved specificity, with sensitivity remaining 0.75 and above. For the youngest age group, the AUC of 0.88 (95% CI 0.77–0.99) using only 8 items was surprisingly well.

Discussion

Strictly speaking, not one single screening instrument investigated appears to meet standards for a satisfactory prediction of an ASD diagnosis in our high-risk sample of very young children, as no instrument demonstrates acceptable diagnostic accuracy for all four indices (Se, Sp, PPV, NPV), at the whole age range, or for the younger and older subgroups. The balance between the sensitivity and specificity of the screens, as expressed by the AUCs, is fair at the most (Cicchetti et al. 1995). In addition to the general inaccuracy of the screens examined, none of the instruments performs clearly better than another in differentiating between ASD and non-ASD. However, it would be too simple and premature to dismiss all these instruments altogether, as each instrument shows specific strengths that should be considered in making decisions about which instrument to use for which purpose. Some caution in interpreting and comparing the results of the three screeners is warranted, as children were included in this study largely by screening positive on one of them (ESAT).

The value of a screening instrument based on its PPV needs to be viewed in the context of the base-rate of the condition studied. Since our study design had led to a high

Table 5 Outcome measures of all individual items of the three screening instruments and the CHAT-key-items for the total group

	Phi	PPV	NPV	Se	Sp
ESAT					
1. Interest in different toys	−0.05	0.56	0.32	0.18	0.78
2. Varied play	−0.13*	0.63	0.24	0.62	0.25
3. Emotions understandable	−0.11	0.61	0.28	0.38	0.50
4. Reaction to sensory stimuli	0.01	0.68	0.33	0.42	0.59
5. Facial emotional expressions	0.03	0.69	0.35	0.37	0.67
6. Eye contact	0.10	0.71	0.39	0.68	0.42
7. Attacts attention	0.13	0.76	0.36	0.38	0.75
8. Stereotypical movement	0.03	0.68	0.35	0.57	0.47
9. Showing and directing attention	0.19**	0.76	0.42	0.56	0.64
10. Interest in other children or adults	0.22**	**0.80**	0.41	0.47	0.76
11. Likes cuddling	0.02	0.68	0.34	0.48	0.54
12. Smile directed to others	0.13*	0.75	0.38	0.39	0.74
13. Enjoys social play	0.10	0.76	0.34	0.25	**0.84**
14. Reacts when spoken to	0.18**	0.76	0.41	0.51	0.68
SCQ					
Reciprocal Social Interaction					
9. Inappropriate facial expression	0.13*	0.77	0.38	0.30	**0.82**
10. Use of other's body	0.11	0.74	0.37	0.39	0.73
19. Friends	0.11	0.69	0.44	**0.85**	0.24
26. Eye gaze	0.15*	0.74	0.39	0.54	0.62
27. Social smiling	0.05	0.71	0.33	0.29	0.76
28. Showing and directing attention	0.19**	**0.83**	0.38	0.32	**0.86**
29. Offering to share	0.12	0.74	0.38	0.50	0.63
30. Seeking to share anjoyment	0.13	0.76	0.37	0.40	0.73
31. Offering comfort	0.26***	0.79	0.46	0.61	0.68
32. Quality of Social overtures	0.12	0.79	0.35	0.21	**0.88**
33. Range of facial expression	0.15*	**0.82**	0.36	0.23	**0.89**
36. Interest in children	0.18**	0.76	0.40	0.49	0.69
37. Response to other children	0.05	0.71	0.33	0.39	0.66
39. Imaginative play with peers	0.10	0.69	0.42	**0.81**	0.27
40. Group play	0.16*	0.72	0.45	0.77	0.38
Communication					
2. Conversation	−0.23*	0.63	0.58	0.66	0.55
3. Stereotyped utterances	−0.07	0.51	0.42	0.67	0.27
4. Inappropriate questions	0.01	0.55	0.46	0.09	**0.91**
5. Pronoun reversal	−0.05	0.50	0.48	0.28	0.68
6. Neologisms	0.01	0.54	0.47	0.38	0.63
20. Social chat	0.15*	0.72	0.43	0.69	0.46
21. Imitation	0.21**	**0.82**	0.40	0.39	**0.82**
22. Pointing to express interest[d]	0.34***	**0.88**	0.44	0.48	**0.87**
23. Gestures	0.03	0.68	0.35	0.69	0.34
24. Nodding to mean 'yes'	0.35***	0.79	0.55	0.77	0.58
25. Head shaking to mean 'no'	0.23***	0.79	0.43	0.55	0.69
34. Imitative social play	0.08	0.72	0.36	0.42	0.66
Restricted, Repetitive, Stereotyped Behaviour					
7. Verbal rituals	0.18*	0.60	0.59	0.77	0.39
8. Compulsions and rituals	−0.15*	0.62	0.24	0.54	0.30
11. Unusual preoccupations	0.16*	0.76	0.39	0.47	0.69
12. Repetitive use of objects	0.26***	0.77	0.47	0.68	0.60

Table 5 Continued

	Phi	PPV	NPV	Se	Sp
13. Circumscribed interests	0.09	0.73	0.36	0.36	0.73
14. Unusual sensory interests	0.09	0.74	0.35	0.27	**0.81**
15. Hand and finger mannerisms	0.22**	0.78	0.42	0.57	0.66
16. Complex body mannerisms	0.07	0.71	0.36	0.53	0.54
35. Imaginative play	0.10	0.71	0.38	0.49	0.69
Not in algorithm					
17. Self injury	0.03	0.70	0.33	0.33	0.70
18. Unusual attachment to objects	−0.06	0.62	0.31	0.21	0.74
38. Attention to voice	0.11	0.73	0.37	0.53	0.58
CSBS-DP Infant-Toddler Checklist					
Emotion an Eye gaze					
1. Understandable emotions	0.05[b]	0.73	0.33	0.10	**0.92**
2. Checking	0.17**[b]	0.71	0.53	**0.91**	0.21
3. Directing smile to others	0.14*[a]	**0.92**	0.34	0.08	**0.99**
4. Following pointing[d]	0.25***[b]	**0.93**	0.38	0.24	**0.96**
Communication					
5. Trying to get attention in order to get help	0.14*[b]	**0.89**	0.35	0.11	**0.97**
6. Trying to get attention of others	0.16*[a]	**0.84**	0.35	0.13	**0.95**
7. Making others laugh	0.20**[b]	**0.80**	0.40	0.44	0.77
8. Directing attention	0.33***[b]	**0.88**	0.44	0.46	**0.87**
Gestures					
9. Giving	0.18**[a]	**0.82**	0.36	0.23	**0.90**
10. Showing	0.25***[b]	**0.85**	0.40	0.40	**0.86**
11. Waving bye-bye	0.26**[a]	**0.89**	0.38	0.25	**0.94**
12. Pointing	0.28***[b]	**0.91**	0.40	0.32	**0.94**
13. Nodding to mean 'yes'	0.33***[b]	0.74	0.64	**0.90**	0.36
Sounds					
14. Using words/sounds to get attention	0.21**[a]	**0.94**	0.35	0.11	**0.99**
15. Stringing sounds	0.27***[a]	**0.80**	0.46	0.63	0.66
16. Using consonant sounds	0.19**[c]	0.73	0.45	0.72	0.47
Words					
17. Use of meaningfully words	0.29***[b]	**0.92**	0.41	0.31	**0.95**
18. Putting two words together	0.33***[b]	**0.86**	0.45	0.51	**0.83**
Understanding					
19. Attention to voice	0.10[b]	**0.90**	0.34	0.06	**0.99**
20. Understanding language without gestures	0.18**[b]	**0.87**	0.37	0.21	**0.94**
Object use					
21. Playing with different toys	0.06[b]	**0.80**	0.33	0.05	**0.97**
22 Appropriate use of objects	0.22**[c]	**0.90**	0.38	0.22	**0.95**
23. Stacking blocks	0.19**[b]	**0.91**	0.37	0.15	**0.97**
24. Imaginative play[d]	0.19**[b]	0.79	0.40	0.44	0.76

The 'sensitivity' and 'specificity' are related to the percentage of children about whom there is already some concern about ASD; since we have no information about screen negative children who were not referred, and whether they are true or false screen negative, true sensitivity or specificity cannot be known. Indices of diagnostic accuracy with good to excellent values are printed in bold (Cicchetti et al. 1995)

* $P < .05$; ** $P < .01$; *** $P < .001$

[a] Sometimes = Positive for autism, in other words: "not yet" and "sometimes" versus "often"

[b] Sometimes = Negative for autism, in other words: "not yet" versus "sometimes" and "often"

[c] None to 8 versus "more than 8"

[d] CHAT-key-item

risk sample which included 67% ASD diagnoses, this consideration could easily lead to devaluating the PPV's found for the various instruments. Taking this into account, the ESAT PPV in the youngest age group was fair (0.75), whereas for the older age group it just did not reach the 0.70 threshold. The CHAT-key-items (high risk criteria) showed excellent PPV, while the performance of both the CSBS-DP and the SCQ was less satisfactory. Overall, the relatively high PPVs established in combination with the low NPVs for all instruments means that a positive screening result is very useful (a screened positive subject has a high chance of actually having ASD), while a negative screening result is not (a screened negative subject has a low chance of actually not having ASD).

Table 6 Outcome measures in different age groups for a selection of 'best' SCQ-items with significant and positive Phi-values

	N	Se	Sp	AUC	95% Confidence interval (AUC)
Total age group: 8–44 months					
SCQ (16 items[a], cut-off = 4)	238	0.89	0.33	0.74	0.68–0.81
SCQ (16 items[a], cut-off = 4)	238	0.82	0.49	0.74	0.68–0.81
Age: 8–24 months					
SCQ (8 items[b], cut-off = 3)	46	0.91	0.55	0.88	0.77–0.99
SCQ (8 items[b], cut-off = 4)	46	0.83	0.82	0.88	0.77–0.99
Age: 25–44 months					
SCQ (15 items[c], cut-off = 4)	192	0.83	0.45	0.74	0.68–0.81
SCQ (15 items[c], cut-off = 5)	192	0.75	0.60	0.74	0.68–0.81

Age groups are based on age at screening with the ESAT. *Se* Sensitivity; *Sp* Specificity; *AUC* Area-Under-the-Curve

[a] Sixteen items included: 7, 9, 11, 12, 15, 20, 21, 22, 24, 25, 26, 28, 31, 33, 36 and 40

[b] Eight items included: 12, 15, 22, 24, 25, 36, 37 and 39

[c] Fifteen items included: 2, 9, 10, 12, 15, 20, 21, 22, 24, 25, 26, 28, 31, 33 and 36

With regard to sensitivities and specificities, in instruments developed for screening a certain condition in a high-risk population, only a minimum of cases with that condition can be missed. It may thus be substantiated that the *sensitivity* of a test is of more value than the specificity. As a consequence of the study design, we a priori expected higher estimates of the ESAT sensitivity and lower estimates of the sensitivity of the other screeners. However, for children of 24 months and younger our study showed the highest sensitivity for the CSBS-DP (0.91). This screener would therefore be a good choice in screening for ASD within this young age group. The ESAT and the SCQ (cut-off 11), both showing high sensitivity as well, could be perceived as good alternatives. In general, high sensitivities of screeners appeared in combination with low specificities, i.e. the proportion of false positives was high. However, the outcome for the CHAT-key-items was reversed; consistent with findings by Scambler, Rogers, and Wehner (2001), these items showed excellent specificity, especially in the oldest age group and using the high-risk criteria. As it combines a high specificity with a high PPV, the CHAT-key-items could be of use for clinicians and researchers wishing to exclude non-ASD subjects. Nonetheless, the outcomes relating to the CHAT-key-items should be interpreted with caution, because in our study the CHAT was not applied in its original form. In general, the strengths and weaknesses of the various instruments must be taken into consideration in deciding which instruments to use for which aim.

Considering the influence of age, in our study no big differences in discriminative power between instruments appeared in general, though the CSBS-DP seems more applicable to children aged 24 months and younger. The fact that norms for the CSBS-DP are only available until the age of 24 months, which made us decide to use the 24 months norms also for children up to 44 months of age, could have influenced outcome measures for the oldest age group.

Most children referred for further assessment were screen positive on the ESAT (87%). A minority was screen negative (13%), but was referred because of clinical concerns. Whereas about 67% (160 out of 238) indeed had ASD, and other non-ASD subjects all had substantial developmental problems that needed professional help, only two referred children appeared to function normally. Obviously, screening with the ESAT enables us to differentiate between normal and abnormal functioning in an age range from 8 to 44 months at least. In itself, this is a remarkable finding; the ESAT was originally developed for screening at 14 months, but also seems of value in older age groups.

With reference to the SCQ, there is an ongoing discussion in the literature about the optimal cut-off for young children. Consistent with previous research, our young sample scored lower on the SCQ than children roughly over 8 years tend to do (Allen et al. 2006; Berument et al. 1999; Corsello et al. 2007). Considering this optimal cut-off for young children, Corsello et al. studied the SCQ used as a secondary screening tool in a young age group (<5 years, N = 201). Using a cut-off of 15 they found a sensitivity of 0.68 with a specificity of 0.74. Using a cut-off of 11 would increase the sensitivity to 0.80, with specificity decreasing to 0.60. Allen et al. (2006) also used the SCQ as a secondary screening instrument with a cut-off of 15 and 11, and found a sensitivity of 0.56 and 0.89 and a specificity of 0.29 and 0.29 respectively in a group of children aged 24–36 months (N = 16). In addition, Wiggins et al. (2007) found a high sensitivity (0.89) together with a surprisingly high specificity (0.89) while using a cut-off of 11 in a clinical sample referred for early intervention (N = 37, age-range 17–45 months). In a recent study, Snow and

Lecavalier ([2008](#)) suggested using a cut-off of 13. Applying this cut-off, they found a sensitivity of 0.85 and a specificity of 0.40 in a sample of 65 children aged 30–70 months and referred for possible ASD. One can derive from our data that in the total age group (8–44 months) as well as in the separate age groups both sensitivity and specificity of the SCQ with a cut-off of 15 are poor. Using a cut-off of 11, sensitivity increases to 0.83 and above, depending on the age group, but specificity decreases to a very low level (0.27 or 0.28). When only a combination of items with positive and significant Phi-values (although these individual values indicate weak associations) is used as in the shorter version of the SCQ, this would help improving the specificity (with sensitivity remaining above 0.80) as compared to using the complete instrument, especially for the age group 8–24 months. Somewhat similar suggestions for improving the SCQ have been put forward by Eaves et al. ([2006a](#)). However, before the suggested alternatives can be used in clinical practice, these findings need to be replicated.

Another ongoing issue is about the exclusion of 6 items from the SCQ that are not applicable to nonverbal children. For example, Berument et al. ([1999](#)) found that removing these items for nonverbal children resulted in a statistically significant, but not meaningful difference for verbal and nonverbal individuals with ASD. They concluded that for the sake of simplicity, a cut-off of 15 would suit both verbal and nonverbal groups. However, Eaves et al. ([2006a](#), [b](#)) found that adjusting the total score for nonverbal children with a correction formula resulted in a better correlation between items and the total score, but changed the results of the screening only slightly (1 child changed categories). In general, as well as in the Corsello et al. ([2007](#)), in our study nonverbal children scored *higher* on the SCQ than verbal children, even though they had missing data on 6 verbal items. An explanation for this finding could be that nonverbal children with ASD may show more severe features of ASD than verbal children. Anyhow, as Corsello et al. ([2007](#)) also suggest, lowering the cut-off score may be a more effective strategy than adjusting scores in order to account for the skipped items for nonverbal children.

The analyses of individual items demonstrates that also no single item of any of the screens at any age achieves acceptable diagnostic accuracy for all four indices (Se, Sp, PPV, NPV) and the association between answering categories and diagnostic grouping remains weak. However, it is possible that items with disappointing discriminating value in the high-risk group examined will have specific value in differentiating between normal and abnormal functioning in a broader sense. Also in the current study, various items did show specific strengths, with most items showing higher specificities than sensitivities. In general, the properties of items in the oldest age group are comparable to those for the whole age group.

Discriminative properties of individual items in the youngest age group can differ somewhat more from their characteristics in the total age group, predominantly influenced by developmental aspects. An interesting issue is the usefulness of inventorying restricted, stereotyped and repetitive patterns of behaviour and/or items on the appropriate use of materials in screens. In both age groups items have been specified with either sensitivities of 0.70 and above or specificities of 0.70 and above could be of use in younger age groups. This is inconsistent with some studies that report repetitive and stereotypical behaviour to be less present in younger children compared to older children. Cox et al. ([1999](#)) for example, examined the stability of ASD clinical diagnosis and diagnosis derived from the ADI-R (Lord et al. [1994](#)) at 20 and 42 months of age. Abnormalities in the domain of repetitive and stereotyped behaviours were not reported at age 20 months in many children with autism, although they were present in most individuals with autism at 42 months. In a comparative study of four diagnostic instruments in toddlers, Ventola et al. ([2007](#)) also reported that many young children (age 16–31 months, $N = 45$) with autism spectrum disorder did not yet display more than one example of restricted interests, maintenance of sameness, or repetitive behaviours on the ADI-R. Lord ([1995](#)) however, found abnormalities such as hand and finger mannerisms, unusual sensory behaviours, unusual preoccupations and whole body mannerisms to be present at both younger and older time points. Further studies should clarify the discriminative value of repetitive and stereotypical behaviour in young children.

Limitations

A limitation to the study presented concerns the fact that the ESAT in combination with concerns in clinicians served as the prescreen. Therefore, one can not tell to which extend the SCQ, CSBS-DP, and CHAT-key-items would have falsely picked up non-ASD cases (false negatives) that the ESAT did not. As the screen negatives, unfortunately, have been lost to follow-up (except for the ones that were ESAT-screen negative but despite referred for further assessment) no truthful information could have been calculated on true sensitivity and true specificity. The 'sensitivity' and 'specificity' mentioned in this study are related to the percentage of children about whom there is already some concern about ASD; a very specific group. In addition, the way the sample was created and consequently its specific characteristics (e.g. high proportion of ASD-cases) influence the generalizability of results negatively. Another limitation of the study is that the ESAT was mostly filled out by a referrer in dialogue with parents, whereas the SCQ and CSBS-DP were filled out by the parents themselves, and on average 2.6 months (SD = 1.7) later than the ESAT. Finally, the interpretation of results is

hampered by the relatively small sample of children between 8 and 24 months of age.

Conclusion

From the literature we know that screening instruments for ASD are of value in discriminating between normal and abnormal development. However, the study presented reveals that screening instruments for ASD and their individual items have unsatisfactory value in discriminating between ASD and non-ASD within the group of children showing abnormal development. Much more research in tailoring more accurate second-level screening instruments for ASD needs to be done before they can be seen to have acceptable clinical utility. However, the question remains how much improvement can still be reached, as ASD-symptoms in infants and young children can be rather non-specific and hard to distinguish from symptoms of other developmental difficulties. In fact, to our less optimistic view, it may be unreasonable to expect second-level screens to discriminate ASD from other psychiatric or developmental disorders in young high-risk populations with greater precision. At this stage, complementary clinical awareness of primary care providers and mental health professionals remains extremely important in early detection. This paper provides new leads for interesting and powerful items in developing or adapting screening instruments. Yet, we should perhaps have to reconsider the aim of developing screening instruments only to discriminate between ASD and non-ASD in populations with severe developmental problems. Even if false-positive for the ASD – non-ASD paradigm, all these children with severe developmental difficulties (and their parents) are highly in need of thorough clinical attention, special management and early intervention.

References

Allen, C. W., Silove, N., Williams, K., & Hutchins, P. (2006). Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders, 37*, 1272–1278. doi:10.1007/s10803-006-0279-7.

Baron-Cohen, B., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? the needle, the haystack, and the CHAT. *The British Journal of Psychiatry, 161*, 839–843. doi:10.1192/bjp.161.6.839.

Baron-Cohen, S., Wheelwright, S., Cox, A., Baird, G., Charman, T., Swettenham, J., et al. (2000). Early identification of autism by the CHecklist for Autism in Toddlers (CHAT). *Journal of the Royal Society of Medicine, 93*(10), 521–525.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: diagnostic validity. *The British Journal of Psychiatry, 175*, 444–451. doi:10.1192/bjp.175.5.444.

Biringen, Z., Robinson, J. L., & Emde, R. N. (2000). Appendix A: the emotional availability scales (2nd ed.; an abridged infancy/early childhood version). *Attachment & Human Development, 2*(2), 251–270. doi:10.1080/14616730050085518.

Bryson, S. E., Rogers, S. J., & Fombonne, E. (2003). Autism spectrum disorders: early detection, intervention, education, and psychopharmacological management. *Canadian Journal of Psychiatry, 48*(8), 506–516.

Charman, T., & Baird, G. (2002). Practitioner review: diagnosis of autism spectrum disorder in 2- and 3-year old children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 43*(3), 289–305. doi:10.1111/1469-7610.00022.

Charman, T., & Baron-Cohen, S. (2006). Screening for autism spectrum disorders in populations. In T. Charman & W. Stone (Eds.), *Social and communication development in autism spectrum disorders: early identification, diagnosis and treatment* (pp. 63–87). New York: The Guilford Press.

Cicchetti, D. V., Volkmar, F., Klin, A., & Showalter, D. (1995). Diagnosing autism using ICD-10 criteria: A comparison of neural networks and standard multivariate procedures. *Child Neuropsychology, 1*, 26–37. doi:10.1080/09297049508401340.

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

Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J., et al. (1999). Autism spectrum disorders at 20 and 42 months of age: stability of clinical and ADI-R diagnosis. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 40*(5), 719–732. doi:10.1111/1469-7610.00488.

Dietz, C., Swinkels, S., van Daalen, E., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum disorder in children aged 14–15 months II: population screening with the Early Screening of Autistic Traits Questionnaire (ESAT). *Journal of Autism and Developmental Disorders, 36*(6), 713–722. doi:10.1007/s10803-006-0114-1.

Eaves, L. C., Wingert, H. D., Ho, H. H., & Mickelson, E. C. (2006a). Screening for autism spectrum disorders with the social communication questionnaire. *Journal of Developmental and Behavioral Pediatrics, 27*(2, Suppl), S95–S103. doi:10.1097/00004703-200604002-00007.

Eaves, L. C., Wingert, H. D., & Ho, H. H. (2006b). Screening for autism: agreement with diagnosis. *Autism, 10*(3), 229–242. doi:10.1177/1362361306063288.

Filipek, P. A., Accardo, P. J., Ashwal, S., Baranek, G. T., Cook, E. H., Jr, Dawson, G., et al. (2000). Practice parameter: screening and diagnosis of autism: report of the quality standards subcommittee of the American academy of neurology and the child Neurology Society. *Neurology, 55*(4), 468–479.

Johnson, C. P., Meyers, S. M., & Council on Children with Disabilities. (2007). Identification and evaluation of children with autism

spectrum disorders. *American Academy of Pediatrics, 120*(5), 1183–1215.

Le Couteur, A., Lord, C., & Rutter, M. (2003). *Autism diagnostic interview-revised*. Los Angeles, CA: Western Psychological Services.

Lord, C. (1995). Follow-up of two-year-olds referred for possible autism. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 36*, 1365–1382. doi:10.1111/j.1469-7610.1995.tb01669.x.

Lord, C., Rutter, M., & LeCouteur, A. (1994). Autism diagnostic interview-revised: A revised version of the diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659–685. doi:10.1007/BF02172145.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Jr, Leventhal, B. L., DiLavore, P. C., et al. (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. doi:10.1023/A:1005592401947.

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

Robins, D. L., Fein, D., Barton, M., & Green, J. A. (2001). 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, 31*, 131–144. doi:10.1023/A:1010738829569.

Rutter, M., Bailey, A., & Lord, C. (2003). *Social communication questionnaire*. Los Angeles, CA: Western Psychological Services.

Scambler, D., Rogers, S. J., & Wehner, E. A. (2001). Can the checklist for autism in toddlers differentiate young children with autism from those with developmental delays? *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 1457–1463. doi:10.1097/00004583-200112000-00017.

Schlichting, J. E. P. T., van Eldik, M. C. M., Lutje Spelberg, H. C., van der Meulen, S. J., & van der Meulen, B. F. (1995). *Schlichting test voor taalproductie*. Lisse: Swets Test Publishers.

Schopler, E., Reichler, R. J., Bashford, A., Lansing, M. D., & Marcus, L. M. (1990). *The psychoeducational profile revised (PEP-R)*. Austin, TX: Pro-Ed.

Siegel, B. (2004). *The pervasive developmental disorders screening test-II*. San Antonio, TX: Psychological Corporation.

Siegel, S., & Castellan, N. J. (1988). *Nonparametric statistics for the behavioral sciences* (2nd ed.). New York: McGraw-Hill Book Company.

Snow, A. V., & Lecavalier, L. (2008). Sensitivity and specificity of the modified checklist for autism in toddlers and the social communication questionnaire in preschoolers suspected of having pervasive developmental disorders. *Autism, 12*(6), 677–694. doi:10.1177/1362361308097116.

Stone, W. L., Coonrod, E. E., Turner, L. M., & Pozdol, S. L. (2004). Psychometric properties of the STAT for early autism screening. *Journal of Autism and Developmental Disorders, 34*, 691–701. doi:10.1007/s10803-004-5289-8.

Swinkels, S. H., Dietz, C., van Daalen, E., Kerkhof, I. H., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum disorder in children aged 14 to 15 months I: the development of the early screening of autistic traits questionnaire (ESAT). *Journal of Autism and Developmental Disorders, 36*(6), 723–732. doi:10.1007/s10803-006-0115-0.

van Eldik, M. C. M., Schlichting, J. E. P. T., Lutje Spelberg, H. C., van der Meulen, B. F., & van der Meulen, S. J. (1995). *Reynell test voor taalbegrip*. Amsterdam: Harcourt Test Publishers.

Ventola, P. E., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J., et al. (2007). Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders, 36*, 839–847. doi:10.1007/s10803-006-0128-8.

Wetherby, A., & Prizant, B. (2002). Communication and Symbolic Behavior Scales Developmental Profile First normed edition. Baltimore, MD: Paul H Brookes Publishing.

Wiggins, L. D., Bakeman, R., Adamson, L. B., & Robins, D. L. (2007). The Utility of the Social Communication Questionnaire in Screening for Autism in Children Referred for Early Intervention. *Focus on Autism and Other Developmental Disabilities, 22*(1), 33–38. doi:10.1177/10883576070220010401.

Zink, I., & Lembrecht, D. (2000). *Nederlandstalige nonspeech test*. Leusden: Acco.