



Universiteit Utrecht

Master Child and Adolescent Psychology

THESIS

**'The validation of three screening instruments for toddlers and pre-schoolers
to detect children with Autism Spectrum Disorders'**

Analysis of the reliability and validity of the Dutch version of the Social Responsiveness Scale
and Modified Checklist for Autism in Toddlers.

Daisy van 't Land, 3160491

July 2012

Jorieke Duvekot

Social Spectrum Study

Erasmus MC – Sophia Children's hospital

Judith Dubas

ABSTRACT

Early detection and intervention of ASD in young children has been shown to be effective for the development of social interaction, communication and a decrease of problems in the future. That is why in this study several screening instruments in detecting ASD in young children were examined. A population of 64 children, aged 18 to 30 months was screened using the Modified Checklist for Autism in Toddlers (M-CHAT) and 19 of them were also screened with a newly developed version of the Social Responsiveness Scale, for special use in toddlers (SRS18-30). A population of 84 children, aged 30 to 48 months was screened using another newly developed version of the SRS, for special use in pre-schoolers (SRS30-48). The brief nature of the three screenings instruments makes it user friendly and provides quick assessment of functioning. The main purpose of this study was to validate these Dutch translated screening instruments by establishing the reliability, sensitivity, specificity, predictive value and convergent validity. The questionnaires were filled out by parents of young children referred for various problems to several mental health centers in the South-West of the Netherlands. The M-CHAT had a sensitivity of 0.71 and specificity of 0.58. The PPV was 0.45 and the NPV 0.80. For the SRS18-30, using a cut-off score of 68, the sensitivity was 0.57 and the specificity and PPV were 1. The NPV was 0.8. The SRS30-48 missed a lot of children later diagnosed with an ASD. The sensitivity for this screen was 0.45, the specificity 0.61. More research into the diagnostic validity of the SRS30-48 showed that a screen result on the SRS30-48 is not significantly related towards an ASD diagnosis or not. The PPV was 0.42 and the NPV was 0.64. The conclusion of this study was that the screening instruments used in the age group of toddlers from 18-30 months, the M-CHAT and SRS18-30, can be used to identify cases of ASD. Care must be taken to the high amount of false positives for the M-CHAT and false negatives for the SRS18-30. The SRS30-48 is not that accurate, possibly due to other symptoms of ASD for pre-schoolers in comparison with children in the age of 4 to 18 (the age group for the original SRS).

CONTENT

Abstract	2
Content	3
Preface	4
Introduction	
Early concerns and early diagnosis	6
Benefits of early diagnosis and early intervention	7
Screening instruments for early detection of ASD	7
Psychometric properties	8
Recent studies into the M-CHAT	9
Recent studies into the SRS	10
Present study	12
Methods	
Population	13
Sample	13
Instruments	14
Procedure	16
Data analysis	16
Results	
Descriptives	18
Hypotheses	19
Discussion	23
References	28
Appendices	
I Nunnally & Bernstein (1994)	30
II Cicchetti, Volkmar, Klin, & Showalter (1995)	31
III Cohen (1988)	32
IV DSM classifications	33

PREFACE

This report is my master thesis for the conclusion of the master program Child and Adolescent Psychology at the Utrecht University. I wrote this thesis as a part of the broader Social Spectrum Study from the Erasmus MC – Sophia Children’s hospital. This is a multicenter study into ASD.

The writing of the thesis has gone through ups and downs. Especially in the beginning I experienced some troubles to start writing things down and have it be checked by my supervisor. In addition, I was the first time for me to write a report in English. But, after 10 months of hard work, especially in the last months, I am satisfied with the result.

I would thank my supervisor from the Erasmus MC, Jorieke Duvekot. For reading and revising the parts of my thesis, giving good advices and freeing up her time in the last weeks before the deadline. Also I would like to thank the Erasmus MC for giving me the opportunity to write my thesis at their research department. It was a good experience to visit different mental health centers to collect data and to experience how much is involved in good research. I would also thank my second reader Judith Dubas for her flexibility to read and grade my thesis as quickly as she could. Finally I would thank my friends, roommates and family for their patience and support.

INTRODUCTION

Autism Spectrum Disorders (ASDs) are also named Pervasive Developmental Disorders (PDD) in the Diagnostic Statistical Manual [DSM-IV TR]; American Psychiatric Association, 2000). Pervasive means that these disorders extensively influence various development domains of the child. A diagnosis of an ASD is, according the DSM-IV TR, made on the basis of deficits in three criteria domains: a) reciprocal social behavior (RSB); b) language development/communication; and c) repetitive/stereotypic behaviors or a restricted range of interests. The exact cause of ASDs is still unclear, but twin and family studies suggest that the origins are the result of genetic factors affecting brain development in very early life (Braird, Charman, Baron-Cohen, Cox, Swettenham, Wheelwright & Drew, 2000). That is a reason why the symptoms usually become manifest in infancy or early childhood, generally before the age of 30 months (Jónsdóttir, Saemundsen, Antonsdóttir, Sigurdardóttir, & Ólason, 2011; Pine, Luby, Abbacchi, & Constantino., 2006). Early detection of ASD symptoms and early diagnosis is important, because, in most cases, children cannot be provided intensive early intervention services until they have been diagnosed with an ASD. Therefore, it is important to use screening instruments to screen children on suspicion of ASD at a young age. This could help making an early diagnosis.

The present study investigates the reliability and validity of two screening instruments; the modified checklist for autism in toddlers (M-CHAT) and two newly developed versions of the social responsiveness scale (SRS): The SRS for toddlers in the age of 18 tot 30 months (SRS18-30) and the SRS for pre-schoolers in the age of 30 to 48 months (SRS30-48). Until now there has not been done any research into the psychometric properties of the Dutch translated version of the M-CHAT and newly developed SRS18-30 and SRS30-48. To see if these instruments are accurate in detecting children with and without ASD symptoms, the psychometric qualities are investigated. By doing research into the psychometric qualities of these two screening instruments, we hope to contribute to an accurate and early detection of the early signs of ASD, what can result in a better long-term prognosis for these young children (Kleinman et al., 2008).

Early concerns and early diagnosis

The most common first concerns of parents of children who were later diagnosed with ASDs were speech and language delays (Chakrabarti, 2009). Abnormal social development was usually the second major concern reported. These concerns from parents of children with ASDs about their children's development start mostly when the child is approximately between the 12 and 18 months (Barton, Dumont-Mathieu, & Fein, 2012). A study conducted in Iceland found this start of developmental concerns later. But still, the majority of parents (76,2%) had developmental concerns before their child has reached the age of three (Jónsdóttir et al., 2011). Even if their child was formal diagnosed after the age of six. This delay between concerns and a formal ASD diagnosis is supported by several studies. Barton et al. (2012) wrote in their report that most children with ASD not receive a formal diagnosis until an average age of four or later. This is partly because there is a delay of several months between the time that parents become concerned about their child's development and seeking professional advice (Jónsdóttir et al., 2011). But ASDs are unfortunately also difficult to detect in young children. Parents do have developmental concerns, but to make a diagnosis, the clinicians have to deal with some factors: 1) the presentation of the symptoms varies from child to child, 2) social and language deficits and delays may not be identified until the child has peer interaction, for example in preschool, 3) low incidence of the disorders leads to a low index of suspicion, and 4) the motor milestones (e.g. rolling over, sitting up, walking) in the child are usually unaffected (Robins, Fein, Barton, & Green, 2001).

Due to these factors, there is doubt about the validity of an early diagnosis. ASDs are difficult to detect in young children and therefore the chance to miss a child or falsely diagnose a child is high. However, a growing number of recent studies have demonstrated that a diagnosis of ASD is mostly stable over time, even when the diagnosis of ASD is made at age 2 (Kleinman, Robins, Ventola, Pandey, Boorstein, Esser, Wilson, Rosenthal, Sutera, Verbalis, Barton, Hodgson, Green, Dumont-Mathieu, Volkmar, Chawarska, Klin, & Fein, 2008). Howlin and Ashgarian (1999) examined a somewhat later age and suggest that the age of the earliest reliable ASD diagnosis was thought to be between 4 and 5 years. Less is known about the stability of an ASD diagnosis at very young age, but the desire to detect autism as early as possible is driving ongoing research to screen children before 2 years, and even in the first year of life (Kleinman et al., 2008).

Benefits of early diagnosis and early intervention

Despite the uncertainty about the validity of the diagnosis occasioned by early identification, the benefits of early diagnosis and detection of ASD outweigh the negative effects. Families consistently express the desire to be informed as early as possible (Barton et al., 2012). Especially children with severe developmental difficulties (and their parents) are highly in need of clinical attention, special management and early intervention (Oosterling, Swinkels, van der Gaag, Visser, Dietz, & Buitelaar, 2009). Early identification of young children and attendant early intervention are on the long-term associated with more positive outcomes in social interaction, communication and cognitive development (Robins et al., 2001; Barton et al., 2012). Another study suggests that early intervention may not only result in gains in functioning and improved life quality for the child and the family, but also results in long-term cost savings for parents and service systems (Motiwala, Gupta, & Lily, 2006). For that reason, it is meaningful to do more research into the early identification of ASD in childhood. By detecting young children with ASD accurately, you can give guidelines for the child and his surrounding family as soon as possible. By giving the parents and family information, tips and therapy they can learn how to deal with the behavioral problems and deficits of the child caused by ASD. This could prevent accumulating problems over time (Nadel, & Poss, 2006).

Screening instruments for early identification of ASD

The early onset of ASD in life has prompted many attempts to direct intensive intervention to children with ASD in toddlerhood. Understanding of the efficacy of these interventions has been hampered by a lack of quantitative tools to measure subtle improvements over time (Pine et al., 2006). In the last 15 years, a lot of research has been already conducted into developing methods and instruments for the early detection of ASD (Barton et al., 2012). The first questionnaire to help identify children who are at risk of developing social-communication disorders was developed by Baron-Cohen et al. in 1992. They named their instrument the Checklist for Autism in Toddlers (CHAT). This was the first screening tool for ASD in young children, up to 18 months of age (Barton et al., 2012). Screening instruments are mentioned to be quick, easy to use and readily interpretable. This screening tool was a 'level 1' screening tool, this means that the accessibility of the instrument is high; it can be administered to all children in care settings and is designed to differentiate children at risk for ASD's from the general population (Canal-Bedia, Carcía-Primo, Martín-Cilleros, Santos-Borbujo, Guisuraga-

Fernández, Herráez-García, del Mar Herráez-García, Boada-Muñoz, Fuentes-Biggi & Posada-de la Paz, 2010). The difference with the so-called 'level 2' screening tools is that the latter are normally applied in the diagnostic trajectory of mental health care centers. In addition, they take more time and training to administer, score and interpret. In recent years, several other autism-specified level 1 and 2 screening instruments for children have been developed. A few examples of level 1 screening instruments are the Social Responsiveness Scale (SRS), the Modified-Checklist for Autism in Toddlers (M-CHAT) and the Pervasive Developmental Disorders Screening Test-II (PDDST-II). Examples of level 2 autism-specified screening instruments are The Social Communication Questionnaire (SCQ) and the Screening Test for Autism in Toddlers (STAT) (Johnson et al., 2007). In the next paragraphs, we will focus on two screening instruments for ASD that are investigated in the present study: the M-CHAT and the SRS.

Psychometric properties

To make an accurate early screening it is important to have reliable, valid and easily accessible screening instruments to clearly identify children with ASD characteristics from those without. In general, a measure of behavior is said to be reliable if its results are repeatable when the behaviors are re-measured. Internal consistency of a measure defines the consistency of the results delivered in a test, ensuring that the various items measuring different areas within a single construct deliver consistent scores (Goodwin, 2005). Cronbach's alpha is known as an internal consistency estimate of reliability of test scores and will generally increase as the intercorrelations among test items increase. To interpret the Cronbach's alpha you can use the scale from Nunnally and Bernstein (1994; Appendix I). For a psychological instrument, reliability is essential. A measure is said to be valid if it measures what it had been designed to measure, and corresponds accurately to the real world (Goodwin, 2005). An important type of (construct) validity is the convergent validity. Convergent validity focuses on the relationship with other psychological constructs, which theoretically should be similar. High correlations between test scores of the different psychological tests would be evidence of a convergent validity (ter Laak & de Goede, 2005). Another kind of validity is the diagnostic validity; the ability of a test to indicate which individuals have the disorder and which do not. Diagnostic validity is measured by the parameters sensitivity and specificity (Zhu, Zeng, & Wang, 2010). Sensitivity is the ability of a screening instrument to accurately identify a high proportion of the children with ASD. Specificity is the extent to which the measure accurately classifies a high proportion of children who do not have ASD. Both parameters should ideally be high to ensure that children are correctly classified

(Zhu et al., 2010). Two other screening parameters are the positive predictive value (PPV) and negative predictive value (NPV). PPV is the proportion of individuals identified by the screener who actually have the disorder. A high PPV indicates a low false positive rate; the number of children who do not have an ASD diagnosis, but have a positive test result. Negative predictive value (NPV) refers to the proportion of individuals with negative screening results who do not have the disorder or diagnose. The PPV and NPV are strongly influenced by the prevalence rate of a disorder in the population under study (Barton et al., 2012).

Recent studies of the M-CHAT

Reliability

There has not been done a lot of research into the reliability of the M-CHAT independent of the developers. The study of Kleinman et al. (2008) determined the internal consistency of the M-CHAT with Cronbach's alpha (α) for all the items and for the subset of six critical items. Their result was an adequate internal reliability for both the entire instrument ($\alpha = .85$) and the six critical items ($\alpha = .84$). This finding is consistent with Robins et al. (2001), who found the internal consistency for the entire screener and for the six critical items to be $\alpha = .85$ and $\alpha = .83$, respectively.

Validity

The developers of the instrument Robins et al. (2001) found a sensitivity of 0.87, specificity of 0.99, and PPV of 0.80. In a study into the use of the M-CHAT for screening 228 young children (18 to 24 months age) for the presence of ASD symptoms in the Arab countries, the sensitivity was 0.86, the specificity 0.80 and the PPV was 0.88 (Eldin, Habib, Noufal, Farrag, Bazaid, Al-Sharbati, Badr, Moussa, Essali, & Gaddour, 2008). These findings were very similar to Robins et al. (2001). The conclusion of this study was that the M-CHAT is an effective tool to use in the early screening of autism in Arab countries. Kleinman et al. (2008) also examined the diagnostic properties of the M-CHAT. They screened 3793 children aged 16-30 months from low- and high-risk sources. The PPV for the combined sample was .36. After use of the follow-up telephone interview the PPV increased to .74. When separating the low- and high-risk group, the values on the screening parameters were unacceptable low for the low-risk group, the PPV was .11. After the follow-up telephone interview this increased to .65. For the high-risk group, the PPV was .76. The conclusion from Kleinman et al. (2008) in this study was; the M-CHAT can be

useful in detecting ASD in children in the age between 16–30 months, but then just for the group with children at high risk for ASD. The Spanish version of the M-CHAT (Canal-Bedia, et al., 2010) showed a sensitivity of 1; a specificity of 0.98; a PPV of 0.35; and a NPV of 1.

Recent studies of the SRS

Reliability

The first published research involving measurements of the original SRS, for children between 4 and 18 years, were about over 1900 children from age 4 until 15 years. The scores on the original SRS were continuously distributed in the population, generally unrelated to intelligence (IQ) and capable of distinguishing children with ASD from those with other child psychiatric disorders (Constantino, Przybeck, Friesen & Todd, 2000; Constantino & Todd, 2003). The original SRS was normed on a sample of 1636 children. To investigate the reliability of the instrument, the internal consistency was calculated (Constantino & Gruber, 2005). The internal consistency for the original SRS total scores ranged from .93 to .97, utilizing Cronbach's alpha, for the normative and clinical samples. The internal consistency of the original SRS is excellent (> .9; Appendix I).

The properties of the German adaptations of the original SRS were analyzed in a study with a large normative and clinical sample (Bölte, Poutska, & Constantino, 2008). From the 1436 individuals who participated, 838 individuals were in the normative and 527 in the clinical sample. The internal consistency in the normative sample was high, both for mothers and fathers (respectively 0.93 and 0.91). In the clinical sample, Cronbach's alpha was also high for combined parent ratings, individual parent ratings did not differ in this sample (both 0.97). Based on these results we can say that the original SRS has a good internal consistency. The internal consistency of the SRS18-30 and SRS30-48 has not yet been examined. The item content of the original SRS in comparison with the SRS18-30 and SRS30-48 differs only on the basis of developmental appropriateness of the wording for rating the behaviors of children in the respective age groups (Pine et al., 2006). That is the reason why in this study the expectation is to find a high internal consistency for the SRS18-30 and SRS30-48.

Validity

The convergent validity between the original SRS and screening instruments has frequently been studied by researchers. Charman, Baird, Simonoff, Loucas, Chandler, Meldrum,

& Pickles (2007) compared the original SRS to two other screening instruments: the Social Communication Questionnaire (SCQ) and the Children's Communication Checklist (CCC). All three screens differentiated highly between children in the ASD and non-ASD groups. In another study of Constantino (Constantino, Davis, Todd, Schindler, Gross, Brophy, Metzger, Shoushtari, Splinter, & Reich, 2003) the original SRS is compared with the Autism Diagnostic Interview-Revised (ADI-R). The original SRS has shown good convergent validity with the Autism Diagnostic Interview-Revised (ADI-R; $r = .7$). The study of Pine et al. (2006) represented a first attempt to test the utility of the SRS in a treatment study involving clinically ascertained preschoolers. In this study the sample consisted of two groups with children in the age range between 36 and 48 months. Group 1 was the randomized control trial and consisted of 23 children. All these children had an ASD diagnosis. In addition to these children, 5 typically developing children were recruited to serve as a contrast group. Group 2, the longitudinal study group, consisted of 45 preschoolers, 15 of them had a diagnosis of PDD-NOS, 13 had a diagnosis of autism and 17 were typically developing preschool siblings of non-ASD child psychiatric patients. Pine et al. (2006) examined the agreement between the SRS30-48 and several other instruments. There was substantial agreement between SRS scores and the Vineland Adaptive Behaviour Composite (Pearson's $r = -0.86$) and scores for social impairment on the Autism Diagnostic Interview-Revised ($r = 0.63$) (Pearson's r , see Appendix III). Pine et al. (2006) concluded that in preschool children, it is possible to obtain brief and reliable quantitative measurements of autistic social impairment, using SRS reports from parents and/or preschool teachers and daycare providers.

According to the screening parameters of the original SRS, the study of Charman et al. (2007) found a sensitivity of 0.75 and the specificity was 0.67. Both parameters were lower than those in the original validation study conducted by Constantino (2005). Sensitivity and specificity were in that study respectively 0.85 and 0.75. The PPV in this study was 0.63 and the NPV 0.81. However, these numbers changed in the different sub-samples of children (children with low IQ vs. high IQ and children with and without behavioral problems). Another study found that a significant number of children with mood and anxiety disorders obtain scores on the original SRS typically found in children with ASD diagnoses as they do not have a diagnosis of any form of ASD (Towbin et al., 2005). That was a reason for Aldridge et al. (2012) to gather further information regarding the use of the original SRS as a screening tool in the particular case of referral to an autism specific assessment service. The sensitivity of the parent form of the original SRS was .91. This means that 91% of the participants diagnosed with ASD scored in the ASD range. The specificity was .08, this is very low and means that 92% of the participants

without ASD fell within the range of ASD on the parent form of the original SRS; the number of false positives was very high. The PPV was 0.71 and the NPV was 0.25.

Aims and hypotheses of the present study

In this study the focus will be on the two autism screening instruments, the M-CHAT and the SRS. In this study two newly developed and translated versions of the SRS will be used: The SRS for toddlers in the age of 18 to 30 months (SRS18-30) and the SRS30-48 for pre-schoolers in the age of 30-48 months (SRS30-48). A lot of research has been done into the M-CHAT and the original SRS, but less research has been done into the newly developed SRS18-30 and SRS30-48 and the Dutch translation of the M-CHAT. In this report, the properties of the Dutch adaptation of the M-CHAT and newly developed SRS18-30 will be analyzed in a clinical sample of toddlers, from 18 to 30 months of age. The properties of the newly developed translated SRS30-48 will be analyzed in a clinical sample of pre-schoolers, from 30 to 48 months of age.

First, the reliability of the three tests will be calculated, by evaluating the internal consistency. Following the literature research, the expectation in this study is that the M-CHAT, SRS18-30 and the SRS30-48 are reliable screening instruments. The hypothesis is finding a good internal consistency ($\alpha \geq .8$, according to Nunnally and Bernstein; Appendix I) for all three screening instruments. Second, the diagnostic validity of the screening instruments was established by calculating the specificity, sensitivity, predictive values and convergent validity of both instruments. To establish convergent validity the SRS18-30 results was compared with the M-CHAT results and vice versa. A strong correlation means high converge and is desirable (Pearson's r value $\geq .7$; Appendix III). Because of the small number of studies into the SRS18-30, the best cut-off score to set the screening parameters is not yet examined. The research question is what will be the appropriate cut-off score of the newly developed SRS18-30? The SRS18-30 has fewer items than the SRS30-48 and the original SRS. The cut-off score in these screening instruments is a total score of 75 or higher. Therefore, in this study the assumption is that the cut-off score of the SRS18-30 will be below 75. In line with the literature the screening parameters for the SRS18-30, the M-CHAT and the SRS30-48 will be established at least as fair if sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) are $\geq .7$; according to Cicchetti et al., 1995. Appendix II).

METHODS

Population and Sample

2.1 Population

The sample for this study came from the broader Social Spectrum Study; this is a multicenter study of ASD. The sample used in this study consisted of children of 18 to 30 and 30 to 48 months of age who are consecutively referred to five youth mental health centers during a period of approximately one year and a half (between December 2010 and May 2012). These mental health centers were in the South West of the Netherlands, located in both rural and urban areas.

2.2 Sample

Age group 18-30 months:

A total number of 63 children aged 18 to 30 months were included in this study. The sample consisted of 46 boys and 18 girls. The mean age of the children was 34.16 months (SD = 9.19). The mean age for boys was 35.39 months (SD = 9.35) and the mean age for girls was 30,82 months (SD = 8.06). The difference in mean age for boys and girls was not significant ($t(61) = 1.782, p = 0.08$). The informant's gender consisted of 15 males and 45 females. Most informants were biological mothers (73.3 %).

Age group 30-48 months:

For the age group of children between 30 and 48 months, a total number of 84 children were included in this study. The sample consisted of 62 boys and 22 girls. The mean age of the children was 40.62 months (SD = 5.52). The mean age for boys was 40.79 months (SD = 5.79) and the mean age for girls was 40.14 months (SD = 4.78). The difference in mean age for boys and girls was not significant ($t(82) = .475, p = .636$). The informant's gender consisted of 8 males and 69 females. The informant's were mostly biological mothers (89.6 %).

Table 1. Description of participants' diagnoses and age in months at screening

Diagnosis	<i>N</i>		Age in months at screening ^a			
	18-30 months	30-48 months	18-30 months		30-48 months	
			M	SD	M	SD
Autism Spectrum Disorder (ASD)	21	31	34.9	7.2	41.0	4.8
Non-ASD	28	43	32.4	9.9	40.3	6.2
No diagnosis on Axis I (V71.09)	8	4	35.9	12.6	41.0	2.7
Delayed diagnosis ASD	2	3	35.0	14.1	41.3	7.2
Delayed diagnosis (non-ASD)	5	3	37.4	6.2	40.3	5.7
Total	64	84				

^a Missing value = 3

Instruments

2.3 Measures

The Modified Checklist for Autism in Toddlers (M-CHAT)

The M-CHAT is the adapted or modified version of the CHAT (*Checklist for Autism in Toddlers*). For use in the United States Robins et al. (2001) adapted the CHAT by eliminating the home visitor observation and adding fourteen parent report items, mostly related to early social-communication and joint attention. The M-CHAT is a parent report checklist containing 23 yes/no items and is designed to identify signs of ASD in children aged 16 – 30 months. It can be completed in 5 – 10 minutes. The measures can be quickly scored, with the results available to providers within minutes (Barton et al., 2012). The original cut-off criteria were maintained in the present study, namely, a child is at risk for ASD (screen positive) when it fails three items or more out of 23 or two out of the following critical items: 2 (interest in other children), 7 (declarative pointing), 9 (brings objects to show), 13 (imitation of action), 14 (responds to name) and 15 (point following). These critical items were identified by a discriminant function analysis of children with and without ASD (Robins et al., 2001). A fail on an item means in this case, answered the question with 'no', aside from item 11, 18, 20 and 22 because these items are reverse-scored.

The Social Responsiveness Scale (SRS)

The original *Social Responsiveness Scale* (Constantino & Gruber, 2005), formerly known as the Social Reciprocity Scale; Constantino, 2005), is a 65-item parent- and teacher-report measure for children between 4 and 18 years old. This screening instrument determined autistic symptoms as quantitative traits, across the entire range of severity in which they occur in nature. It can be used in the general population, but also in educational and clinical settings (Bölte, Poutska, & Constantio, 2008). In the present study, newly developed versions based on the original SRS were used; the parent-report questionnaires of the SRS18-30 for toddlers in the age of 18 to 30 months and the SRS30-48 for young children in the age of 30 to 48 months. The SRS18-30 has 51 items. These items differed from the items of the SRS30-48 and the original SRS4-18; however, they do have the same line and are appropriate for the age group of children in the age of 18-30 months. Two items about the social communication and speech of the toddler are added in which parents can fill out the words and sentences their child use to communicate. For the children between 30 and 48 months, the SRS30-48 for young children was used. This version of the SRS was almost the same as the original SRS. It had the same number of items (65), but the item content of the two versions differed on the basis of developmental appropriateness of the wording for rating the behaviors of children in the respective age groups (Pine et al, 2006). Similarities for the three SRS versions (SRS18-30, SRS30-48 and the original SRS) were that all inquired children's social impairments in a natural social setting based on the past six months. Each item on the scale examined an observed aspect of RSB (Reciprocal Social Behavior) that is rated on a Likert scale format. The answer options are labeled 0 = never true, 1 = sometimes true, 2 = often true, or 3 = almost always true. The questionnaires can be completed in 15 to 20 minutes. The treatment subscales are *social awareness*, *social cognition*, *social communication*, *social motivation*, and *autistic mannerisms*. For screening and diagnosis use of the total score is recommended (Constantino & Gruber, 2005). Higher scores on the SRS indicated greater severity of social impairment. For the original SRS, a total SRS raw score of 75 or higher differentiated best between children with and without a clinical diagnosis of any ASD (Constantino & Gruber, 2005). In the present study a total raw score of 75 is used as a cut-off for determining whether a child has a high probability of having ASD for the SRS30-48. The best cut-off score for the SRS18-30 is examined in this study.

2.4 Procedure

Between December 2010 and May 2012 all parents of young children (18-30 months and 30-48 months) who were referred to five different youth mental health care centers in the South-West of the Netherlands received a questionnaire package prior to intake appointment. The questionnaire package for parents contained several screening measures for ASD. Parents of infants from 18 to 30 months of age received a package including the SRS18-30 and the M-CHAT. Parents of infants from 30 to 48 months received a questionnaire package with just the SRS30-48. For both age groups the package included also a registration form that assesses demographic information and prior health care utilization. After screening, assessment and diagnosis were carried out by an interdisciplinary team at the mental health care center. The composition of the team varied by institution, but always included at least a developmental pediatrician and a clinical child psychologist. The diagnosis was based on the results of diagnostic instruments and developmental tests, combined with medical data and clinical observations from team members. All diagnoses were made according to the Diagnostic and Statistical Manual of Mental Disorders-4th Edition (DSM-IV TR). The DSM-IV TR sorts pervasive developmental disorders, also known as Autism Spectrum Disorders (ASD) into five subcategories: Autistic Disorder (299.00), Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS; 299.80), Asperger's Disorder (299.80), Rett's Disorder (299.80) and Childhood Disintegrative Disorder (299.10) (APA, 2000; for detailed criteria see Appendix IV). In this study all these disorders fell within the category ASD.

2.5 Data analysis

Descriptive analysis of the collected data by region, gender and age was made. In addition, correlations were computed between age in months and gender. For each screening instrument group differences in mean sum scores from children with and without a diagnosis of ASD were established using univariate analysis of variance and t-tests. Level of significance was set as $p < .05$. Cronbach's Alpha was used to assess internal consistency, by conducting a reliability analysis. This analysis computes if the items within the scale measure the same construct. If so, they should all be fairly strongly correlated with each other. Before analyzing the convergent validity by correlating the SRS18-30 scores with the M-CHAT scores, a scatterplot was created. In a scatterplot consistent patterns in the data are easier to see. Subsequently, a parametric test of correlation was conducted. In addition, a Receiver-Operator-Characteristic (ROC) analysis

was run, in order to investigate the best cut-off score for the SRS18-30. A ROC curve was drawn by plotting the sensitivity (Sen) against the 1-specificity (Sp) for every potential cut-off score of the SRS18-30. A binary logistic regression analysis (Chi-square (X^2)) was performed to see if the screen results could significantly predict an ASD diagnosis or not. Screening parameters, such as sensitivity, specificity and predictive values, were assessed for each instrument (SRS18-30, SRS30-48 and M-CHAT) by using the formulas in table 1. This table shows the frequencies or counts in a contingency table. There are two variables (screen result and ASD diagnosis), each with two levels (screen positive or negative for ASD and an ASD diagnosis or non-ASD diagnosis), so the table is called a 2*2 contingency table.

Table 2. 2x2 Contingency table screen result * clinical diagnosis

Screen result	Clinical diagnosis		
	ASD	Non-ASD	
Positive	True Positive (TP)	False Positive (FP)	<i>Positive predictive value</i> = TP / (TP + FP)
Negative	False Negative (FN)	True Negative (TN)	<i>Negative predictive value</i> = TN / (FN + TN)
	<i>Sensitivity</i> = TP / (TP + FN)	<i>Specificity</i> = TN / (FP + TN)	

RESULTS

Descriptives

M-CHAT

The M-CHAT was filled out by parents and caregivers of 64 toddlers. The total M-CHAT score of children with an ASD diagnosis was significantly higher ($M = 7.52$) than children without an ASD diagnosis ($M = 3.02$), $t(62) = -4.581$, $p < .001$, $r = .50$ (see table 3). There was no significant difference found in the mean total M-CHAT score for boys and girls, $t(92) = -.416$, $p = .678$.

SRS18-30

Of the sample of children between 18 and 30 months, 19 children were screened with the SRS18-30. The total SRS18-30 score of children with an ASD diagnosis was higher ($M = 66.4$) than children without an ASD diagnosis ($M = 43.5$). But, there was no significant difference between the total SRS18-30 score among children with an ASD diagnosis and children without, $t(17) = -2.069$, $p = .054$, $r = .44$ (see table 3). There was no significant difference in the mean total score of the SRS18-30 for boys and girls, $t(31) = -1.528$, $p = .137$.

SRS30-48

The SRS30-48 was filled out by parents and caregivers of 84 pre-schoolers. The total SRS30-48 score of children with an ASD diagnosis was higher ($M = 71.7$) than children without an ASD diagnosis ($M = 65.3$). There was no significant difference between the total SRS30-48 score among children with an ASD diagnosis and children without, $t(82) = .499$, $p = .334$, $r = .05$ (see table 3). There was no significant difference in the mean total SRS30-48 score for boys and girls was, $t(106) = -.560$, $p = .577$.

Table 3. Mean scores of the screening instruments for ASD and non-ASD

Screening instruments	ASD			Non ASD		
	M	SD	N	M	SD	N
M-CHAT	7.52*	5.3	21	3.02*	2.6	43
SRS18-30	66.4	33.9	7	43.5	14.4	12
SRS30-48	71.7	30.2	31	65.3	27.7	53

* = significant difference $p < .001$

Hypothesis 1:

The internal consistency of the screening instruments M-CHAT, SRS18-30 and SRS30-48 is good ($\alpha \geq .8$, according to Nunnally and Bernstein; Appendix I).

Cronbach's Alpha was used to determine the internal consistency for the 23-item M-CHAT as well as the six critical items of the current sample. Internal reliability for all items of the M-CHAT was $\alpha = .89$. The internal reliability of the critical items was $\alpha = .83$. Using Cronbach's Alpha, the internal consistency for the SRS18-30 was $\alpha = .95$ and for the SRS30-48 was $\alpha = .92$.

The internal consistence was high for all the measures used.

Hypothesis 2:

The convergent validity between the M-CHAT and SRS18-30 results is high, a strong correlation is expected (Pearson's r value of .7 to 1; Appendix III).

There was a significant positive correlation between SRS18-30 Total score and M-CHAT Total score ($r = .77$, $N = 34$, $p < .05$). It was a moderate correlation: 58.8 % of the variation is shared.

Hypothesis 3:

The cut-off score of the SRS18-30 is below 75.

A cut-off score for the SRS18-30 below 75 is more appropriate. Table 4.1 shows that when using a cut-off score of ≥ 68 the sensitivity of the screen was .57 and the specificity of the screen was 1.00. Using a cut-off score of ≥ 43 gave a sensitivity of .71 and a specificity of .50. The best cut-off score for the SRS18-30 is depends on the requirements you set for the indices sensitivity and specificity.

Table 4.1. Different cut-off values and their sensitivity and specificity values for the SRS18-30

Cut-off value	Sen.	Spec.
39	.71	.42
43	.71	.50
47	.57	.50
51	.57	.67
68	.57	1.0
81	.43	1.0

Hypothesis 4:

The screening parameters for both the M-CHAT and the SRS18-30 and SRS30-48 will be established at least as fair ($\geq .7$, according Cicchetti et al. 1995; Appendix II)

Chi-square

Of the 45 children completed the M-CHAT, 15 screened positive and had an ASD diagnosis (23.4%). The relationship between the M-CHAT screen results and an ASD diagnosis was significant: $X^2(1, N = 64) = 4.939, p < .05$. The association was of weak strength: $\phi = .27$.

Of the 19 children screened with the SRS18-30, with a cut-off score of 68, 4 children screened positive and were diagnosed with an ASD (21.1%). The analysis for the SRS18-30 showed that 2 cells had expected count less than 5, so an exact significance test was selected for Pearson's chi-square. There was a relationship between the SRS18-30 screen results and an ASD diagnosis: $X^2(1, N = 19) = 8.686$, exact $p = .009$. The association was of moderate strength: $\phi = .67$ (see appendix III) and 45.6 % (ϕ^2) of the variance is explained.

Of the 84 children completed the SRS30-48, while using the cut-off score of 75, 14 children screened positive and were diagnosed with ASD (16.7%). The relationship between the SRS30-48 screen results and an ASD diagnosis or not was not significant: $X^2(1, N = 84) = 0.247, p = .619$.

Screening parameters

The screening parameters of the different screening instruments are summarized in Table 5. Applying the criteria of Cicchetti et al. (1995; Appendix II) to the results in Table 5, not a single screening instrument demonstrated acceptable diagnostic accuracy for all four indices (Se, Sp, PPV, NPV). The AUCs of the instruments were poor to fair (with values between 0.56 and 0.76). None of the examined screening instruments seemed to have satisfactory discriminative power in differentiating between ASD and non-ASD in children at a very young age.

Table 5. Outcome measures of the M-CHAT, SRS18-30 and SRS30-48

Screening instrument	N	Se	Sp	PPV	NPV	AUC	95% CI (AUC)
M-CHAT	64	0.71	0.58	0.45	0.80	0.76	0.62 - 0.90
SRS18-30 (cut-off = 43)	19	0.71	0.50	0.45	0.75	0.67	0.35 - 0.98
SRS18-30 (cut-off = 68)	19	0.57	1.00	1.00	0.80	0.67	0.35 - 0.98
SRS18-30 (cut-off = 75)	19	0.43	1.00	1.00	0.75	0.67	0.35 - 0.98
SRS30-48 (cut-off = 75)	84	0.45	0.60	0.40	0.65	0.56	0.43 - 0.69

Se = Sensitivity; Sp = Specificity; PPV = Positive Predictive Value; NPV = Negative Predictive Value; AUC = Area Under the Curve; CI = Confidence Interval; Indices of diagnostic accuracy with fair to good values are in bold (Cicchetti et al. 1995)

DISCUSSION

The main purpose of this study was to validate the ASD screening instruments M-CHAT, SRS18-30 and SRS30-48 in a sample of young children in the Netherlands by assessing the reliability, convergent validity, optimal cut-off score and diagnostic accuracy. The expectations were to find a good internal consistency for the three screening instruments. A high convergent validity between the M-CHAT and SRS18-30. A cut-off score for the SRS18-30 below 75. And, fair screening parameters for the M-CHAT, SRS18-30 and SRS30-48.

Reliability

The study showed that all three instruments (M-CHAT, SRS18-30, and SRS30-48) had a good internal consistency. This meant that the scores on similar items were related. For the M-CHAT, this is consistent with the findings of Robins et al. (2001). For the SRS18-30 and SRS30-48 this is also consistent with studies about the internal consistency of the original SRS (Constantino & Gruber, 2005). Due to the minimal changes in the items for the version for toddlers (SRS18-30) and preschoolers (SRS30-48), it was expected that a good internal consistency for these versions was found as well. The high internal consistency can probably be explained by the fact that the items are all based on different areas of one construct; ASD symptoms in young children.

Convergent validity

There was a strong and significant correlation between the total scores of the M-CHAT and SRS18-30. An explanation for this high convergent validity is that the two screening instruments are almost measuring the same construct; ASD symptoms in young children. The study of Pine et al. (2006) also observed a strong correlation between scores on the SRS for preschoolers (SRS30-48) and the ADI-R and Vineland Adaptive Behavior Composite (VABS).

Optimal Cut-off score of the SRS18-30

Because the newly developed version of the SRS for toddlers (SRS18-30) had fewer items, we investigated the optimal cut-off score. An optimal cut-off score is a score that will

identify almost all children with ASD, with a minimal amount of false positives and a good value for all the screening parameters ($\geq .7$, Appendix II). But, sensitivity will always have a trade-off with specificity. Oosterling et al. (2009) and Barton et al. (2012) suggest in their study that for a screening instrument, the sensitivity is of more value than the specificity. The cut-off score can be modified for different clinical purposes, but in the case of screening for ASDs, there is greater risk in missing children than in pursuing evaluation of a child who does not have the disorder. Screening instruments developed for screening a certain condition in a high-risk population, in this case ASD, can only miss a minimum of cases (desirable is a low value of false negatives). In this study, the sample could be considered 'high risk', as young children referred to the mental health care center often have developmental problems and a relatively large proportion of the sample received a diagnosis of ASD (35,5%). Thus, estimates of specificity may be less critical in the selection of an optimal cut-off score for the SRS18-30. A cut-off of 43 had a fair sensitivity (.71) and a poor specificity (.50). When we raised the cut-off score towards 68, the sensitivity was poor (.57) and the specificity excellent (1.0). When analysing both contingency tables, the cut-off score of 43 had a great amount of false positives, what resulted in the low specificity. The contingency table when using the cut-off score of 68 showed no false positives and just one more false negative than after using the cut-off score of 43. In addition, the two variables screen result and ASD diagnosis or not were not significant associated with each other for the SRS18-30 while using a cut-off score of 43. For the cut-off score of 68 screen result and ASD diagnosis or not were significant associated. That is why in this study there has been chosen for a higher cut-off score of 68.

Diagnostic validity

None of the three screening instruments examined in this study had all fair screening parameters. This meant that there were a lot of false positives and false negatives in comparison with the true positives and true negatives. This marked that children with a negative screen were diagnosed with ASD and children with a positive screen were diagnosed without ASD. This is not desirable for a screening instrument. Explanations for these high amounts of false negatives and false positives are given below. The AUC found in this study is in line with the unsatisfying screening parameters. The area under the curve (AUC) gives an overall indication of the diagnostic accuracy while interpreting the sensitivity and 1-specificity levels. An AUC value closer to 1, indicated high values of sensitivity and specificity; the screening reliably distinguished among young children with and without an ASD diagnosis, whereas values at .50 indicate the predictor is no better than chance (Zhou, Obuchowski, & Obuschowski, 2002). Other

studies found an AUC of .99 for the M-CHAT (Canal-Bedia, 2011) and an AUC of .83 for the original SRS (Bölte et al., 2008). The lowest AUC in this study, was found for the SRS30-48. The AUC value is close to .50 (.56), which indicates that the predictor is not much better than chance. In 56% of the cases, a positive screen on the SRS30-48 predicts an ASD diagnosis and vice versa.

Possible explanations for Screen False Negatives

The low sensitivity of the SRS18-30 and the SRS30-48 is the result of a relatively high amount of false negatives (negative screen result but rather an ASD diagnosis) and low amount of true positives. This is not consistent with the original SRS, Constantino & Gruber (2005) found satisfying values for all screening parameters. A possible explanation of the higher value of false negatives for the SRS30-48 is the use of the same cut-off score as the original SRS. But, by lowering the cut-off score, a disadvantage is the possible growth of false positives and therefore more deterioration of the PPV and specificity. Another explanation for both the SRS18-30 and the SRS30-48 is that caregivers under-report the symptoms of ASD of their child because they are possibly used to it. Previous studies show that ethnic minorities also might under-report the symptoms of their child (Begeer, Bouk, Boussaid, Terwogt, & Koot, 2009). This sample also included ethnic minorities. The high amount of false negatives for the SRS18-30 and SRS30-48 is not desirable in screening children for ASD because children with ASD will be missed by the screen and therefore could not be provided with appropriate help for their disorder.

Possible explanations for Screen False Positives

In this study a large amount of false positives were found for the M-CHAT, consistent to Robins et al. (2001) and Kleinman et al. (2008). Because of that, Robins et al. (2001) introduced a structured telephonic follow-up interview of positive screens. This interview is designed to reduce the false positive rate, which avoids unnecessary referrals. In this study the follow-up interview has not been included, because it is a very time-consuming appliance and the M-CHAT was used in this study as a screening instrument because of her brevity. Another explanation for the large amount of false positives found for the M-CHAT and SRS30-48 is possibly the interpretation of the questions among parents. It is likely that some parents did not understand the questions or interpret them different and therefore over-report ASD symptoms for their child. Another explanation for over-reporting of the symptoms is that funding and intervention opportunities are reliant on a definitive diagnosis. It is possible that parents have sought out information relating to autism and, therefore, inadvertently attribute some of their

child's behavioral difficulties as autistic-like in nature (Aldridge, 2012). For the SRS30-48, over-reporting is a plausible explanation because children without ASD had almost the same high mean total scores as children with ASD. Another possible explanation for the high false positive rate is the relatively young age group. Typical development of these young children is highly variable and some children show early developmental concerns that later resolve (Barton et al., 2012; Pine et al., 2006). These studies even suggest a possible 'sensitive period', in which young children are more influenced by their environment and therefore the ASD symptoms decrease. In addition, recent research from Pandey et al. (2008) suggests that most children who falsely screen positive for ASD at 18 months are not typically developing, and mostly show signs of other development disorders. Maybe younger children under the age of four do have other ASD symptoms.

Strengths and limitations

Some strengths of the current study are that data were systemically collected at different health care centers in the South West of the Netherlands. In addition, a formal diagnosis of ASD was often based on standardized, well-founded autism diagnostic tools, combined with clinical assessment, thus minimizing case misclassification. Clinical awareness of primary care providers and mental health professionals remains extremely important in early detection (Oosterling et al., 2009). The results of this study are tempered by several limitations. For instance, the relatively small sample of children in both age groups. The study group is not large enough to generalize the results to a population. In this study, reliability was investigated through an internal reliability analysis. In order to examine reliability, it is more extensively to calculate also the inter-rater reliability and the test-retest reliability. In addition, as mentioned before in this discussion section, another limitation is that we did not use the follow-up telephone interview for the M-CHAT to review failed items. In this study, we also did not look at child characteristics that can affect the score on a screening instrument, for example IQ and behavioral problems. Another limitation is the use of just the parent-report for the SRS18-30 and the SRS30-48. Teachers are likely to be less invested in whether a child receives a formal diagnosis and they are perhaps better aware of the range of normal, social behaviors in children (Aldridge, 2012). A last unfortunate limitation is the number of missing data, not all items were answered and for some children, the diagnostic process was not yet completed several by the mental health care center.

Conclusion

Overall, the results of this study suggest that the M-CHAT and SRS18-30 are useful screening instruments in the clinical and educational setting in detecting ASD in children of 18 to 30 months, but care must be taken to the high amount of false positives and false negatives. The SRS30-48 needs more research and is not a recommended screening instrument in identifying children with and without ASD following the results of this study. By doing more research with a bigger sample the diagnostic accuracy of all three instruments could be assessed better. Perhaps more research into the items to prevent over- and under-reporting due to unclear wording. Subsequently, all three screening instruments could not be used exclusively for diagnosis. The brief nature of the three screenings instruments makes it user friendly and provides quick assessment of functioning. But, the same advantageous feature of brevity limits its utility and the thoroughness of the information obtained.

Further research

A possibility for further research is a longitudinal large-scale study into the three screening instruments. This could possibly address the study its limitations and also give a good picture of the course of development of the children screened positive for an ASD, in particular children who received a delayed diagnosis. Especially more research into the newly developed versions, SRS18-30 and SRS30-48, is desirable, because of the limited amount of studies that previously have been done. In addition, a challenge for further research is making sure that parents who may have difficulties reading or whose preferred language for communication is not Dutch receive the screening in ways that are appropriate. This could result in a reduction of the non-response and missing items. The comparison of the screening values in a high and low risk sample or possibly a sample with more girls could also be an opportunity for further research.

REFERENCES

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC
- Baird, G., Charman, T., Baron-Cohen, S., Cox, A., Swettenham, J., Wheelwright, S., & Drew, A. (2000). A screening instrument for autism at 18 months of age: a 6-year follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 39, 694 – 702
- Barton, M.L., Dumont-Mathieu, T., & Fein, D. (2012). Screening Young children for autism spectrum disorders in primary practice. *Journal of Autism and Developmental disorders*, 42, 1165-1174.
- Begeer, S., Bouk, S.E., Boussaid, W., Terwogt, M.M., & Koot, H.M. (2009). Underdiagnosis and referral bias of autism in ethnic minorities. *Journal of Autism and Developmental Disorders*, 39, 142 – 148
- Bölte, S., Poutska, F., & Constantino, J.N. (2008). Assessing autistic traits: cross-cultural validation of the social responsiveness scale (SRS). *Autism Research*, 1, 354-363
- Chakrabarti, S. (2009). Early identification of autism. *Indian Pediatrics*, 46, 412-414
- Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., & Pickles, A. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *The British Journal of Psychiatry*, 191, 554 – 559
- Cicchetti, D.V., Volkmar, F., Klin, A., & Showalter, D. (1995). Diagnosing autism using CD-10 criteria: A comparison of neural networks and standard multivariate procedures. *Child Neuropsychology*, 1, 26 – 37
- Constantino, J.N., & Todd, R.D. (2000). Genetic structure of reciprocal social behavior. *American Journal of Psychiatry*, 157, 2043 – 2045
- Constantino, J.N., Przybeck, T., Friesen, D., & Todd, R.D. (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *Journal of Developmental and Behavioral Pediatrics*, 21, 2 – 11
- Constantino, J.N., Davis, S.A., Todd, R.D., Schindler, M.K., Gross, M.M., Brophy, S.L., Metzger, L.M., Shoushtari, C.S., Splinter, R., & Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of autism and developmental disorders*, 33, 427 – 433.
- Constantino, J.N., & Gruber, C.P. (2005). *The Social Responsiveness Scale*. Los Angeles, CA: Western Psychological Services.
- Fombonne, E. (2009). Epidemiology of pervasive developmental disorders. *Pediatric Research*, 65, 591-598.
- Goodwin, C.J. (2005). *Research in Psychology: methods and design*. Hoboken: Wiley Sons.
- Jónsdóttir, S.L., Saemundsen, E., Antonsdóttir, I.S., Sigurdardóttir, S., & Ólason, D. (2011). Children diagnosed with autism spectrum disorder before or after the age of 6 years. *Research in Autism Spectrum Disorders*, 5, 175 – 184.
- Kleinman, J.M., Robins, D.L., Ventola, P.E., Pandey, J., Boorstein, H.C., Esser, E.L., Wilson, L.B., Rosenthal, M.A., Sutera, S., Verbalis, A.D., Barton, M., Hodgson, S., Green, J., Dumont-Mathieu, T., Volkmar, F., Chawarska, K., Klin, A., & Fein, D. (2008). The modified checklist for autism in toddlers: A follow-up study investigating the early detection of autism spectrum disorders. *Journal of autism and developmental disorders*, 38, 827-839

- Motiwala, S.S., Gupta, S., & Lily, M.B. (2006). The cost-effectiveness of expanding intensive behavioural intervention to all autistic children in Ontario. *Health Care Policy*, 1, 136 – 151
- Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Oosterling, I.J., Swinkels, S.H., Gaag, van der, R.J., Visser, J.C., Dietz, C., & Buitelaar, J.K. (2009). Comparative analysis of three screening instruments for autism spectrum disorder in toddlers at high risk. *Journal of autism and developmental disorders*, 39, 897 – 909.
- Pine, E., Luby, J., Abbacchi, A., & Constantino, J. (2006). Quantitative assessment of autistic symptomatology in preschoolers. *Autism*, 10, 344-352
- Robins, D.L., Fein, D., Barton, M.L., & 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.
- Zhou, X. H., Obuchowski, N. A., & Obushcowski, D. M. (2002). *Statistical methods in diagnostic medicine*. Wiley & Sons: New York.
- Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *Health Care and Life Sciences*, 1-9

APPENDICES

APPENDIX I

Nunnally and Bernstein (1994)

A commonly accepted rule for describing internal consistency using Cronbach's alpha:

Cronbach's alpha	Internal consistency
$\alpha \geq .9$	Excellent
$.9 > \alpha \geq .8$	Good
$.8 > \alpha \geq .7$	Acceptable
$.7 > \alpha \geq .6$	Questionable
$.6 > \alpha \geq .5$	Poor

APPENDIX II

Cicchetti, Volkmar, Klin, & Showalter (1995)

Criteria of the various indices of diagnostic accuracy:

Value	Diagnostic accuracy
0.90 – 1.00	Excellent
0.80 – 0.89	Good
0.70 – 0.79	Fair
< 0.70	Poor

APPENDIX III

Cohen, 1988

A commonly accepted rule for describing Pearson's r correlations:

Pearson's r	Strength of correlation
≥ 0.5	Large correlation
0.3 – 0.4	Medium correlation
0.1 – 0.2	Small correlation

APPENDIX IV

ASD classifications according to the Diagnostic and Statistical Manual of Mental Disorders (4th ed. (American Psychiatric Association, 2000)

299.00 Autistic Disorder

A. A total of six (or more) items from (1), (2), and (3), with at least two from (1), and one each from (2) and (3):

(1) qualitative impairment in social interaction, as manifested by at least two of the following:

- a) marked impairment in the use of multiple nonverbal behaviors, such as eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction
- b) failure to develop peer relationships appropriate to developmental level
- c) a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people (e.g., by a lack of showing, bringing, or pointing out objects of interest)
- d) lack of social or emotional reciprocity

(2) qualitative impairments in communication, as manifested by at least one of the following:

- a) delay in, or total lack of, the development of spoken language (not accompanied by an attempt to compensate through alternative modes of communication such as gesture or mime)
- b) in individuals with adequate speech, marked impairment in the ability to initiate or sustain a conversation with others
- c) stereotyped and repetitive use of language or idiosyncratic language
- d) lack of varied, spontaneous make-believe play or social imitative play appropriate to developmental level

(3) restricted, repetitive, and stereotyped patterns of behavior, interests, and activities as manifested by at least one of the following:

- a) encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus
- b) apparently inflexible adherence to specific, nonfunctional routines or rituals
- c) stereotyped and repetitive motor mannerisms (e.g., hand or finger flapping or twisting or complex whole-body movements)
- d) persistent preoccupation with parts of objects

B. Delays or abnormal functioning in at least one of the following areas, with onset prior to age 3 years: (1) social interaction, (2) language as used in social communication, or (3) symbolic or imaginative play.

C. The disturbance is not better accounted for by Rett's disorder or childhood disintegrative disorder.

299.80 Pervasive Developmental Disorder, Not Otherwise Specified (PDD-NOS)

This category should be used when there is a severe and pervasive impairment in the development of reciprocal social interaction or verbal and nonverbal communication skills, or when stereotyped behavior, interests, and activities are present, but the criteria are not met for a specific pervasive developmental disorder, schizophrenia, schizotypal personality disorder, or avoidant personality disorder. For example, this category includes "atypical autism" - presentations that do not meet the criteria for autistic disorder because of late age of onset, atypical symptomatology, or subthreshold symptomatology, or all of these.

299.80 Asperger's Disorder (or Asperger Syndrome)

A. Qualitative impairment in social interaction, as manifested by at least two of the following:

- 1) marked impairment in the use of multiple nonverbal behaviors, such as eye-to-eye gaze, facial expression, body postures, and gestures to regulate social interaction
- 2) failure to develop peer relationships appropriate to developmental level
- 3) a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people (e.g., by a lack of showing, bringing, or pointing out objects of interest to other people)
- 4) lack of social or emotional reciprocity

B. Restricted, repetitive, and stereotyped patterns of behavior, interests, and activities, as manifested by at least one of the following:

- 1) encompassing preoccupation with one or more stereotyped and restricted patterns of interest that is abnormal either in intensity or focus
- 2) apparently inflexible adherence to specific, nonfunctional routines or rituals
- 3) stereotyped and repetitive motor mannerisms (e.g., hand or finger flapping or twisting, or complex whole-body movements)
- 4) persistent preoccupation with parts of objects

C. The disturbance causes clinically significant impairment in social, occupational, or other important areas of functioning.

D. There is no clinically significant general delay in language (e.g., single words used by age 2 years, communicative phrases used by age 3 years).

E. There is no clinically significant delay in cognitive development or in the development of age-appropriate self-help skills, adaptive behavior (other than in social interaction), and curiosity about the environment in childhood.

F. Criteria are not met for another specific pervasive developmental disorder or schizophrenia.

299.80 Rett's Disorder (or Rett Syndrome)

A. All of the following:

- 1) apparently normal prenatal and perinatal development
- 2) apparently normal psychomotor development through the first 5 months after birth
- 3) normal head circumference at birth

B. Onset of all of the following after the period of normal development:

- 1) deceleration of head growth between ages 5 and 48 months
- 2) loss of previously acquired purposeful hand skills between ages 5 and 30 months with the subsequent development of stereotyped hand movements (i.e., hand-wringing or hand washing)
- 3) loss of social engagement early in the course (although often social interaction develops later)
- 4) appearance of poorly coordinated gait or trunk movements
- 5) severely impaired expressive and receptive language development with severe psychomotor retardation

299.10 Childhood Disintegrative Disorder

A. Apparently normal development for at least the first 2 years after birth as manifested by the presence of age-appropriate verbal and nonverbal communication, social relationships, play, and adaptive behavior.

B. Clinically significant loss of previously acquired skills (before age 10 years) in at least two of the following areas:

- 1) expressive or receptive language
- 2) social skills or adaptive behavior
- 3) bowel or bladder control
- 4) play
- 5) motor skills

C. Abnormalities of functioning in at least two of the following areas:

- 1) qualitative impairment in social interaction (e.g., impairment in nonverbal behaviors, failure to develop peer relationships, lack of social or emotional reciprocity)
- 2) qualitative impairments in communication (e.g., delay or lack of spoken language, inability to initiate or sustain a conversation, stereotyped and repetitive use of language, lack of varied make-believe play)
- 3) restricted, repetitive, and stereotyped patterns of behavior, interests, and activities, including motor stereotypies and mannerisms

D. The disturbance is not better accounted for by another specific pervasive developmental disorder or by schizophrenia.