Comparing the Performance Characteristics of Autism Spectrum Disorder Screening Measures in Toddlers

Amanda Palmer

Vanderbilt University

Treatment and Research Institute for Autism Spectrum Disorders

Abstract

The current study represents a preliminary investigation of a new screening measure, the Vanderbilt Scales for Autism Spectrum Disorder (ASD). This measure's psychometric performance in toddlers referred for developmental evaluation (N=159, ages 18-36 months) was compared to the most common screening measure (i.e., the Modified Checklist for Autism in Toddlers [M-CHAT]) and a structured observation of behavior. Results indicated that the Vanderbilt Scales had a lower sensitivity than the M-CHAT, but a higher specificity and positive predictive value. Adding a structured observation of behavior resulted in a higher positive predictive value than when either respective screening measure was used alone. Results highlight the limits of self-report screening measures and the potential value of hybrid screening methodologies utilizing parent report and structured observations.

Comparing the Performance Characteristics of ASD Screening Measures in Toddlers

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder whose current diagnostic criteria (DSM-5: APA, 2013) are established along two domains of impairment: 1) social communication and social interaction; and 2) restricted, repetitive patterns of behavior, interests, or activities. To be diagnosed with ASD, individuals must exhibit all three deficits outlined in the social communication and social interaction domain (i.e., deficits in social-emotional reciprocity; deficits in nonverbal communicative behaviors; and deficits in developing, maintaining, and understanding relationships) and at least two of the four deficits in the restricted, repetitive behaviors domain (i.e., stereotyped or repetitive motor movements; insistence on sameness and routines; highly restricted interests; and hyper- or hypo-reactivity to sensory input). These symptoms must be present in early development, cause clinically significant impairments, and not be better explained by an intellectual disability or other condition (DSM-5: APA, 2013).

According to the CDC (2012), about 1 in 88 children in the United Sates has an ASD with a prevalence rate almost five times higher for males (1 in 54) than for females (1 in 252). ASD has a strong genetic component, with heritability estimated to be between 40 and 90 percent (Bailey, Le Couteur, & Gottesman, 1995; Folstein & Rosen-Sheidley, 2001; Marshall & Scherer, 2012). Being the sibling of another child with ASD greatly increases the risk of being diagnosed with ASD, with recurrence estimated between 6.7 to 18.7 percent. This risk varies by gender and increases twofold when two or more older siblings have been diagnosed with ASD (Gronborg, Schendel, & Parner, 2013; Ozonoff et al., 2011). Despite these strong neurogenetic underpinnings, there is

currently no biological test for ASD and diagnosis depends on behavioral measures and assessments.

Rationale for Early Screening

An increasing body of literature supports significant gains in cognitive and adaptive functioning for young children with ASD who receive intensive, autism-specific early intervention services (Dawson et al., 2010; Warren et al., 2011). Although there has been growing emphasis on early screening and detection of ASD, the average age of diagnosis (both mean and median) in the U.S. is between 4 and 5 years (CDC, 2012). While some children may be able to access low-intensity, non-specific intervention services through early intervention and medical systems while waiting for a diagnostic evaluation, the number of intervention hours is typically considerably less than what is recommended. This is because the healthcare system often requires a diagnosis to obtain access to intensive, autism-specific services (Lord & Richler, 2006; Stahmer & Mandell, 2007). By inhibiting access to early intervention services, delays in diagnosis can have potentially harmful effects on a child's development. Delays in diagnosis can also have negative familial and societal consequences (i.e., parental stress, cost of care, etc.).

The autism society estimates that early diagnosis and effective early intervention services can reduce the lifetime care costs associated with ASD (estimated to be $3.2 million (Ganz, 2007)), by two-thirds (Autism Society of America, n.d.). Additionally, early diagnosis has the potential to increase awareness among parents so they may be better informed about the recurrence risk to later-born children and able to recognize early signs of ASD in siblings (Holt & Christensen, 2013; Jain, Juneja, & Sairam, 2013; Jaspers et al., 2013). These benefits of early diagnosis and intervention, combined with

the prevalence rate of 1 in 88 (CDC, 2012), makes the accurate and early identification of young children with ASD a relevant and important issue.

Current Screening Practice

In an effort to lower the age of diagnosis and take advantage of early intervention services, the American Academy of Pediatrics has recommended universal screening for ASD at 18 and 24 months of age (Johnson & Myers, 2007). Despite the multiple hypothesized benefits of early detection and diagnosis, and practice parameters suggesting early screening, it is estimated that only 42 to 55 percent of pediatricians regularly screen for ASD in toddlers (Carbone et al., 2013). This may be partly due to a lack of data about the most effective protocol for pediatricians to conduct screening to accurately identify ASD.

In community practice, The Modified Checklist for Autism in Toddlers (M-CHAT: Robins, Fein, Barton, & Green, 2001) is one of the most widely used screening measures and has been shown to identify many children with ASD at young ages. However, the M-CHAT often results in substantial over-identification of children with other developmental concerns, particularly when clinicians do not utilize the embedded follow-up interview or other validation procedures (Chlebowski, Robins, Barton, & Fein, 2013; Miller et al., 2011**)**.

The M-CHAT is a 23-item yes/no parent report measure, with accompanying follow-up interview, designed to identify symptoms of ASD in children ages 16-30 months. The measure tests for the presence of deficits in language, arousal modulation, sensory responsiveness, early joint attention skills, motor functions, and social/emotional development (Robins et al., 2001). Follow-up interview questions are only administered

for those M-CHAT items in which the parental responses indicate behavior that demonstrates risk for ASD, and are used to clarify these parental responses.

In a recent study examining the use of the M-CHAT in a community pediatric practice, Miller et al. (2011) demonstrated that the measure can identify children with ASD earlier than parental/provider concern or basic developmental surveillance, but the use of the follow-up interview was required to reduce a substantial false positive rate. Without the use of the follow-up interview, over 100 children from a total sample of 796 were possibly "unnecessarily" referred for evaluations. Although research supports the critical role of this follow-up interview in the use of the M-CHAT, a study examining community pediatric screening processes has suggested that many clinicians do not use the follow-up interview. Swanson et al. (2013) examined the screening practices of 27 community pediatric advisors and found that although 91 percent indicated routine use of ASD-specific screening measures, all cases involved the use of the M-CHAT without the follow-up interview. Another study, Chlebowski et al. (2013), examined the use of the M-CHAT as an autism-specific screener in a large pediatric-based sample. Of the 18,989 children screened with the M-CHAT, 1,737 children (9.1%) screened positive and required the M-CHAT follow-up interview. This large number of positive screens, and required follow-up interviews, makes it difficult for service systems to keep up with need in the pediatric community, offering a possible reason why many pediatric advisors fail to use the follow-up interview in conjunction with the M-CHAT.

As the awareness of the potential positive effects of early diagnosis and intervention is increasing, there is greater emphasis on early screening for all children during pediatric visits. Ideally, children referred for an ASD evaluation would be seen

promptly after referral, but limitations in community resources and clinics often result in families having to wait for long periods of time (6 to 12 months or more) to be seen for a comprehensive diagnostic evaluation (Swanson et al., 2013). These delays cause considerable family stress and may limit access to specific ASD early intervention services.

As pediatricians increasingly use screening tools to make clinical referrals for ASD evaluations, the ability of these measures to distinguish between ASD and non-ASD symptoms becomes even more important. The long wait times are compounded by the use of screening measures, like the M-CHAT, that stress sensitivity over specificity, resulting in high rates of false positive screens. Sensitivity is the ability of a test to accurately identify individuals who manifest the syndrome of interest (i.e., children with autism will screen positive on the measure). Specificity is the ability of a test to accurately identify individuals who do not manifest the syndrome of interest (i.e., children without autism will screen negative on the measure) (Gregory, 2007). Screening measures often have high sensitivity but low specificity, meaning that they identify symptoms not necessarily specific to ASD (Pool & Hourcade, 2011). Consequently, children at risk for other language, developmental, or behavioral disorders often test positive on these screening measures (false positives) and are unnecessarily referred for autism-specific evaluations. Pediatricians, aware of these long wait times for diagnostic evaluations, may be even less likely to routinely screen children for ASD.

Creation of a New Screener at Vanderbilt

Clinicians and researchers at Vanderbilt created a new screening measure, The Vanderbilt Scales for ASD, in an attempt to address some of the challenges of the M-

CHAT (i.e., false positives and the limitations surrounding the follow-up interview in pediatric practice). The scale consists of 23 items pulled from developmental checklists and interviews commonly used at Vanderbilt clinics. The items from these checklists with the highest weights, as determined by a standardized canonical function, were chosen to be included in the Vanderbilt Scales. The standardized canonical function measures the contribution of the item to the discrimination between the two groups (ASD and non-ASD) by taking into account the percentage of children with and without ASD who screened positive on the item (Poulsen & French, n.d.). Each item discriminating between ASD and non-ASD was examined, rephrased for clarity/brevity, and coupled with either a yes/no or compatible Likert scale response option that could potentially be used to develop risk scores. Unlike the M-CHAT, the Vanderbilt Scales utilizes a graded rating system, giving parents an expanded method for reporting their child's abilities. For example, the item on the M-CHAT, "Does your child ever use his/her index finger to point to ask for something? Yes or No?" is compared to the item on the Vanderbilt Scales, "How often does your child point with his/her index finger to request objects that are out of reach? Rarely, Sometimes, Often, Frequently?" The continuum of response options for the items on the Vanderbilt Scales allows parents to more accurately depict how frequently their children perform those behaviors, yielding a score that better reflects their children's abilities.

Additionally, clinicians and researchers at Vanderbilt have been investigating the limits of parent report data with regards to ASD screening (Taylor, Vehorn, Noble, Weitlauf, & Warren, in press; Warren et. al. 2012) and have been exploring if adding a few simple, structured behavioral observations may in fact improve the performance

characteristics of a screening protocol for ASD. Since clinicians place the greatest weight on a child's behavior when determining an ASD diagnosis, it would follow that behavioral observations could be useful in the screening process to identify children at risk for ASD.

Purpose/Hypotheses

The current study sought to 1) create a valid scoring system for the Vanderbilt Scales for ASD, 2) determine if the Vanderbilt Scales for ASD would possess improved performance characteristics over the M-CHAT in accurately identifying children with and without ASD, and 3) investigate the clinical value of a structured observation of behavior in conjunction with the parent report screening measures. It was hypothesized that the Vanderbilt Scales for ASD would possess improved performance characteristics over the M-CHAT, specifically a higher sensitivity, specificity, and positive predictive value (PPV), which is the likelihood that a positive screen reflects the true presence of the syndrome of interest (Glascoe, n.d.). Additionally, it was hypothesized that adding a structured observation of behavior to the screening process would additionally enhance performance characteristics.

Methods

Participants

Participants in the study were caregivers and their children, ages 18-36 months, who were participating in first-time diagnostic appointments or autism research protocols at a university based autism clinic. Participants were referred by a variety of community resources due to clinical concerns or risk status (i.e., siblings). The M-CHAT and Vanderbilt Scales for ASD were obtained on 159 children (n= 21 TD, n= 91 ASD, n= 47

Other DD). Structured observation of behavior (see Autism Diagnostic Observation Schedule below) was available for 155 of these children. In addition, children received a clinical best estimate diagnosis based on an evaluation that involved a clinical interview, cognitive assessment, adaptive behavior assessment, and a research reliable administration of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000). The four missing ADOSs were from diagnostic appointments at the university based autism clinic. Two of these children were diagnosed with global developmental delay and two were diagnosed with language delays. Clinicians involved in these cases felt confident in their non-ASD status and did not feel that the administration of the ADOS would add anything to the clinical picture; therefore it was not administered.

Measures: Screening

The Modified Checklist for Autism in Toddlers (M-CHAT). The M-CHAT (See Appendix A) is a screening tool consisting of a 23-item yes/no parent-report questionnaire. The items are designed to assess the presence of key symptoms of autism, including deficits in: sensory processing (under sensitive to noise), motor functioning (unusual finger movements), social interchange (poor eye contact, lack of responsive smile), early joint attention (bringing objects to parents to share interest, pointing, and following adults' points), and early language and communication (pointing to request). Screening positive on the measure (being identified as at risk for ASD) is defined as screening positive on any three items or two of the six critical items: Does your child take an interest in other children?; Does your child ever use his/her index finger to point to indicate interest in something?; Does your child ever bring objects over to you (parent) to show you something?; Does your child imitate you? (e.g., you make a face-will your

child imitate it?); Does your child respond to his/her name when you call?; If you point at a toy across the room, does your child look at it? (Robins, 2001).

The Vanderbilt Scales for ASD. The Vanderbilt Scales for ASD (See Appendix B) is an early screening measure utilizing graded parental ratings of early core symptoms of ASD. The measure consists of 12 yes/no questions and 11 graded response questions. Item examples: "Is your child using phrases to communicate? Yes/ No;" "How often does your child look at things you point at? Rarely, Sometimes, Often, Frequently?" Items 16-23 were experimental items included in the measure to test a hypothesis regarding response bias for another study. Previous work suggests that the accuracy and validity of self-report and parent-report screening measures can be improved through the use of embedded metrics that account for response characteristics that may influence the scores obtained by the measures, such as over-reporting and under-reporting (Butcher, Graham, & Ben-Porath, 2001; Reynolds & Kamphaus, 2006). These eight items in the Vanderbilt Scales were included to account for parent over-reporting and under-reporting (Taylor et al., in press).

Autism Diagnostic Observation Schedule. The Autism Diagnostic Observation Schedule (ADOS) is a semi-structured, standardized assessment consisting of a series of activities involving social interaction, communication, play, and imaginative use of materials for individuals suspected of having an ASD (Lord et al., 2000). Different modules of the ADOS are available based on the child's age and language level. Module 1 (no words or single words) or module 2 (phrase speech) was administered to 152 participants. There are a total of 28 and 29 coded items in modules 1 and 2 respectively. Each item is coded on a scale from 0-3, with 0 representing typical use of the behavior

and 3 representing the absence of or abnormal use of the behavior. The ADOS-2 (Lord et al., 2012) was published during data collection and three of the participants were administered the toddler module (new to the ADOS-2). This module is for children 12-30 months of age, regardless of language level, and contains 42 items. Participants (n=155) received an ADOS evaluation during their diagnostic appointment. Three of the items, which could identify potential "red-flag" behaviors of ASD, be easily administered in a pediatric setting, and may have value beyond parental assessment, were chosen as a structured observation of behavior. Their administration and coding is explained below.

Pointing: Pointing for the purpose of requesting and/or for shared attention. This behavior can be assessed by holding up two items, such as two different toys, and asking the child which one they want or by observing anytime the child points to share interest during the session (i.e., out the window or to an object out of reach). Coding: 0) child points with their index finger to the object and coordinates the point with a directed gaze to the object and person; 1) child uses an approximation of a point; and/or points in the absence of coordinated gaze; and/or coordinates only pointing that includes touching the object with coordinated gaze or vocalization; 2) child only points while touching an object without coordinated gaze or vocalization; 3) child does not point (Lord et al., 2012).

Response to name: This behavior is assessed by coding the child's response to hearing their name called. Coding: 0) child looks toward the clinician and makes eye contact immediately on either the first or second attempt at calling their name; 1) child looks toward the parent/caregiver and makes eye contact after the first or second attempt, or makes eye contact with the clinician after the clinician's third or fourth attempt; 2)

child does not immediately make eye contact with either clinician or parent/caregiver after his/her name is called in six attempts, but shifts gaze briefly (no eye contact), shifts gaze after a delay, or looks at least once when an interesting or familiar vocalization or verbalization is made (e.g., tongue clicking, "I'm gonna get you"); 3) child does not look toward either the clinician or the parent/caregiver after any purely verbal or vocal attempt to gain attention (Lord et al., 2012).

Response to joint attention: This behavior is assessed by coding the child's response to the clinician's use of gaze or pointing to direct the child's attention to a distant object, such as a toy on a chair. Coding: 0) child uses the clinician's gaze as a cue to turn his/her eyes or face in the direction of the target without the need for pointing; 1) child follows clinician's pointing by looking at or toward the target; 2) child does not follow the clinician's gaze or pointing, but looks at the object when activated; 3) child does not orient to the object even when the object is activated (Lord et al., 2012).

Measures: Diagnostic Evaluation

As part of the evaluation process, children received a clinical best estimate diagnosis from participating psychological providers based on an assessment that involved a clinical interview, cognitive assessment, adaptive behavior assessment, and a research reliable administration of the Autism Diagnostic Observation Schedule.

Composite IQ index. The Mullen Scales of Early Learning (MSEL; Mullen, 1995) is a standardized cognitive assessment for children from birth to 5 years and yields an overall ability index termed the Early Learning Composite.

Adaptive behavior. Adaptive behavior was assessed by the Vineland Adaptive Behavior Scales—Second Edition (VABS-II; Sparrow, Cicchetti, & Balla, 2005), a

parent interview that assesses social, communication, daily living, and motor skills. It provides standard scores for these domains as well as an overall adaptive behavior composite.

Autism symptom severity. The ADOS (Lord et al., 2000) is a semi-structured standardized clinical observation system that measures social, communication, and repetitive behaviors associated with autism. Scoring of the instrument yields an overall 'comparison score" which is operationalized as a severity of autism symptoms calibrated for age and language functioning.

Completion of Measures and Diagnostic Procedures

The M-CHAT and Vanderbilt Scales were either mailed to the participants before the evaluation or they were filled out the same day as the visit during a break in the clinical evaluation. In all cases, the M-CHAT was administered without the follow-up interview. Clinical data collected during the assessment included measures of cognitive ability, adaptive behavior, and autism symptoms. The clinical diagnoses were reviewed and, for the purposes of this study, the children were assigned to one of three categories: typical (n=21 TD), other developmental disability (n=47 Other DD), or ASD (n=93).

Data Analysis

Data analysis included creating a valid scoring system for the Vanderbilt Scales for ASD and coding responses for the M-CHAT and structured observation of behavior to obtain scores on these measures. Sensitivity, specificity, and positive predictive value were calculated for each measure to assess its performance.

Scoring

***Vanderbilt Scales for ASD*.** The items within the Vanderbilt Scales that corresponded to the DSM-5 criteria for ASD were used to obtain an overall score on the measure. Items 4, 5, 7 a-c, 8 a-c, 9, 10, 11 a&b, and 12 correspond to the social communication and social interaction impairment domain and items 15 a-e correspond to the restricted, repetitive patterns of behavior, interests, and activities domain. While items 1, 2, 3, 6, 13, and 14 were omitted from the scoring because they do not correspond to the new DSM-5 criteria for ASD, they do serve to provide useful information for the clinician conducting the screening assessment and may help the clinician develop a better picture of the child's abilities and development. The eight items included in the Vanderbilt Scales as embedded metrics to account for response bias (items 16-23) were not included in the scoring.

All items included in the scoring consisted of four response options: "rarely," "sometimes," "often," and "frequently." For items in which the absence of the behavior in question indicated atypical development, "rarely" and "sometimes" were assigned codes of 1, while "often" and "frequently" were assigned codes of 0. This coding applied to items 4, 7 a-c, 8 a-c, 9, 10, 11 a&b, and 12. Items 5 and 15 a-e were reverse coded because the presence of the behavior in question indicated atypical development. So, for these items, "rarely" and "sometimes" were assigned codes of 0 while "often" and "frequently" were assigned codes of 1.

The codes for all items were summed to yield an overall score for the measure, with a higher overall score associated with the presence of more ASD symptoms. Signal detection methodology, similar to that used in Stone et al. (2004), was then used to

identify a cutoff score. A receiver operating characteristic (ROC) curve was used to plot the sensitivity and specificity for each possible cutoff score and the score that demonstrated the optimal sensitivity and specificity was chosen. Reliability for the Vanderbilt Scales, using this cutoff score, was determined using Cronbach's alpha.

M-CHAT. The M-CHAT was scored according to the published measure (Robins et al., 2001). Screening positive (being identified as at risk) on the M-CHAT was defined as screening positive on two of the six critical items or any three of the 23 items in the measure.

Structured observation of behavior. For each of the three ADOS items, codes of 0 and 1 were scored as "0", and codes of 2 and 3 were scored as "1". Screening positive (being identified as at risk) on the structured observation of behavior was defined as receiving a score of "1" on at least one ADOS item.

Calculating sensitivity, specificity, and positive predictive value (PPV). When investigating the performance of screening measures, sensitivities, specificities, and positive predictive values are often examined. Sensitivity is the ability of the screening measure to accurately identify individuals who manifest ASD. Specificity is the ability of the screening measure to accurately identify individuals who do not manifest ASD (Gregory, 2007). Positive predictive value is the extent to which screening positive on the measure reflects the true presence of ASD (Glascoe, n.d.). In screening development, from a statistical point of view, a measure that has a sensitivity and specificity of 70-80%, and a PPV of 50%, is considered to be a valid screener. Population based studies on ASD screening often do not yield adequate measures of sensitivity and specificity due to the time and cost involved in following all participants from screening to diagnosis, and

often only yield measures of PPV. It is important to note that in large, good quality US studies of the M-CHAT, the PPV ranges from 48-54% (Warren et al., 2014).

For this study, a crosstabs analysis was used to determine the number of participants in each diagnostic category (typical, other developmental disability, and ASD) that screened positive or negative on each screening measure. Sensitivity, the proportion of participants who were diagnosed with ASD and screened positive on the measure, was calculated by dividing the number of participants who screened positive and were given a diagnosis of ASD (true positives) by the total number of participants who received a diagnosis of ASD (true positives and false negatives). Specificity, the proportion of participants who were not diagnosed with ASD and screened negative on the measure, was calculated by dividing the number of participants who screened negative and were not given a diagnosis of ASD (true negatives) by the total number of participants who did not receive an ASD diagnosis (true negatives and false positives). Positive predictive value, the probability that a positive screen reflects the true presence of ASD, was determined by dividing the number of participants who screened positive and received a diagnosis of ASD (true positives) by the total number of participants who screened positive. This analysis was conducted for the Vanderbilt Scales for ASD, M-CHAT, structured observation of behavior, simultaneous use of the structured observation of behavior and the Vanderbilt Scales for ASD (i.e., screening positive on both measures), and the simultaneous use of the structured observation of behavior and the M-CHAT (i.e., screening positive on both measures).

Results

Scoring and Reliability of the Vanderbilt Scales

A receiver operating characteristic (ROC) curve (Figure 1), plotting the sensitivity and specificity for each possible cutoff, showed a cutoff value of nine to have the optimal combination of sensitivity (81.32%) and specificity (66.18%). Screening positive (being identified as at risk) was defined as receiving a score equal to or greater than nine. Screening negative (being identified as not at risk) was defined as receiving a score less than nine. Internal reliability, as determined by Cronbach's alpha, was found to be adequate (α=.803) (Pool & Hourcade, 2011).

Performance Characteristics of Screening Procedures

Sensitivity, specificity, and positive predictive value were calculated for each screening measure, the structured observation of behavior, and for the simultaneous use of the structured observation of behavior with each screening measure. A crosstabs analysis was run to determine the number of participants in each diagnostic category who screened positive or negative on each screening procedure (See Table 1). Sensitivities, specificities, and positive predictive values were then calculated for the sample. The sensitivity of the Vanderbilt Scales, 81.32%, was less than the M-CHAT, 86.81%, but its specificity, 66.18%, was higher than the specificity of the M-CHAT, 52.94%. Additionally, the Vanderbilt Scales resulted in a higher positive predictive value than the M-CHAT: 76.29% and 71.17% respectively. The structured observation of behavior yielded a sensitivity of 79.12%, a specificity of 71.88%, and a positive predictive value of 80.00%. Simultaneous testing requiring positive screens on both the Vanderbilt Scales and the structured observation of behavior resulted in a sensitivity of 68.13%, a

specificity of 89.06%, and a positive predictive value of 89.86%. Simultaneous testing requiring positive screens on both the M-CHAT and the structured observation of behavior resulted in a sensitivity of 74.73%, a specificity of 84.38%, and a positive predictive value of 87.18% (See Table 2). The combination of parent-report measure and structured observation of behavior resulted in the greatest specificity and positive predictive value.

Discussion

Early detection of ASD and subsequent early intervention contributes to significant gains in cognitive and adaptive functioning (Dawson et al., 2010; Warren et al., 2011), reduces the societal costs of ASD (Autism Society of America, n.d.), and contributes to awareness and knowledge in parents (Holt & Christensen, 2013; Jain, Juneja, & Sairam, 2013; Jaspers et al., 2013). Despite this growing awareness and emphasis on early detection, many pediatricians fail to conduct routine screening for ASD (Carbone et al., 2013), and if they do conduct screening, it typically involves the use of the M-CHAT without the accompanying follow-up interview (Swanson et al., 2013). This has been shown to lead to substantial numbers of false-positive screens and over-referrals, creating long wait-lists for autism-specific diagnostic evaluations (Miller et al., 2011). These long wait times cause families considerable stress and delay access to early intervention services, which can be harmful to the child's development. This study sought to investigate the use of a new autism-specific screening measure, the Vanderbilt Scales for ASD. It was hypothesized that this measure would possess improved performance characteristics over the M-CHAT, and have the potential to reduce over-referrals for autism-specific diagnostic evaluations. This study also sought to investigate

the potential clinical value of a hybrid screening protocol utilizing a parent-report measure as well as a structured observation of the child's behavior. Since the diagnosis for ASD is behaviorally based, a screening protocol that takes into account an observation of the child's behavior could more effectively identify children in need of an autism-specific evaluation.

The results of the study indicate that the M-CHAT, being slightly more sensitive than the Vanderbilt Scales, correctly identified a greater percentage of the children with ASD as being at risk. However, the Vanderbilt Scales was more specific, meaning that it correctly identified a greater percentage of the children without ASD as being not at risk, including those with developmental concerns other than ASD. The Vanderbilt Scales also had a higher positive predictive value, meaning that it was more likely that a positive screen on the Vanderbilt Scales reflected the presence of ASD than a positive screen on the M-CHAT. So, without a significant reduction in sensitivity, the Vanderbilt Scales better distinguished between ASD and non-ASD symptoms. Successfully separating children with ASD specific concerns from children with other developmental concerns will help reduce the number of unnecessary referrals for autism-specific evaluations, thereby decreasing wait times and facilitating earlier access to intervention services.

Simultaneous testing requiring positive screens on both the screening measure and structured observation of behavior resulted in a higher specificity and positive predictive value than when either respective screening measure was used alone. The screening protocol requiring positive screens on both the Vanderbilt Scales and structured observation of behavior resulted in the greatest specificity and positive predictive value. However, these increases in specificity and positive predictive value did come at a cost of

decreased sensitivity. By making the screening protocol more stringent, it was better able to identify children without ASD as not at risk, but also failed to identify a greater percentage of children with ASD as being at risk.

It is important to acknowledge that the current study represents only an initial pilot investigation comparing the use of the M-CHAT and Vanderbilt Scales for ASD, and evaluating the possibility and value of adding a structured observation of behavior to a screening protocol for ASD. As such, there are limitations of the current approach in terms of the methods and participants. The use of the structured observation of behavior in the current study involved secondary data analysis. The scores for these three items were taken from each participant's larger ADOS administration. In a community application of this structured observation of behavior, the clinician would only administer the three items, not the entire ADOS. In order to see whether this screening protocol would be feasible, a future study should investigate the use of this structured observation of behavior in conjunction with a parent report screening measure within community pediatric practices. Also, this study involved the use of the M-CHAT without the follow-up interview. So, it is possible that this could account for the lower specificity of the M-CHAT as compared to the Vanderbilt Scales for ASD. However, since many clinicians fail to use the follow-up interview or other validation procedures in conjunction with the M-CHAT (Swanson et al., 2013), the results of the use of the M-CHAT within this research are in line with community practice. Since the Vanderbilt Scales does not require the use of any validation procedures, it may prove to be an easier, more effective screening tool for community clinicians. It is important to note however, that in January 2014, the M-CHAT-R/F was released for free use. This revised measure is a two stage

screening process, consisting of 20 parent report items and a brief structured follow-up interview if the child screens positive. The follow-up interview obtains additional examples to assess risk and is designed to be completed by community clinicians, and may aid in the reduction of false positives (Robins et al., 2014).

Another limitation of the study is that the results pertain to a referred population, and therefore may not be generalizable to a community pediatric sample. The participants in this study were either referred for autism-specific diagnostic evaluations by their community clinicians, or were younger siblings of children with ASD participating in research protocols. The small sample size is also a limiting factor when applying the results to larger, general community populations.

The results of the current study add to the existing body of research on screening measures and protocols and suggest multiple areas for future research within this topic. One possible future direction is to further assess the Vanderbilt Scales for ASD in a larger, community pediatric sample. The current study suggests that the Vanderbilt Scales for ASD is a reliable measure that has the potential to decrease over-referrals while still successfully identifying many children at risk for ASD at young ages. Additionally, it does not require the use of any validation procedures, which makes it a particularly attractive option for use in community pediatric settings since many clinicians report regular use of the M-CHAT, but fail to use the embedded follow-up interview or other validation procedures (Swanson et al., 2013). Another future direction is to investigate a screening protocol that utilizes a parent report measure, clinician follow-up, and structured observation of behavior to identify children in need of an autism-specific evaluation. While the simultaneous use of the screening measure and structured

observation of behavior sacrificed some sensitivity, it greatly increased specificity and positive predictive value. Perhaps by exploring how these three screening elements interact, a more efficient and effective screening protocol could be established.

Even though there is evidence that ASD can be reliably diagnosed in community settings during the second year of life (Warren & Stone, 2011), the average age of diagnosis in the US still remains between four and five years of age (CDC, 2012). While awareness and research on early detection is growing, there is still a lack of consensus on the most effective way to reach children at this young age and screen for ASD. Screening during well child visits to the pediatrician will help reach these children beginning at 18 months, but an effective screening and referral protocol must be established to reduce this gap between the age of screening and diagnosis.

References

Autism Society of America. *Autism through the lifespan: Infants and toddlers.* Retrieved Jan. 23, 2014, from autismsociety.org/living-with-autism/lifespan/infants-toddlers

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5 ed. Washington DC: American Psychiatric Association 2000.

Autism Spectrum Disorders (ASDs) Data and Statistics Center for Disease Control and Prevention. Atlanta: CDC; 2012. Retrieved from http://www.cdc.gov/ncbddd/autism/data.html

Bailey, A., Le Couteur, A., Gottesman, I., et al. (1995). Autism as a strongly genetic disorder: evidence from a British twin study. *Psychological Medicine, 25* (1), 63-77.

Butcher, J. N., Graham, J. R., & Ben-Porath, Y. S. (2001). Minnesota Multiphasic Personality Inventory-2. Minneapolis: University of Minnesota Press.

Carbone, P. S., Murphy, N. A., Norlin, C., Azor, V., Sheng, X. & Young, P. C. (2013). Parent and pediatrician perspectives regarding the primary care of children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 43* (4), 964-972.

Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the Modified Checklist for Autism in low-risk toddlers. *Pediatrics, 131*(4), 1121-1127.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J.,...& Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: the Early Start Denver Model. *Pediatrics, 125*(1), 17-23.

Folstein, E. & Rosen-Sheidley, B. (2001). Genetics of autism: complex aetiology for a heterogeneous disorder. *Nature Review Genetics, 2* (12), 943-955.

Ganz, M. L. (2007). The lifetime distribution of the incremental societal costs of autism. *Archives of Pediatrics and Adolescent Medicine*, *161,* 343-349.

Glascoe FP. Standards for screening test construction. Available at: www.dbpeds.org.

Gregory, R. J. (2007). *Psychological testing: History, principles, and applications* (5th ed.) Boston: Pearson.

Gronborg, T. K., Schendel, D. E., & Parner, E. T. (2013). Recurrence of autism spectrum disorders in full- and half-siblings and trends over time: a population-based cohort study. *JAMA Pediatr*ics, *167* (10), 947-953.

Holt, J. M. & Christensen, K. M. (2013). Utahns' understanding of autism spectrum disorder. *Disability and Health Journal*, *6*(1), 52-62.

Jain, R., Juneja, M., Sairam, S. (2013). Children with developmental disabilities in India: Age of initial concern and referral for rehabilitation services, and reasons for delay in referral. *Journal of Child Neurology*, *28* (4), 455-60.

Jaspers, M., de Winter, A, F., Buitelaar, J.K., Verhulst, F. C., Reijneveld, S. A., & Hartman, C. A. (2013). Early childhood assessments of community pediatric professionals predict autism spectrum and attention deficit hyperactivity problems. *Journal of Abnormal Child Psychology, 41* (1), 71-80.

Johnson, CP & Myers, SM (2007). Identification and evaluation of children with autism spectrum disorders. Pediatrics, 120(5): 1183–1215.

Lord, C. & Richler, J. Early diagnosis of autism spectrum disorders. In: Charman, T. & Stone, W., ed Social & communication development in autism spectrum disorders New York: Guilford; 2006:35-62.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., 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 spectrum of autism. *Journal of Autism and Developmental Disorders,* 30(3): 205–223.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule* (2nd ed.). Los Angeles, CA: Western Psychological Services.

Marshall, C. R. & Scherer, S.W. (2012). Detection and characterization of copy number variation in autism spectrum disorder. *Methods in Molecular Biology, 115-135.*

Miller, J. S., Gqabrielsen, T., Villalobos, M., Alleman, R., Wahmhoff, N., Carbone, P. S., and Segura, B. (2011). The each child study: Systematic screening for autism spectrum disorders in a pediatric setting. *Pediatrics, 127 (5),* 868-872.

Mullen, E. M. (1995) *Mullen Scales of Early Learning.* Circle Pines, MN: American Guidance Service.

Ozonoff, S., Young, G. S., Carter, A., Messinger, D., Yirmiya, N., Zwaigenbaum, L., ... & Stone, W. L. (2011). Recurrence risk for autism spectrum disorders: a Baby Siblings Research Consortium study. *Pediatrics, 128* (3), 488-495.

Pool, J. L., & Hourcade, J. L. (2011). Developmental screening: A review of contemporary practice. *Education and Training in Autism and Developmental Disabilities, 46*(2), 267-275.

Poulsen, J. & French, A. Discriminant Function Analysis. Retrieved from http://userwww.sfsu.edu/efc/classes/biol710/discrim/discrim.pdf

Reynolds, C. R., & Kamphaus, R. W. (2006). Behavior Assessment System for Children, Second Edition. Upper Saddle River: Pearson Education.

Robins, D. L., Casagrande, B. S., Barton, M., Chen, C. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics,133* (1), 37-45.

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.

Sparrow, S. S., Cicchetti, D. V., and Balla, D. A. (2005) *Vineland Adaptive Behavior Scales, Second Edition (Vineland-II)* Minneapolis, MN: Pearson Assessments.

Stahmer, A. C. & Mandell, D. S. (2007). State infant/toddler program policies for eligibility and services provision for young children with autism. *Administration and Policy in Mental Health, 34* (1), 29-37.

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 Disorders, 34 (6)*, 691-701.

Swanson, A., Warren, Z., Stone, W., Vehorn, A., Dohrmann, E., & Humberd, Q. (in press). The diagnosis of autism in community pediatric settings: Does advanced Straining facilitate practice change. *Autism: International Journal of Research and Practice.*

Taylor, C.M., Vehorn, A., Noble, H., Weitlauf, A. & Warren, Z.E. (in press). Can Metrics of Reporting Bias Enhance Early Autism Screening Measures? *Journal of Autism and Developmental Disorders.*

Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-VanderWeele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics, 127* (5), 1303-1311.

Warren, Z. & Stone, W. (2011). Best practices: Early diagnosis and psychological assessment. In: Autism Spectrum Disorders. New York: Oxford university Press; 2011:1271-82.

Warren, Z.E., Vehorn, A., Dohrmann, E., Nicholson, A., Sutcliffe, J., & Veenstra-VanderWeele, J. (2012). Accuracy of phenotyping children with autism based on parent report: What specifically do we gain phenotyping 'rapidly'? *Autism Research*, 5, 31-38.

Warren ZE, Weitlauf AS, Vehorn A, Taylor C, Sathe NA, Krishnaswami S, Fonnesbeck C, McPheeters ML. Screening for Autism Spectrum Disorder in Young Children. Evidence Report/Technology Assessment. 2014. (Prepared by Vanderbilt Evidence-based Practice Center under Contract No. 290201200015i) Rockville, MD: Agency for Healthcare Research and Quality.

Table 1

Summary of the crosstabs analysis for each screening measure showing the number of positive and negative screens for each diagnostic group.

Screening Procedure	Positive Screen	Negative Screen
M-CHAT (n=159)		
ASD	79	12
TD	0	21
DD	32	15
Vanderbilt Scales (n=159)		
ASD	74	17
TD	0	21
DD	23	24
Observation of Behavior (n=155)		
ASD	72	19
TD	3	18
DD	15	28
Observation of Behavior and Vanderbilt Scales (n=155)		
ASD	62	29
TD	0	21
DD	7	36
Observation of Behavior and M-CHAT (n=155)		
ASD	68	23
TD	0	21
DD	10	33

Table 2

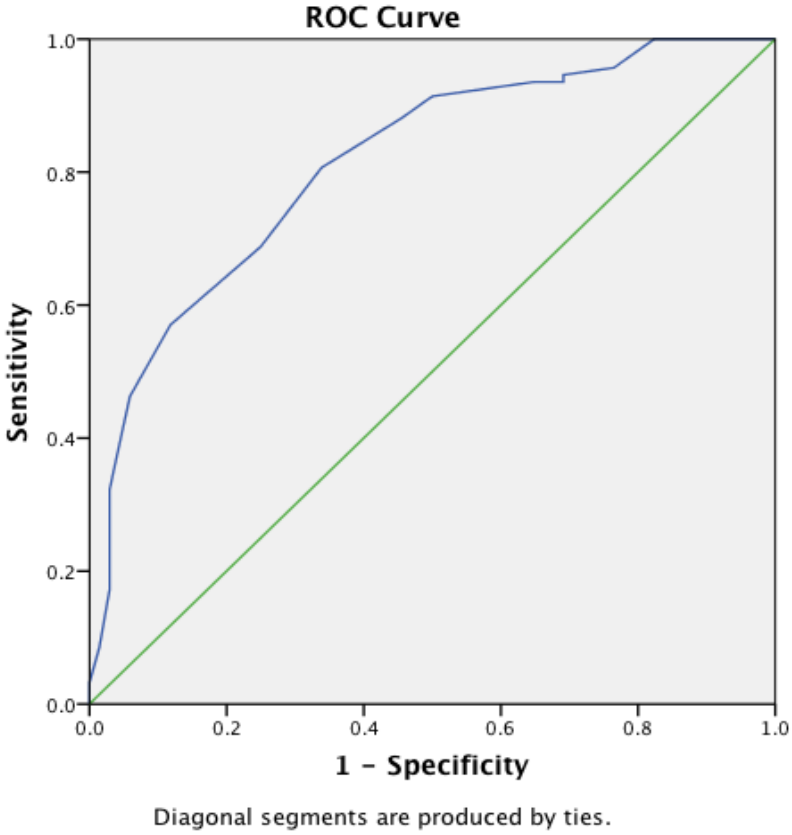
Summary of the performance characteristics (sensitivity, specificity, and positive predictive value) for all screening procedures.

	Sensitivity	Specificity	PPV
Vanderbilt Scales	81.32%	66.18%	76.29%
M-CHAT	86.81%	52.94%	71.17%
Observation of Behavior	79.12%	71.88%	80.00%
Observation of Behavior and Vanderbilt Scales	68.13%	89.06%	89.86%
Observation of Behavior and M-CHAT	74.73%	84.38%	87.18%

Figure 1

Receiver Operating Characteristic Curve: A plot of the sensitivities against the specificities for the different possible cutoff scores. Each point on the curve corresponds to a different sensitivity/specificity pair.

Figure 1


ROC Curve
Sensitivity
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0
1 – Specificity
Diagonal segments are produced by ties.

Appendix A

M-CHAT

Please fill out the following about how your child usually is. Please try to answer every question. If the behavior is rare (e.g., you've seen it once or twice), please answer as if the child does not do it.

1. Does your child enjoy being swung, bounced on your knee, etc.?	Yes	No
2. Does your child take an interest in other children?	Yes	No
3. Does your child like climbing on things, such as up stairs?	Yes	No
4. Does your child enjoy playing peek-a-boo/hide-and-seek?	Yes	No
5. Does your child ever pretend, for example, to talk on the phone or take care of a doll or pretend other things?	Yes	No
6. Does your child ever use his/her index finger to point, to ask for something?	Yes	No
7. Does your child ever use his/her index finger to point, to indicate interest in something?	Yes	No
8. Can your child play properly with small toys (e.g. cars or blocks) without just mouthing, fiddling, or dropping them?	Yes	No
9. Does your child ever bring objects over to you (parent) to show you something?	Yes	No
10. Does your child look you in the eye for more than a second or two?	Yes	No
11. Does your child ever seem oversensitive to noise? (e.g., plugging ears)	Yes	No
12. Does your child smile in response to your face or your smile?	Yes	No
13. Does your child imitate you? (e.g., you make a face-will your child imitate it?)	Yes	No
14. Does your child respond to his/her name when you call?	Yes	No
15. If you point at a toy across the room, does your child look at it?	Yes	No
16. Does your child walk?	Yes	No
17. Does your child look at things you are looking at?	Yes	No
18. Does your child make unusual finger movements near his/her face?	Yes	No
19. Does your child try to attract your attention to his/her own activity?	Yes	No
20. Have you ever wondered if your child is deaf?	Yes	No
21. Does your child understand what people say?	Yes	No
22. Does your child sometimes stare at nothing or wander with no purpose?	Yes	No
23. Does your child look at your face to check your reaction when faced with something unfamiliar?	Yes	No

Appendix B

Vanderbilt Scales for ASD

Please complete the following questions about your child's development.

1) Are you worried about your child's development?
Yes No
Check all areas of you concern:
___ Speech/language ___Motor development
___ Social interactions ___Learning ___Autism

2) Is your child using words to communicate?
Yes No If yes, how many: _____

3) Is your child using phrases to communicate?
Yes No

4) How often will your child use vocalizations or words
to get your attention?
Rarely Sometimes Often Frequently

5) How often will your child use your hands to ask for something (i.e., places your hand on doorknob, refrigerator handle, remote)?
Rarely Sometimes Often Frequently

6) Has your child ever had words or early language skills (i.e., babbling) that seemed to go away?
Yes No If yes, how old was your child when this happened? _____

7) How often does your child point with his/her index finger?
To objects that are interesting
Rarely Sometimes Often Frequently
To request objects that are out of reach
Rarely Sometimes Often Frequently
To objects s/he is touching (e.g., pictures or storybook)
Rarely Sometimes Often Frequently

8) How often does your child use these gestures?
Shakes his head "no"
Rarely Sometimes Often Frequently
Nods his head "yes"
Rarely Sometimes Often Frequently
Reaches to be picked up
Rarely Sometimes Often Frequently

9) How often does your child look at things you point at?
Rarely Sometimes Often Frequently

10) How often does your child bring you objects to show you?
Rarely Sometimes Often Frequently

11) How often does your child look at you to make sure you're watching him/her?
Rarely Sometimes Often Frequently
In a new place or with a new person
Rarely Sometimes Often Frequently

12) How often is your child interested in children his/her age?
Rarely Sometimes Often Frequently

13) How often does your child appear to be more interested
in toys/objects than interacting with you?
Rarely Sometimes Often Frequently

14) How often do you see your child pretending while playing(e.g, feeding/hugging a doll, putting a stuffed animal in a car, making up a story)?
Rarely Sometimes Often Frequently

15) Does your child seem to be *intensely* interested in:
Playing with certain types of toys
Rarely Sometimes Often Frequently
Parts of objects or toys (i.e., wheels, light-switches, buttons)
Rarely Sometimes Often Frequently
The way objects look or feel (i.e., looking closely at objects, feeling textures of toys/objects)
Rarely Sometimes Often Frequently
Organizing toys (i.e., lining-up objects)
Rarely Sometimes Often Frequently
Activating toys (i.e., repeatedly pushing buttons, spinning or wobbling objects)
Rarely Sometimes Often Frequently

16) My child frequently holds his breath when upset?
Yes No

17) My child will laugh and smile when happy.
Yes No

18) My child can roll or throw a ball?
Yes No

19) I feel that nothing I do improves my child's behavior.
Yes No

20) My child frequently makes up complex stories when pretending.
Yes No

21) My child understands everything we say.
Yes No

22) My child will often imitate complex actions with dolls/action figures (i.e., brushes dolls teeth, puts diapers on stuffed animals)?
Yes No

23) Parenting my child is never challenging.
Yes No