Psychometric Properties of Diagnostic Assessment Instruments for Autism Spectrum Disorders in a Community Sample Aged 2 Through 17 Years

Sheryl Lozowski-Sullivan

Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Clinical Psychology Commons, Cognitive Psychology Commons, and the School Psychology Commons

Recommended Citation

https://scholarworks.wmich.edu/dissertations/434

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
PSYCHOMETRIC PROPERTIES OF DIAGNOSTIC ASSESSMENT INSTRUMENTS FOR AUTISM SPECTRUM DISORDERS IN A COMMUNITY SAMPLE AGED 2-17 YEARS

by

Sheryl Lozowski-Sullivan

A Dissertation
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Doctor of Philosophy
Department of Psychology
Advisor: C. Richard Spates, Ph.D.

Western Michigan University
Kalamazoo, Michigan
June 2011
Recent estimates of the incidence and prevalence of autism spectrum disorders (ASDs) indicate substantial increases over the past 20 years (Centers for Disease Control and Prevention, 2009; Fombonne, 2009; Matson & Kozlowski, 2011; Schreibman & Koegel, 2005; Wing & Porter, 2002), in part because of the availability of significantly improved diagnostic assessment instruments (Lord & Corsello, 2005; Schreibman, 2005). In highly structured research settings, some of these diagnostic instruments correlate well with each other; however, few studies have examined the relation between these diagnostic tools in traditional clinical practice (Mazefsky & Oswald, 2006; South et al., 2002). This study examined archival client data for 77 cases from two outpatient clinics serving children with suspected ASD to investigate the psychometric properties of diagnostic assessments when used in clinical practice. The obtained psychometric properties were compared to published psychometric properties from research sites. Obtained reliability and validity measures were much lower for the current study than those in published psychometric studies. This study also examined whether several published findings in the ASD literature were replicated in the community sample. Unlike some previously published studies, no significant differences existed between the
receptive and expressive language scores for those children with a final diagnosis of ASD and those with either a non-ASD diagnosis or no diagnosis. The current study did not find significant differences between the adaptive behavior composite for children with a final diagnosis of ASD and those with either a non-ASD diagnosis or no diagnosis. Similar to published studies (Mazefsky & Oswald, 2006; South et al., 2002), this study found that the mean GARS score and the mean GADS score for those with a final diagnosis of autism/PDDNOS or Asperger’s, respectively, underestimated the probability of being on the autism spectrum when compared to more extensive assessment using direct observation tools. These results highlight the importance of (a) utilizing the multi-method-multi-trait assessment process in community, clinic-based settings to safeguard against overgeneralization from test scores alone, and (b) evaluating psychometric properties of diagnostic instruments in their most commonly used settings in addition to extant research protocols.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>iv</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Diagnostic Criteria</td>
<td>2</td>
</tr>
<tr>
<td>Common Clinical Presentation</td>
<td>5</td>
</tr>
<tr>
<td>Adaptive Profile</td>
<td>8</td>
</tr>
<tr>
<td>Language</td>
<td>9</td>
</tr>
<tr>
<td>Diagnostic Tools for Autism Spectrum Disorders</td>
<td>12</td>
</tr>
<tr>
<td>Overview of Psychometric Properties</td>
<td>13</td>
</tr>
<tr>
<td>Instruments for Assessing Autism Spectrum Disorders</td>
<td>16</td>
</tr>
<tr>
<td>Autism Diagnostic Interview-Revised</td>
<td>16</td>
</tr>
<tr>
<td>Autism Diagnostic Observation Schedule-Generic</td>
<td>19</td>
</tr>
<tr>
<td>Gilliam Autism Rating Scale</td>
<td>21</td>
</tr>
<tr>
<td>Gilliam Asperger's Disorder Scale</td>
<td>22</td>
</tr>
<tr>
<td>Instruments for Assessing Functioning</td>
<td>23</td>
</tr>
<tr>
<td>Wechsler Preschool and Primary Scale of Intelligence</td>
<td>24</td>
</tr>
<tr>
<td>Wechsler Abbreviated Scale of Intelligence</td>
<td>25</td>
</tr>
<tr>
<td>Wechsler Intelligence Scale for Children-IV</td>
<td>26</td>
</tr>
<tr>
<td>Peabody Picture Vocabulary Test-III</td>
<td>26</td>
</tr>
<tr>
<td>Expressive Vocabulary Test</td>
<td>28</td>
</tr>
<tr>
<td>Vineland Adaptive Behavior Scales-II</td>
<td>28</td>
</tr>
<tr>
<td>Psychometric Studies of Autism Spectrum Assessment Instruments</td>
<td>29</td>
</tr>
<tr>
<td>Current Study</td>
<td>32</td>
</tr>
<tr>
<td>METHOD</td>
<td>33</td>
</tr>
</tbody>
</table>
Table of Contents – Continued

Cases Reviewed.............................................................................................................. 33
  Variables Coded......................................................................................................... 33
  Coding Procedure and Inter-coder Agreement....................................................... 34
  Statistical Analyses.................................................................................................. 35
RESULTS.......................................................................................................................... 39
  Data.............................................................................................................................. 39
    Demographic Data..................................................................................................... 39
    Diagnostic Assessment Data.................................................................................... 40
    Statistical Analyses.................................................................................................. 44
DISCUSSION.................................................................................................................... 57
  Similarities with Other Published Psychometric Studies....................................... 63
  Similar Findings for Descriptive Characteristics of ASD....................................... 66
SUMMARY OF CONTRIBUTIONS AND LIMITATIONS............................................... 69
  Future Research and Clinical Recommendations................................................ 70
REFERENCES.................................................................................................................. 74
APPENDICES
  A. Data Coding Sheet................................................................................................. 87
  B. HSIRB Approval Letter.......................................................................................... 91
LIST OF TABLES

1. Sample Tools in a Multi-Method/Multi-Trait ASD Assessment..................... 17
2. Diagnostic Assessments by Year.................................................................. 40
3. Time 1/Time 2-Assessment Diagnoses.......................................................... 42
4. Diagnostic Instrument Descriptive Data......................................................... 43
5. Comparison of ADI-R, ADOS, GARS and GADS Published Internal Consistency Reliability, and Current Study Internal Consistency Reliability............................................................................. 45
6. Comparison of GARS, GADS Published Inter-rater Reliability and Current Study Inter-rater Reliability.............................................................. 46
7. Analysis of Variance for Those with an ASD Final Diagnosis and Those with a non-ASD Final Diagnosis on the ADI-R......................................................... 47
8. Analysis of Variance for Those with an ASD Final Diagnosis and Those with a non-ASD Final Diagnosis on the ADOS......................................................... 47
9. GARS/GADS Quotients by Final Diagnostic Classification................................. 49
10. Analysis of Variance for ASD Characteristics among Those with a Final Diagnosis of Asperger’s................................................................. 49
11. Analysis of Variance for ASD Characteristics among Those with a Final Diagnosis of Autism............................................................................. 49
12. Current Study Estimates of Concurrent Validity (r) for ASD Characteristics........................................................................................................... 50
13. Case Example of 8-Year-Old Female: Diagnostic Assessment Data.............. 53
14. Case Example of 5-Year-Old Male: Diagnostic Assessment Data.................. 55
INTRODUCTION

Several studies suggest that the incidence and prevalence of autism spectrum disorders (ASDs) are rising (Fombonne, 2009; Matson & Kozlowski, 2011; Wing & Porter, 2002) due to several factors, including (a) refinement of diagnostic criteria, (b) increasing awareness of ASD (Fombonne, 2009; Matson & Kozlowski, 2011), (c) diagnostic “switching” from mental retardation to pervasive developmental disorder (PDD; Fombonne, 2009), and (d) development of more accurate assessment tools (Dixon, Garcia, Granpeesheh, & Tarbox, 2009). The incidence rate of ASDs among children ages 4 to 17 years is estimated to be 60 to 70 out of 10,000 children (Centers for Disease Control and Prevention [CDC], 2009), or 1 in 150 children. Utilizing 2002 United States population estimates, Fombonne (2009) estimated 486,000 to 567,000 U.S. residents under age 20 have a PDD. The prevalence of a disorder can increase without a concomitant increase in incidence, primarily due to changes in the method of diagnosis (Fombonne, 2005). In a review of worldwide prevalence studies conducted between 1966 and 2004, Fombonne (2009) found an overall increase in ASD prevalence rates based primarily on changes in the definition of ASD and increased awareness of ASD.

In all epidemiological studies reviewed by Fombonne (2005, 2009), boys were consistently identified as having autism more frequently than girls. The mean male to female ratio was 4.3:1 across studies reviewed. In terms of cognitive functioning, Fombonne reported the median proportion of children with autism also categorized as having severe-to-profound mental retardation was 38.5%, the proportion with mild-to-
moderate impairment was 29.3%, and the proportion without cognitive impairment was 29.6%. Thus, approximately 70% of children with autism have some degree of cognitive impairment (Fombonne, 2005, 2009). Utilizing diverse methods of assessment (e.g., medical examination, parental report), the median rate of at least one medical disorder associated with autism was 5.5 across 16 studies. Three medical disorders frequently associated with autism include (a) epilepsy, (b) Down syndrome, and (c) fragile X.

Although children with ASDs often present with complex medical, cognitive and social challenges, intensive behavioral intervention, when provided at a young age, has been shown to substantially improve outcomes for children with ASDs (Boyd, Odom, Humphreys, & Sam, 2010; Bregman, Zager, & Gerdtz, 2005; Eikeseth, 2009; Howlin, 2005; Odom, Rogers, McDougle, Hume, & McGee, 2007; Schreibman & Ingersoll, 2005; Schreibman & Koegel, 2005; Volkmar, Paul, Klin, & Cohen, 2005). The success of these interventions has created a strong impetus for developing early identification tools (Chawarska, Klin, Paul, & Volkmar, 2007; Lord, 2010; Lord & Corsello, 2005; Matson & LoVullo, 2009; Risi et al., 2006; Ventola et al., 2006), resulting in several new or revised diagnostic instruments over the past two decades to assess children with an ASD (Lord & Corsello, 2005; Risi et al., 2006). Those tools typically focus on the explicit diagnostic criteria for the three ASDs (Hartley, & Sikora, 2009) and common presenting concerns that are useful for screening in broad populations.

**Diagnostic Criteria**

Children diagnosed with ASDs demonstrate impairments in communication and social behavior, as well as restricted, repetitive, or stereotypical behaviors, interests, and activities (American Psychiatric Association [APA], 2000). The ASDs include (a) autistic
disorder, (b) Asperger's disorder (AS), and (c) pervasive developmental disorder not otherwise specified (PDD-NOS), all of which are subsumed under pervasive developmental disorders (PDD) in the DSM-IV-TR (APA, 2000; Volkmar & Klin, 2005). Two other PDDs, Childhood Disintegrative Disorder and Rett Syndrome, are generally not considered part of the autism spectrum (Volkmar & Klin, 2004; Volkmar & Wiesner, 2004), though they have been included as part of the PDD grouping in previous iterations of the DSM system.

According to the DSM-IV-TR (APA, 2000), the diagnostic criteria for autistic disorder requires that a child exhibit a total of six (or more) symptoms from three core categories: (a) qualitative impairment in social interactions (e.g., marked impairment in the use of multiple nonverbal behaviors such as eye-to-eye gaze, facial expression, body postures, failure to develop peer relationships, lack of social or emotional reciprocity); (b) qualitative impairments in communication (e.g., delay or lack of spoken language, marked impairment in ability to initiate or sustain conversation, stereotyped and repetitive language or idiosyncratic language, lack of make-believe play or social imitative play); and (c) restricted, repetitive, and stereotyped patterns of behavior, interests, and activities (e.g., preoccupation with one or more stereotyped and restricted patterns of interest, inflexible adherence to specific, nonfunctional routine or ritual, stereotyped and repetitive motor mannerisms, preoccupation with parts of objects). Of those six symptoms, at least two must be from the social domain, while one is required in each of the other two domains. Autism, sometimes referred to as "classical autism," is the most clearly defined of the ASDs (Matson & Boisjoli, 2007; Tidmarsh & Volkmar, 2003).

The criteria for a diagnosis of AS are similar to those for autism (i.e., at least two
impairments in the social interaction domain and one in the domain of restricted, repetitive, and stereotyped patterns of behavior), except that no overall delay in language skills, cognitive development, self-help skills, or adaptive behavior is evident (APA, 2000). Though overall delays are not evident, individuals diagnosed with AS often have idiosyncratic and peculiar language characteristics and uneven cognitive profiles, with a large portion displaying symptoms consistent with a non-verbal learning disorder. These symptoms include strengths in verbally mediated skills (e.g., vocabulary, knowledge, verbal memory and output) and deficits in nonverbal skills (e.g., visual spatial problem-solving, visual motor coordination (Klin, Pauls, Schultz, & Volkmar, 2005). Two additional distinctions differentiate autistic disorder from AS. First, AS is often apparent later than autism, and second, children with AS often desire social contact, whereas children with autistic disorder are frequently indifferent to social contact (Ozonoff, Goodlin-Jones, & Solomo, 2005; Sciutto & Cantwell, 2005; Tidmarsh & Volkmar, 2003). Children with AS often want to engage socially with others but have difficulty understanding social norms (Ozonoff et al., 2005; Matson & Boisjoli, 2007; Myles & Simpson, 2002). The difference in social awareness and desire to engage socially between children with AS and autism creates a debate among experts. Some suggest that AS is a milder form of autism (Volkmar & Klin, 2005), while others suggest that AS is fundamentally different than autism because individuals with AS desire social contact but are strikingly socially inept (Myles & Simpson, 2002).

The diagnostic category PDD-NOS is a sub-threshold category for individuals who display some characteristics but do not meet all the criteria for autistic disorder or AS (Ozonoff et al., 2005; Matson & Boisjoli, 2007; Volkmar, Lord, Bailey, Schultz, &
Klin, 2004). PDD-NOS is diagnosed when an individual displays a severe impairment in
the development of reciprocal social interaction associated with either verbal or
nonverbal communication skills or with the presence of stereotyped behavior, interests,
and activities (APA, 2000). An individual diagnosed with PDD-NOS may exhibit
behavior very similar to individuals diagnosed with autism or AS but not to the extent
that he or she meets all of the criteria for one of those disorders. For this reason, PDD-
NOS is often called “atypical autism” and is sometimes considered a milder form of
This diagnostic category of PPD is used often clinically, but research on it is in its
infancy compared to research on autism (Matson & Boisjoli, 2007; Volkmar & Klin, 2005).

**Common Clinical Presentation**

In addition to the characteristics described above, several other common features
have been observed in individuals with ASDs that are not specifically detailed as
diagnostic criteria. For example, children with ASDs often have lower-than-expected
adaptive skills (i.e., motor skills, daily living, communication, and socialization) based on
their intellectual functioning (Matson, Mayville, et al., 2003). Assessments of children
with ASDs show a wide variability of intellectual functioning skills with one or two
superior skills (e.g., learning disabilities interspersed with exceptional splinter skills) and
frequent occurrence of problematic and disruptive behaviors (Myles & Simpson, 2002;
Volkmar & Klin, 2005). Affective disorders (e.g., affective ability, inappropriate
affective responses, depression, and anxiety) are also relatively common in individuals
with ASDs (Klin, McPartland, & Volkmar, 2005; Ozonoff et al., 2005; Volkmar & Klin,
2005). In addition, sensory and motor features are commonly associated with ASDs.
Sensory features involve a variety of behavioral manifestations of sensory processing, including (a) hyposensitivity, (b) hypersensitivity, (c) over-responsivity and under-responsivity, and (d) preoccupation with sensory stimuli. Individuals diagnosed with an ASD often have deficits in the ability to conceptualize, plan, and execute voluntary, goal-directed actions (Baranek, Parham, & Bodfish, 2005). Motor coordination of individuals with ASDs is often characterized by (a) odd posture or gait, (b) poor body awareness, and (c) clumsiness (Howlin, 2005; Klin, Paul, et al., 2005).

The clinical presentation of ASDs can vary greatly across the developmental span and across individuals (Lord & Corsello, 2005; Volkmar & Klin, 2005). The most common presenting issues for children aged 2-12 are detailed here, but interested readers are directed to LeBlanc, Riley, and Goldsmith (2008) for a review of behavioral characteristics in adolescence and adulthood. Childhood is typically divided into two broad categories: early childhood (0 to 6 years) and later childhood (6 to 12 years). The earlier period serves as the primary window of identification for children with autism and PDD-NOS with language delays, and the later period is more commonly associated with identification of AS.

A substantial body of research documents the onset of ASD characteristics within the first three years of life (Carter, Davis, Klin, & Volkmar, 2005; Chawarska, Klin, et al., 2009; Chawarska & Volkmar, 2005; Coonrod & Stone, 2005; Eaves & Ho, 2004). By age 2, developmental abnormalities are recognized by 80 to 90% of parents. Parents often present children to medical professionals with concerns of slower-than-expected speech development and concerning behaviors such as hand and finger mannerisms and other repetitive behaviors. Loss of language skills and decreased interest in activities and social
interactions may also prompt parents to seek advice, as approximately 25 to 45% of children with autism regress after initially developing language and social interaction skills (Chawarska & Volkmar, 2005; Tager-Flusberg, Paul & Lord, 2005). By the fourth birthday, children with autism typically display a number of impairments in social interaction (e.g., abnormal eye contact, limited interest in other children, limited social smile, limited range of facial expressions) and communication (e.g., poor response to name, failure to respond to gestures, use of others’ body as a tool), as well as stereotypical behaviors and repetitive patterns (e.g., inappropriate use of objects, repetitive interests or play, unusual sensory behaviors).

The period of middle childhood is typically marked by progress for those who have been identified early or by initial identification for those with more subtle clinical presentation. For those who were identified in early childhood, middle childhood (ages 6 to 12) often results in progress in (a) social and adaptive skills, (b) play skills (i.e., cooperative and pretend), and (c) language and communication skills for those individuals who do not have co-morbid mental retardation. However, overall progress for the child with ASD is typically slower than what is expected for normally developing age-mates. For those school-age children with mental retardation and ASD, social and adaptive skills may regress or plateau. For children with AS, initial identification may occur in middle childhood, as children with AS often come to the attention of professionals after entry into school when their difficulties in understanding social cues (e.g., eye contact, social cues, personal space) and developing peer relations become apparent. Although language and intellectual delays are not evident, other behavioral and social issues are typically dominant, with school adjustment, academic achievement, and
emotional development (i.e., recognizing, understanding, and expressing affect) proving problematic. Many school age children with ASDs also have hyperactive and inattentive behaviors that create difficulties in the transition from home or preschool to primary school (Ozonoff et al., 2005; Loveland & Tunali-Kotoski, 2005).

**Adaptive Profile**

As mentioned previously, approximately 70% of children on the spectrum also have mild to moderate cognitive impairment (Fombonne, 2005), which often affects adaptive behaviors (e.g., self-care, safety). Adaptive behavior development is directly related to cognitive functioning and severity of ASD symptoms (Loveland & Tunali-Kotoski, 2005). For children with mild to moderate cognitive impairments and ASD, a profile of socialization deficits is typically observed. For very young children, social adaptive skills such as interest in others, simple interactive games, and demonstration of affection are affected (Chawarska, Klin, et al., 2009; Klin, Saulnier, Tsatsanis, & Volkmar, 2005).

School-age children on the spectrum typically have difficulties with (a) sharing, (b) cooperative play, (c) approaches of other children, and (d) pretend play, for example (Loveland & Tunali-Kotoski, 2005). Moreover, children with autism typically have lower scores than their age-mates and IQ-matched controls in socialization and communication from infancy into school age (Chawarska, Klin, et al., 2009; Klin, Volkmar, & Sparrow, 1992; Lord & Schopler, 1989; Stone, Ousley, Hepburn, Hogan, & Brown, 1999) and greater variability in adaptive skills (Matson, Mayville, et al., 2003; Klin et al., 1992) than children not on the autism spectrum. Children with AS tend to have better developed adaptive skills than those with autism, a difference that may be related to cognitive
functioning or more highly developed verbal skills in many children diagnosed with AS (Loveland & Tunali-Kotoski, 2005).

**Language**

Children with ASDs show patterns of delay in language differing from normally developing children and children with specific language disorders (Tager-Flusberg, Paul, & Lord, 2005). Children on the spectrum typically speak late and develop speech at a slower rate than their typically developing age peers (Chawarska, Klin, et al., 2009; Tager-Flusberg, Paul, & Lord, 2005). Approximately 25% of children with autism (not AS, PDD-NOS or other developmental delays) experience a loss of language skills after acquisition of words between 12 and 18 months (Gerenser, 2009; Tager-Flusberg, Paul, & Lord, 2005). Children diagnosed with AS do not exhibit general delays in language or cognition but show deficits in social interactions (Loveland & Tunali-Kotoski, 2005).

Children with ASDs show more severe receptive language difficulties than typically developing children and children with specific language impairments (Loucas, et al., 2008; Paul, 2005). Children on the spectrum as young as 12 months are less responsive to their name, to their mother’s or any person’s voice than typically developing children (Lord & Richler, 2006; Paul, 2008). By age 2, most children on the spectrum have sufficiently delayed language to alert parents there may be a problem with the child (Lord & Richler, 2006; Paul, 2005; Tager-Flusberg, Paul, & Lord, 2005; Loveland & Tunali-Kotoski, 2005).

Approximately 20% of children with autism never acquire functional language; however, increasing awareness of ASDs and early intervention appears to be reducing the proportion of children on the spectrum without functional language (Tager-Flusberg,
Paul, & Lord, 2005). Specifically, children on the autism spectrum often demonstrate normal articulation later than typically developing peers. The development of categorization skills (e.g., food, animals, clothes) in children on the spectrum is similar to typically developing children but often lacks classes for mental states (e.g., think, remember) and social emotional terms (e.g., angry, sad) (Gerenser, 2008; Tager-Flusberg, Paul, & Lord, 2005).

While many normally developing children, children with below average IQs, and children on the spectrum use words that approximate or associate with the actual word such as, “bow-wow” for dog and “puter” for computer, only children on the autism spectrum use neologisms or odd phrases for which the related word is not easily identifiable (e.g., “hot rain” for steam) (Tager-Flusberg, Paul, & Lord, 2005). Another classic feature of autism is echolalia—immediate or delayed exact repetition of spoken language, usually a phrase such as, “It’s time for bed,” or “To infinity and beyond.” Echolalia is not exclusive to autism spectrum disorders, occurring in normally developing children, children without eyesight, individuals with dementia, and children with other language impairments (Gerenser, 2008; Tager-Flusberg, Paul, & Lord, 2005).

Additional language abnormalities in children with ASDs include pronoun reversals (e.g., “he” for the self-referent “I”), vocal quality, intonation (e.g., monotony), and stress patterns. Comprehension of words in children with ASDs is delayed relative to production skills with particularly substantial impaired understanding of language in social contexts (i.e., pragmatics) due to the underlying social communication deficits (Paul, 2008). Moreover, sustained delay in verbal comprehension differentiates children on the spectrum from those children with a language disorder (Loucas et al., 2008).
Researchers continue to more precisely define the features of language impairments in children with ASDs (e.g., Louvas, et al., 2008). Children with autism and PDD-NOS typically have delayed receptive and expressive vocabulary compared to their non-ASD age peers (Louvas et al.). Hundry, et al., (2010) found greater impairment in receptive than in expressive language on the Preschool Language Scale (PLS) and the Vineland Adaptive Behavior Scale- 2nd Edition (VABS-II) with substantial individual variation. However, discrepancies in expressive versus receptive language may be a function of the specific feature(s) of language (e.g., lexical, syntax, pragmatics) measured by the instrument (Hundry et al.). Kjelgaard & Tager-Flusberg (2001) investigated language functioning in 89 children with autistic disorder. Comparing the means of the PPVT-III and EVT, which primarily measure lexical language (i.e., word meanings or vocabulary), they found no significant difference (t-test) between receptive and expressive language in children with autism.

Of the children with autism who had a discrepancy of at least one standard deviation between PPVT-III and EVT scores, 13 (16%) had higher PPVT-III scores and three (4%) had higher EVT scores. Like Hundry et al., Kjelgaard & Tager-Flusberg found significant variability in language skills among children with autism, ranging from children whose vocabularies, grammatical knowledge, and articulation skills were in the normal range to non-verbal children. Conversely, children with AS often have language skills that exceed average (Klin, Pauls, Schultz, & Volkmar, 2005). Increasingly, comprehension is the focus of language impairment and a target for early intervention in children on the autism spectrum (Hundry, et al., 2010; Kjelgaard & Tager-Flusberg). Klin, Saulnier, et al. (2005) recommended that a profile of assets and deficits for the
individual child be developed based on multiple instruments, with results of assessment being interpreted based on multiple lines of converging evidence in order to guide individualized treatment (Schreibman, 2005).

**Diagnostic Tools for Autism Spectrum Disorders**

Significant progress has been made in the development and refinement of ASD diagnostic tools (Lord, 2010; Lord & Corsello, 2005), such that diagnoses of ASDs in early childhood can be reliable, valid, and stable over time (Coonrod & Stone, 2005; Lord & Corsello, 2005). A variety of diagnostic tools have been developed in the past 30 years for the two primary purposes of screening and diagnosis. Screening large populations of children for possible autism is recommended in coordination with developmental screening (Lord & Corsello, 2005). Two such instruments used for screening include the Modified-Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, Barton, & Green, 2001) and the Australian Scale for Asperger’s disorder (Attwood, 1997). Broad-based screening is a critical pathway for identifying and diagnosing children with ASDs (Coonrod & Stone, 2005).

Diagnosis is a prerequisite for access to clinical intervention services and specialized educational services for children with autism as well as other childhood mental health issues. Accurate diagnoses (Chawarska et al., 2007; Lord & Corsello, 2005; Risi et al., 2006; Ventola et al., 2006) and behavioral intervention in early childhood (Bregman et al., 2005) have been shown to substantially improve outcomes for children with ASDs (Howlin, 2005; Schreibman & Koegel, 2005; Volkmar et al., 2005). For this reason, development of accurate and reliable assessment tools for children with ASDs for use in both research and clinical settings is imperative. Another purpose of diagnostic
assessment is to delineate individual functioning in various domains (e.g., intellectual and adaptive functioning, expressive and receptive verbal abilities, social competence, self-care, activities of daily living) in order to inform and monitor treatment interventions.

Instruments used to diagnose ASDs have benefited from the interplay between research refining the diagnostic categories of ASD and clinical utility (Lord, 2010; Lord & Corsello, 2005); however, researchers continue to struggle with further refinement of the subcategories comprising ASDs (e.g., Hundry et al., 2010). General issues affecting ASD diagnostics include (a) difficulties differentiating developmental changes from diagnostic variables, (b) studies with relatively small sample sizes, (c) difficulties in controlling for language delay when examining other variables, (d) absence of biological markers, and (e) the degree of overlap between autism and non-autism ASDs (i.e., AS and PDD-NOS; Matson, & Boisjoli, 2007; Lord, 2010; Lord & Corsello, 2005). Current assessment instruments used to diagnose children for ASDs focus on the core symptom areas of impairment using a variety of standardized, observational, and criterion-referenced tools measuring communication, behavior, and psychological functioning (Lord, 2010; Volkmar et al., 2005).

**Overview of Psychometric Properties**

The *psychometric properties of a scale* refer to the statistical measures that indicate whether a measure adequately assesses the construct it purports to measure. The broad delineations of psychometric properties refer to reliability and validity (Kazdin, 2003), which are estimated using a variety of techniques described below. A measure may be reliable but not valid, but it cannot be valid without being reliable. That is, reliability is a necessary but not sufficient condition for validity (Anastasi & Urbina, 1997).
Reliability refers to the consistency of the measure or agreement between two sets of independently derived sets of scores (Anastasi & Urbina, 1997). Broadly defined, this involves scores obtained by the same person when the person retakes the same or a parallel, equivalent version of a test at a different time or under different testing conditions (Anastasi & Urbina, 2003; Kazdin, 2003). Reliability is expressed as a correlation coefficient (r) and is a combination of error variance and true variation. Reliability coefficients are interpreted in terms of the percentage of score variance attributable to different sources. For example, a reliability coefficient of .90 indicates that 90% of the variance in test scores is due to true variance and 10% of the variance in test scores is due to error variance. The widely-accepted cut-off in social sciences is that the correlation coefficient should be .70 or higher. Correlations lower than .70 are considered poor, .70 to .79 are fair, .80 to .89 are good, and > .90 are excellent (Cicchetti, 1994).

Specifically within psychometrics, Bracken (1987) suggested that technically adequate measures should be internally consistent and stable over time as evidenced by (a) internal consistency of .90 or greater for the instrument’s total score, (b) internal consistency of .80 or greater on median subtests, and (c) total test temporal stability of .90 or greater. In addition, effect size is a descriptive statistic that measures the strength of a relationship between two variables. Cohen (1992) provides effect size guidelines for the social sciences, including the one-way analysis of variance applicable to inferential statistics analyses in the current study: (a) small = .10, (b) medium = .25, and (c) large = .40.

The most commonly used techniques for estimating reliability include (a) test-retest reliability (i.e., consistency over time), (b) alternative-form reliability (i.e., consistency across versions), (c) inter-rater or inter-scorer reliability (i.e., consistency
across respondents; Anastasi & Urbina, 1997; Kazdin, 2003), and (d) internal consistency. *Internal consistency* refers to the degree of homogeneity of the items in an instrument (Kazdin, 2003) and is estimated using (a) split-half reliability (i.e., items of test divided into two halves and correlated), (b) Kuder-Richardson, or (c) coefficient alpha. Kuder-Richardson is typically used with dichotomous items (i.e., yes/no, right/wrong), whereas when scaled responses (e.g., sometimes, often, never) are employed, coefficient alpha is typically used (Anastasi & Urbina, 1997; Kazdin, 2003). Internal consistency and instruments used to estimate internal consistency as they apply to particular ASD diagnostic instruments are described in greater detail in the paragraphs that follow.

*Validity* refers to the degree to which a test measures what it purports to measure and how well it does so. Unlike reliability, which generally is relatively high or low on a fixed scale, validity in psychological assessment is based on the specific use for which it is being considered (Anastasi & Urbina, 1997). That is, the question is not simply one of how well different administrations correlate but whether the instrument is valid for use with a particular sample (e.g., developmentally-delayed children ages 4 to 12).

Researchers typically examine the type of validity that is suited to their particular research purposes by comparing performance of a specific measure to another independently determined measure of the domain of interest (i.e., a “gold standard” measure). The types of validity can be classified as (a) construct validity, (b) content validity, (c) criterion validity (i.e., concurrent validity and predictive validity), (d) face validity, (e) convergent validity, and (f) discriminant validity (Kazdin, 2003). These types of validity are described further as they apply to the ASD diagnostic instruments detailed below.
Instruments for Assessing Autism Spectrum Disorders

The current state-of-the-art ASD diagnostic tools include observational, interview, and informant checklists (Klin et al., 2005; Lord & Corsello, 2005; Lord et al., 2000; Ozonoff et al., 2005) employed in accordance with a multi-method assessment process. That is, the use of multiple assessment methods in differential diagnosis safeguards against overgeneralization from test scores alone (Anastasi & Urbina, 1997).

Psychological assessment involves the process of integrating information from multiple sources, including (a) psychological tests; (b) interviews including history of development, behaviors, symptoms, and description of current symptoms and problems; and (c) standardized observation. Multiple measures of a trait are needed to ensure that the results are not restricted to the trait as assessed by a particular measure and method (Campbell & Fiske, 1959; Kazdin, 2003).

Table 1 illustrates tools used in the psychological assessment of children considered for an ASD using multiple methods (i.e., interview, observation, checklists) to measure multiple traits (i.e., characteristics and functioning). Each measure and its selected psychometric properties are described below. Not all psychometric properties (e.g., normative data on special groups) are included in the discussion; however, those psychometric properties pertinent to the proposed current study are included.

**Autism Diagnostic Interview-Revised**

The Autism Diagnostic Interview-Revised (ADI-R; Le Couteur, Lord, & Rutter, 2003) is a 93-item caregiver interview assessing (a) communication, (b) reciprocal social interaction, (c) play, and (d) developmental history for a possible diagnosis of an ASD.
Table 1

*Sample Tools in a Multi-Method/Multi-Trait ASD Assessment*

<table>
<thead>
<tr>
<th>Method</th>
<th>Trait</th>
<th>Functioning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview</td>
<td>Autism Diagnostic Interview-Revised (ADI-R)</td>
<td>Vineland Survey Interview</td>
</tr>
<tr>
<td>Observation</td>
<td>Autism Diagnostic Observation Schedule-Generic (ADOS-G)</td>
<td>Wechsler Peabody Picture Vocabulary Test (PPVT-III)/Expressive Vocabulary Test (EVT)</td>
</tr>
<tr>
<td>Checklists</td>
<td>Gilliam Autism Rating Scale (GARS)</td>
<td>Vineland Survey Rating Form</td>
</tr>
<tr>
<td></td>
<td>Gilliam Asperger's Disorder Scale GADS</td>
<td></td>
</tr>
</tbody>
</table>

The ADI was originally developed as a research diagnostic tool (Lord & Corsello, 2005), modified to include a broader and lower age range of children, and shortened to 93 items (Lord, Rutter, & Le Couteur, 1994). The revised version of the ADI, the ADI-R, is considered a necessary component (Tidmarsh & Volkmar, 2003) or the “gold standard” interview method of an ASD assessment (Lord & Corsello, 2005).

Inter-rater reliability was calculated using percentage exact agreement and weighted kappas for all individual rater pairs. Inter-rater reliability demonstrates that different raters scored the same subject on the same instrument and achieved similar results. This means that the reliability property of inter-rater or inter-scorer reliability (i.e., consistency across respondents) was assessed by six pairs of raters viewing and coding interviews of the reliability participants. Participants were 10 autistic and 10 mentally handicapped or language-impaired participants involved in the psychometric
analysis of the ADI-R. Weighted kappa exceeded .63 on the social algorithm items. Multi-rater weighted kappas exceeded .69 for all 13 communication items. Weighted kappa levels exceeded .69 for all seven items in restricted, repetitive behaviors. Interrater agreement for rater pairs exceeded 92% for all pairs (Lord et al., 1994). Thus, based on Bracken’s (1987) criteria, this represents adequate temporal stability.

Internal consistency of the ADI-R algorithms was assessed using Cronbach’s alpha. This is a reliability estimate of the degree of homogeneity of the items in an instrument (Kazdin, 2003). Internal consistency estimates were reported to be .95 in the social domain, .84 in the communication domain, and .69 in the restricted, repetitive behaviors domain (Lord et al., 1994). Thus, the internal consistency within the social and communication domains is adequate; however, according to Bracken’s (1987) criteria for psychometric properties, the internal consistency of the restricted, repetitive behaviors domain falls short of adequate.

Validity analyses were conducted with 25 autistic and 25 non-autistic subjects (i.e., the subjects in reliability study plus an additional 15 autistic and 15 non-spectrum subjects). The appropriateness of the ADI-R algorithm was tested by comparing individual items and area summary scores across diagnostic groups and by comparing the number of children who met autism criteria. This comparison represents discriminant validity, indicating how well the instrument discriminated between those children with autism and non-spectrum children having similar characteristics (e.g., language delay). Within the domains of social, communication, and restricted and repetitive behaviors, the algorithms for each group (i.e., autistic and non-autistic) were compared using one-way fixed analysis of variance (ANOVA). Significant differences between the algorithm
scores for children with autism and non-spectrum children within each of the three domains were reported at the p < .0001 level (Lord et al., 1994), suggesting excellent discriminant validity.

**Autism Diagnostic Observation Schedule-Generic**

The Autism Diagnostic Observation Schedule-Generic (ADOS-G) is an observational instrument that provides a summary of an individual’s behavioral strengths and limitations utilizing a standardized administration (Lord et al., 2000). The child’s spontaneous social and communicative behaviors are evoked using various standard play items and situations and naturalistic presses from the evaluator. The ADOS-G has four modules based on development and language levels. These range from no expressive or receptive language to fluent speech. Behaviors are coded in four areas: (a) Communication (ADOS-C), (b) Social Reciprocal Interaction (ADOS-SRI), (c) Play (ADOS-P), and (d) Stereotyped Behaviors and Restricted Interests (ADOS-SB). Diagnostic algorithms and thresholds based on the DSM-IV differ across these modules. The ADOS-G algorithms indicate a diagnosis of autism, ASD (including PDD-NOS), and non-ASD. Unlike the ADI-R, the ADOS-G provides a non-autism ASD diagnostic algorithm based on a lower threshold for each domain and the total score (Lord et al., 2000).

Forty-five to 59 subjects were included in the reliability analyses in each module (Lord et al., 2000). One half of the participants had autism, one third had PDD-NOS, and one sixth had non-spectrum disorders. The reliability of domain scores was computed across pairs of raters for algorithm subtotals and total scores for each module separately. For the social domain, correlations ranged from .88 to .97. For the communication
domain, correlations ranged from .74 to .90. For the social-communication total scores, correlations ranged from .84 to .98. For the restricted, repetitive behaviors, correlations ranged from .75 to .90 (Lord et al., 2000). Thus, based on Cicchetti’s (1994) criteria, these are acceptable correlations. However, Bracken’s (1987) more stringent criteria suggest that internal consistency correlations that fall short of .90 are not adequate.

In this same study, internal consistency was assessed using Cronbach’s alpha. Cronbach’s alpha ranged from .86 to .91 in the social domain and .74 to .84 in the communication domain (Lord et al., 2000). In the restricted, repetitive behaviors domain, the estimate of internal consistency for Modules 1 and 2 ranged from .63 to .65, and for Modules 3 and 4, the range was .47 to .56. For the social-communication totals, the Cronbach’s alphas ranged from .91 to .94 for all modules (Lord et al., 2000). Internal consistency in the restricted, repetitive behaviors domain across all modules fell short of both the Bracken (1987) and Cicchetti (1994) criteria for acceptable homogeneity within this domain. Internal consistency within the social and communication domains as well as social-communication are acceptable, according to Cicchetti.

Discriminant validity was evaluated within the domains of social, communication, and restricted and repetitive behaviors as well as within the social-communication total (Lord et al., 2000). The algorithms for each group (i.e., autistic and non-autistic) were compared using one-way, fixed ANOVAs. Significant differences between the algorithm scores for children with autism and non-spectrum children in the two domains of social and communication and in the social-communication total were reported at the p < .01 level (Lord et al., 2000). This indicates that the social and communication domains of the ADOS-G differentiate well between those individuals on the autism spectrum from those
not on the spectrum in all four modules. Differences in algorithm scores for the restricted and repetitive behaviors domain between children on the autism spectrum and those not on the spectrum were not significant for any of the four modules. This indicates that the restricted and repetitive behaviors domain of the ADOS-G (all modules) does not differentiate well between those individuals on the autism spectrum and those not on the spectrum.

**Gilliam Autism Rating Scale**

The Gilliam Autism Rating Scale (GARS; Gilliam, 1995) is an informant-report questionnaire of 42 items designed to assess the probability that a child has autism. Persons aged 3 to 22 can be assessed for probability of autism using this instrument. The 42 items are grouped into four subscales: (a) social interaction, (b) communication, (c) stereotyped behaviors, and (d) developmental disturbances. Items are rated from never observed to frequently observed. Each subscale is totaled and converted to a standard score with a mean of 10 and standard deviation of 3. An overall standard score or Autism Quotient with a mean of 100 and standard deviation of 15 is derived from four, three, or two of the four subscales. Autism Quotient scores are categorized according to 1 of 7 levels of probability of having autism, ranging from low probability to high probability.

Internal consistency for the three subscales and the total score were estimated with coefficient alphas of .84 for stereotyped behavior, .86 for communication, .88 for social interaction, and .84 for the Autism Index. Inter-rater reliability for the GARS was demonstrated by 16 participants being rated by their teacher as well as by their parents; these assessment results were correlated for each pair of ratings. The correlations were all .72 and above (p < .01) with the Autism Index correlation of .88 (Gilliam, 1995).
Concurrent validity on the GARS was established through comparison to the Autism Behavior Checklist (ABC). Correlations on ABC and GARS subscales (i.e., social interaction, communication, stereotyped behaviors) were significant at the p < .01 level. The correlation coefficient between the GARS Asperger’s Diagnostic Quotient (ADQ) ABC total was .94 (p < .01), which demonstrates excellent concurrent validity. Discriminant validity was reported as the percentage of cases diagnosed previously by school personnel correctly classified (i.e., autism or non-spectrum), which was 90% based on the GARS ADQ (Gilliam, 1994) and demonstrates excellent discriminant validity.

Gilliam Asperger’s Disorder Scale

The Gilliam Asperger’s Disorder Scale (GADS) is an informant-report questionnaire of 32 items designed to assess the probability that a child has autism. Persons ages 3 to 22 can be assessed for probability of autism using this instrument. It is composed of four subscales: (a) social interaction, (b) restricted patterns, (c) cognitive patterns, and (d) pragmatic skills. Items are rated from never observed to frequently observed. Each subscale is totaled and converted to a standard score with a mean of 10 and standard deviation of 3. An overall standard score or Autism Quotient with a mean of 100 and standard deviation of 15 is derived from four, three, or two of the four subscales. An ADQ of < 69 indicates low-not probable, 70 to 79 indicates borderline, and > 80 indicates high-probable (Gilliam, 1994).

Internal consistency for the four subscales was estimated with coefficient alphas of .70 or higher. The reliability of the ADQ estimated with coefficient alpha was .94, indicating strong internal consistency; that is, it is consistently measuring the same
construct (i.e., AS). Test-retest reliability (i.e., stability reliability) was .93 (p < .01) for the ADQ, which is excellent. This means that informants rated the 10 participants twice within a two-week interval; these quotient scores were then correlated with one another. Inter-rater reliability for the GADS was demonstrated by 16 participants being rated by their teacher as well as by their parents; these assessment results were correlated for each pair of ratings. The correlations were all .72 and above (p < .01) with the ADQ correlation .89 (Gilliam, 1994). These correlation estimates are acceptable based on the Cicchetti (1994) criteria.

The GADS manual (Gilliam, 1994) provides criterion validity, which indicates the effectiveness of an instrument in predicting an individual’s performance on specific activities (Anastasi & Urbina, 1997). With respect to the GADS, criterion validity is a measure of the effectiveness in predicting which people are likely to have AS and which people are not likely to have AS. Criterion validity necessitates a comparison criterion; the author chose the GARS to compare to the GADS based on autism and AS being similar diagnostic categories. The correlation coefficient between the GARS Autism Quotient and the GADS ADQ was .58 (p < .01), which indicates poor concurrent validity. Discriminant validity was based on differences in mean subscale and ADQ scores on the GARS between AD and non-AD. Using these scores, a computer classifying the participants (i.e., AD or non-AD) was correct with 83% accuracy (Gilliam, 1994). This indicates good reliability based on the Cicchetti (1994) criteria.

**Instruments for Assessing Functioning**

In addition to the instruments described above that are specific to ASD symptoms, instruments that assess overall and specific language functioning are typically used in a
comprehensive ASD assessment, as Table 1 illustrated (Lord & Corsello, 2005; Risi et al., 2006). Autism spectrum-specific instruments allow for comparison of the individual child to children on the autism spectrum with respect to the three core characteristics of ASDs. Instruments assessing functioning provide a comparison of the individual child to children in the general population on important components of overall functioning (e.g., intellectual abilities, adaptive skills) (Klin, Saulnier, Tsatsanis, & Volkmar, 2005). Instruments often used in ASD assessments that compare functioning of children with a possible ASD to normative samples include (a) measures of cognitive functioning such as the Wechsler scales of intellectual functioning (e.g., Wechsler Preschool and Primary Scale of Intelligence-III [WPPSI-III], Wechsler Intelligence Scale for Children-IV [WISC-IV]), (b) measures of receptive and expressive language (e.g., Peabody Picture Vocabulary Test [PPVT-III], Expressive Vocabulary Test [EVT]), and (c) measures of adaptive functioning (e.g., The Vineland Adaptive Behavior Scales-II [VABS-II]). Measures of functioning displayed in Table 1 are described below. Only those psychometric properties pertinent to the current study are discussed.

Wechsler Preschool and Primary Scale of Intelligence

The WPPSI-III (Wechsler, 2002) was developed to assess the intelligence of children between age 2 years, 6 months and 7 years, 3 months. A child’s general intellectual functioning is represented using a Full Scale IQ (FSIQ) with a mean of 100 and a standard deviation of 15. The FSIQ is derived from a combination of verbal items (the Verbal IQ [VIQ]) and relatively nonverbal items (Performance IQ [PIQ]). Additional tasks can be used to obtain a General Language Composite (GLC) and, with older children, a Processing Speed Quotient (PSQ; Wechsler, 2002).
The psychometric properties of the WPPSI-III are strong based on both the Cicchetti (1994) and Bracken (1987) criteria. The average FSIQ internal consistency was .96, with each age group interval having an internal consistency coefficient of .95 or higher. Test-retest reliability (with an average interval of 26 days) ranged from .86 to .92. Concurrent validity has been examined by comparing the WPPSI-III to six other standardized test measures: (a) WPPSI-R, (b) WISC-III, (c) Bayley Scales of Infant Development (BSID-II), (d) Differential Ability Scales (DAS), (e) Wechsler Individual Achievement Test-II (WIAT-II), and (f) Children’s Memory Scale (CMS). Correlations of composite scores ranged from .31 to .89, demonstrating wide variability in the concurrent validity across comparison instruments. In addition, validity evidence for the WPPSI-II in several specific groups of children was presented. For example, children with autism had significantly lower VIQs than PIQs when compared to typically developing children (Wechsler, 2002).

**Wechsler Abbreviated Scale of Intelligence**

The Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999) was designed as a short, reliable measure of intelligence for ages 6 to 89 years. The WASI provides a FSIQ as well as estimates of VIQ and PIQ consistent with the other Wechsler scales. It contains four subtests with means of 100 and standard deviations of 15, which can be used to estimate likely ranges of the WISC-III or WAIS-III scores for 68% and 90% confidence intervals. Test-retest reliability ranged from .83 to .95. Split-half reliability ranged from .92 to .96 for the Indexes and FSIQ. These correlations suggest good to excellent reliability based on the Cicchetti (1994) criteria. Concurrent validity was assessed by comparing the WASI to the WIAT, WAIS-II, and WISC-III, with
coefficients ranging from .66 to .88 for subtests and .76 to .92 for Indexes and the FSIQs. These correlations demonstrate inadequate to good reliability for the subtests and adequate to excellent reliability for the Indexes and the FSIQs (Wechsler, 1999).

**Wechsler Intelligence Scale for Children-IV**

The WISC-IV (Wechsler, 2003) is designed to assess the cognitive and problem-solving abilities of children for children between ages 6 years and 16 years, 11 months. It consists of 10 core and five supplemental subtests and yields the (a) FSIQ, (b) Verbal Comprehension Index (VCI), (c) Perceptual Reasoning Index (PRI), (d) Working Memory Index (WMI), and (e) Processing Speed Index (PSI; Wechsler, 2003).

The psychometric properties of the WISC-IV are very strong. Internal consistency was estimated using the split-half reliability, resulting in coefficients of .96 or .97 at each age group for the FSIQ. Index coefficients ranged from .94 (VCI) to .88 (PSI). The correlations above .90 were excellent, with the index correlation of .88 on the PSI being good. Concurrent validity was examined by comparing the WISC-IV to four other standardized test measures: (a) the WISC-III, (b) the WPPSI-III, (c) the WAIS-III, and (d) the WASI). Correlations of the FSIQ scores were all .89 or higher, which is considered good. In addition, validity evidence for the WISC-IV in several specific groups of children was presented. For example, children with autism scored significantly lower than matched controls on all composites (Cicchetti, 1994).

**Peabody Picture Vocabulary Test-III**

The Peabody Picture Vocabulary Test-III (PPVT-III; Dunn, Dunn, Williams, & Wang, 1997) is designed to measure one specific aspect of language functioning—receptive vocabulary—in ages 2 years, 6 months to 90 years. Two forms of the PPVT-III
exist; each have 204 items grouped into 17 sets of 12 items. For each set of items, the three easiest are first, the six most difficult items are randomly placed in the middle of the set, and three additional easy items end the set (Dunn et al., 1997).

The psychometric properties of the PPVT-III are strong. Alternate-form reliability coefficients between Forms IIIA and IIIB ranged from .88 to .96 for standard scores, demonstrating good to excellent reliability (Cicchetti, 1994). Alternative-form reliability can be used to assess reliability when two parallel forms of the construct of interest are available, as with the PPVT-III (Forms A and B). Internal consistency measured by Cronbach’s alpha was .95 (mean for Forms A and B). Based on both Cicchetti (1994) and Bracken (1987) criteria, this demonstrates excellent internal consistency. Test-retest reliability for several age groups all exceeded .91. Internal validity was assessed by (a) item homogeneity, (b) item growth curves, and (c) age differentiation. Criterion validity studies were completed comparing the PPVT-III to the WISC-III, as well as three other measures. The PPVT-III scores correlated slightly higher with the WISC-III VIQ (.91 to .92 for the two forms) than with the FSIQ (.90 for the two forms). These correlations suggest the concurrent validity is excellent. Scores on the PPVT-III among special groups (i.e., learning-disabled, speech-impaired, child and adult mentally retarded, hearing-impaired, gifted and talented individuals) were evaluated compared to the general population; however, children on the autism spectrum were not specifically included in the validation studies (Dunn et al., 1997).

**Expressive Vocabulary Test**

The EVT (Williams, 1997) is designed to assess other aspects of language functioning, expressive vocabulary, and word retrieval for ages 2 years, 6 months to 90
years. The EVT consists of 190 items. The first 38 items are labeling items, while for the remaining items, the examiner presents a picture and a word, and the examinee produces a synonym (Williams, 1997).

The psychometric properties of the EVT are not as strong as the psychometric properties of the PPVT-III. The median internal reliability coefficient was .95 with all age-based intervals in the .90s, which demonstrates excellent internal reliability. Test-retest reliability estimates (with a mean interval of 42 days) fell between .77 and .90 for the sample of 226 participants, which are acceptable (Cicchetti, 1994). Criterion validity studies were completed comparing the EVT to the WISC-III, as well as two other measures. The EVT scores correlated higher with the WISC-III VIQ (.72) than with the FSIQ (.68). The former correlation is adequate; however, the correlation between the EVT and the FSIQ is inadequate based on the Cicchetti (1994) criteria.

**Vineland Adaptive Behavior Scales-II**

The VABS-II (Sparrow, Cicchetti, & Balla, 2005) is designed to assess adaptive functioning for ages birth to 90 years. Both parent/caregiver and teacher rating forms are available. The parents/caregiver rating can be administered by Survey Forms (checklist rating) and Interview and Rating Forms (semi-structured interview). The VABS-II assesses adaptive behavior in the four domains of (a) communication, (b) daily living skills, (c) socialization, and (d) motor skills (Sparrow et al., 2005).

Internal consistency using the split-half reliability estimates were .93 and greater for the Adaptive Behavior Composite in all age groups, which demonstrates excellent reliability (Cicchetti, 1994). Test-retest reliability was assessed using the two parent/caregiver forms (interval ranged from 13 to 34 days) and resulted in Adaptive
Behavior Composite correlation coefficients of .81 to .96 across six age ranges. This demonstrates good to excellent reliability based on the Cicchetti (1994) criteria.

Concurrent and discriminant validity for the VABS-II was assessed by comparing it to other adaptive measures (VABS, Adaptive Behavior Assessment System [ABAS], Behavior Assessment System for Children-II [BASC-II], WISC-III, and WASI-III). The VABS-II correlated relatively highly with other measures of adaptive behavior, while it correlated poorly (as expected) with measures of intelligence. For example, the correlation between the VABS-II Adaptive Behavior Composite and the WISC-III FSIQ was .12, which demonstrates that the VABS-II and the WISC-III measure different constructs (Sparrow et al., 2005).

**Psychometric Studies of Autism Spectrum Assessment Instruments**

Studies of ASD assessment instruments fall into four main categories of study types: (a) issues surrounding obtaining an ASD diagnosis (e.g., screening, practice parameters, effort, and problems; Eaves & Ho, 2004; Holzer et al., 2006; Sciutto & Cantwell, 2005), (b) differential diagnosis of PDD subtypes (Dixon et al., 2009; McConachie, LeCouteur, & Honey, 2005; NoterDaeme, Mildenberger, Sitter, & Amorosa, 2002; Sciutto & Cantwell, 2005), (c) initial psychometric reliability and validity studies used to develop and/or revise instruments (Campbell, 2005; Lord et al., 1994), and (d) symptoms and diagnosis across the lifespan (McGovern & Marian, 2005; Moore & Goodson, 2003; Turner, Stone, Pozdol, & Coonrod, 2006). Studies designed to initially validate or revise ASD are summarized in the paragraphs describing each instrument individually. Two studies have compared the psychometric properties of ASD instruments when used in community samples for the purposes of diagnosis and
intervention. These two studies are presented in detail below.

A small number of experimental studies have compared the ADI-R and ADOS-G (Chawarska et al., 2007; Risi et al., 2006; Ventola et al., 2006). For example, Risi et al. (2006) compared sensitivity and specificity for the ADOS-G and ADI-R in a large sample of children across diagnostic centers. The purpose of the Risi et al. study was to propose standard criteria for the combined use of the ADI-R and ADOS-G to diagnose cases of autism. Data from 1,039 diagnostic evaluations at two Midwestern university-based clinics specializing in suspected ASD were used. ADI-R total scores correlated .57 with the ADOS-G total scores. In the subsample of children younger than 36 months, the ADI-R total score correlated .60 with the ADOS-G total score. In the subsample of children older than 36 months with profound mental retardation, the ADI-R total scores correlated .28 with the ADOS-G total score with a range of .23 to .95 for subdomains (Risi et al., 2006). In the area of ASD diagnostics, Risi et al. showed that the combined information from the ADI-R and the ADOS-G better reflect consensus clinical judgments of autism and ASD than any single instrument.

However, few studies have evaluated ASD instruments in clinical, community-based settings (Mazefsky & Oswald, 2006; South et al., 2002). Mazefsky and Oswald (2006) reviewed data from 75 assessments of children with suspected autism at a university medical center clinic specializing in ASD evaluations. Children were between the ages of 22 months and 8 years. Assessment instruments used included the (a) ADOS-G, (b) ADI-R, (c) GARS, and (d) multidisciplinary clinical evaluation. The clinical evaluation team included (a) a licensed clinical psychologist, (b) a child psychiatrist, (c) an education specialist, (d) a speech/language pathologist, and (e) an occupational
therapist, all of whom had 5 to 20 years of experience with children with autism. Although many of the children were already receiving early intervention or special education, the purpose of the evaluation was to clarify the child’s disorder and recommend interventions (Mazefsky & Oswald, 2006).

Agreement between the ADOS-G and the team diagnosis occurred in 58 of 75 cases, resulting in 77% agreement. In 12 of the 75 cases (16%) the ADOS-G resulted in false positives (i.e., ADOS-G indicated autism when the team assigned another PDD). In 5 of the 75 cases (6%), the ADOS-G resulted in a false negative (i.e., the ADOS-G indicated PDD other than autism when the team assigned a diagnosis of autism).

Agreement between the ADI-R and the team diagnosis occurred in 55 of 75 cases (73% agreement). Twelve of 17 cases (16%) resulted in false positives, and 8 of 75 cases (11%) resulted in false negatives based on the ADI-R algorithm. The mean GARS Autism Quotient for those children with a clinical diagnosis of autism was 87.87 (SD = 10.71), statistically different (p < .001) than the mean for the standardization sample (mean = 100). The majority of children with autism were classified by the GARS as falling at a “below average” or lower probability of having autism. Thus, in this study, the GARS underestimated the probability of having autism (Mazefsky & Oswald, 2006).

South et al. (2002) reported similar community-based findings regarding the GARS, utilizing 119 cases from four university specialty clinics. These children were diagnosed by one or more “internationally recognized autism experts” as meeting the criteria for autism based on one or more of the following assessment instruments: (a) ADI-R, (b) ADOS-G, and (c) VABS-II. The sample mean for the GARS Autism Quotient was 90.10, significantly below the standardized mean of 100 (South et al., 2002).
Administration of the GARS, if used as a diagnostic measure as South et al. suggest, may occur in a non-specialty setting and has a risk of a high false negative rate as the sensitivity index is only .48.

Current Study

The purpose of the current study was to compare the findings obtained with a community sample in a clinical setting to initial psychometric studies of various ASD instruments conducted in research environments and to compare psychometric properties of ASD instruments to one another. The data for this community sample were archival data from two outpatient clinics serving children with possible ASDs that were all supervised by the same licensed clinical psychologists over the span of several years. Several relationships were analyzed in this archival study.

First, various psychometric properties, such as (a) internal consistency, (b) inter-rater reliability, (c) concurrent validity, and (d) convergent and discriminant validity derived from this community sample were compared to published psychometric properties. Second, correlations between ASD instruments that have not been evaluated in the community-based published literature were investigated, including comparisons of the ADI-R and ADOS-G and comparisons of the ADOS-G and ADI-R with the GARS and the GADS. Lastly, comparisons of the receptive language between children with a final diagnosis of ASD and those with a final diagnosis of non-spectrum as well as the comparison of expressive language between the groups (i.e., autism and non-spectrum) were evaluated.
METHOD

Cases Reviewed

Seventy-seven cases were reviewed for this study. The majority of these cases (92%; 71 cases) represents clients of a small, university-based outpatient specialty clinic serving families of children aged 2 to 17 suspected of or identified with ASDs. Less than 10% of these cases (6 cases) represent clients of a university outpatient generalist psychology training clinic serving a community population of approximately 250,000. All clients presented for a diagnostic evaluation of a potential ASD, and typical referral sources included (a) school professionals, (b) other mental health providers, (c) other Western Michigan University (WMU) training clinics, and (d) various community sources. All evaluations were supervised by the same licensed clinical psychologist who specializes in ASDs. Graduate student training at the Center for Autism (CFA) included a four-hour training with a licensed clinical psychologist and an apprentice model with less experienced therapists observing more experienced therapists or the licensed psychologist administering, scoring, and interpreting the diagnostic instruments prior to independent use. In addition, the supervising Ph.D. was onsite for the vast majority of client contacts supervising and providing assistance with the evaluations. Cases were discussed weekly with the supervising clinical psychologist. One graduate student attended clinical training for the ADOS-G, and held trainings for other student therapists on the ADOS.

Variables Coded

Demographics. Each case file was coded for several demographic variables, including (a) age of the child at intake, (b) sex of the child, (c) primary caregiver (mother,
father, grandparent, guardian, adoptive parent), (d) mother/father ages, (e) caregiver relationship to the child (biological, foster, adoptive, guardian), (f) number of siblings, (g) diagnosis at intake (if applicable), and (h) final diagnosis.

**Assessment scores.** Total score, subscale scores, and, in some cases, individual items for all of the ASD diagnostic instruments were compiled into Excel spreadsheets (see Appendix) by trained coders. These included the ADI-R and ADOS-G individual test items and subscales for (a) social reciprocity, (b) communication, and (c) restricted, repetitive behaviors. The ADOS-G Module number (1-4) was coded. The GARS or GADS was coded for the (a) informant (up to four informants), (b) Autism Quotient score, (c) quotient age equivalent, (d) quotient percentile, (e) subscale standard scores (i.e., social interaction, communication, and restricted, repetitive behaviors), and (f) individual test item responses for each subscale. The Wechsler scale (WPPSI-II, WISC-IV, or WASI) scores were recorded for the FSIQ. The WPPSI-II and the WASI scores for VIQ and PIQ were recorded. In addition, the WPPSI-II processing speed quotient (PSQ) and composite language scores were recorded. The WISC-IV (a) VCI, (b) PRI, (c) PSI, and (d) WMI were recorded. The VABS-II informant(s) and type (survey or interview) was coded. Adaptive behavior measured by the VABS-II total score age equivalent and percentile was recorded. Receptive and expressive language functioning measured by the PPVT-III and EVT age equivalent and percentile scores were also recorded.

**Coding Procedure and Inter-coder Agreement**

All variables were coded by examining the individual records for each case (see Appendix). The researcher served as the primary coder for all files. Each case file was reviewed by instrument and coded based on the Data Coding Sheet. Then, the assessment
report included in each case file was reviewed and compared to the data coded. Fewer than 10 discrepancies between the instrument (e.g., ADQ score) and the assessment report were found. In those cases, the instrument was used to code the data.

Three undergraduate research assistants were trained by the researcher as secondary coders for coding a subset of cases. Inter-coder agreement (ICA) was assessed for all variables for 30% of cases (23 cases). ICA was calculated using point-by-point agreement (i.e., number of agreements divided by the number of agreements plus disagreements, multiplied by 100%) for each case. An agreement is defined as an exact match between coders (e.g., exact value entered for the GADS Autism Diagnostic Quotient). The mean ICA percentage was 93%, with a range of 85%-100% ICA. Discrepancies were reviewed by the researcher prior to finalizing the data entry by comparing the two coded entries to the paper protocol. All discrepancies were reconciled in the database prior to conducting analyses; that is, the original data (e.g., GARS questionnaire, ADI-R Diagnostic Algorithm) were reviewed for discrepantly coded items and remediated based on the original data.

**Statistical Analyses**

Analyses of the finalized database involved examining the psychometric properties of individual instruments as well as conducting between instrument comparisons of indicators of characteristics and functioning level (see Table 1). First, four instruments designed to assess the characteristics of ASDs (i.e., ADI-R, ADOS-G, GARS, GADS) were examined. Each of these four instruments was evaluated for the psychometric properties of reliability and validity. Second, multiple instruments designed to assess the level of cognitive, adaptive, or language functioning (i.e., Wechsler scales,
VABS-II, PPVT-III, EVT) were examined. Unless otherwise specified, correlations were estimated using the Pearson Product Moment Correlation; that is, correlations between continuous variables or variables measured at an interval level (Trochim, 2000).

Psychometric properties of ASD instruments. For instruments assessing the characteristics of ASD (i.e., ADI-R, ADOS-G, GARS, GADS), reliability was estimated by internal consistency (Cronbach’s alpha) and by inter-rater reliability for the GARS and GADS only. Internal consistency reliability using Cronbach’s alpha is essentially the mean of all possible split-half estimates without calculating each split-half but utilizing the mathematical formula derived by Cronbach (Trochim, 2006). Internal consistency estimates are reported in the social reciprocity, communication, and restricted, repetitive behaviors domains for the ADI-R and the ADOS. Internal consistency for the GARS was estimated with coefficient alphas for (a) stereotyped behavior, (b) communication, (c) social interaction, and (d) the Autism Index score. The consistency of the GADS subscales of (a) social interaction, (b) restricted patterns, (c) cognitive patterns, (d) pragmatic skills, and (e) the ADQ are estimated with coefficient alphas.

Inter-rater reliability was estimated for the GARS and the GADS. In cases with multiple raters of the GARS (e.g., mother, father, grandmother), the inter-rater reliability was estimated in the subscales of (a) social interaction, (b) communication, (c) stereotyped behaviors, and (d) the Autism Quotient. Mother raters were compared to father raters and then to the next closest relative (i.e., grandmother, aunt). That is, correlations for each pair of raters were analyzed and aggregated into a mean overall inter-rater reliability, resulting in four reliability coefficients (i.e., social interaction, communication, stereotyped behaviors, and Autism Quotient). The mean correlations
were compared to the means presented in the GARS manual. Inter-rater reliability utilizing the same process was estimated for the GADS (i.e., social interactions, communication, stereotyped behaviors, pragmatic skills, and ADQ).

Discriminant validity (i.e., how well the instrument discriminated between those children with autism spectrum disorders and those with no autism spectrum disorder) for the current sample was estimated similarly to the published discriminant validity for the ADI-R and the ADOS-G within the domains of social, communication, and restricted and repetitive behaviors, using one-way fixed analysis of variance (ANOVA).

**Analysis of ASD characteristics.** As in the South et al. (2002) and Mazefsky and Oswald (2006) studies, the sample mean GARS Autism Quotients for those children with a final diagnosis of autism were compared to the standardized mean of 100. Additionally, the measurement of characteristics or traits (i.e., social, communication, restricted, repetitive behaviors) measured by the ADI-R (interview method), ADOS-G (observation method), and GARS or GADS (checklist method) was compared using ANOVA to specifically address how the measurement of these characteristics is different or similar across methods.

For instruments assessing the characteristics of ASD (i.e., ADI-R, ADOS-G, GARS, GADS), validity was estimated using concurrent validity, one type of criterion validity. **Concurrent validity** refers to the degree to which different measures of the same construct measured at the same time correlate. In the case of the ASD instruments, several different subscales purport to measure the ASD core constructs of (a) social impairment, (b) communication impairments, and (c) restricted and repetitive behaviors. Correlations between subscale scores on social, communication, and restricted and
repetitive behaviors on the ADI-R and the ADOS-G (three correlations), the ADI-R and
the GARS (three correlations), and the ADOS-G and the GARS (three correlations)
estimated concurrent validity for each subscale. Similarly, correlations between subscale
scores on social and restricted and repetitive behaviors between the ADI-R and the
GADS (two correlations) and the ADOS-G and GADS (two correlations) estimated
concurrent validity for these two subscales. Discriminant validity shows that constructs
that should not be related are, in reality, not related. For example, subscale scores
measuring different core characteristics of ASD (e.g., social impairments and
communication impairments) should theoretically be poorly correlated. Generally, a poor
correlation is less than .20 (Trochim, 2000).

**Analysis of functioning measures.** Statistical analyses of instruments assessing
adaptive and language functioning (i.e., Wechsler scales, VABS-II, PPVT-III, EVT) were
undertaken. Scores on the VABS-II (i.e., adaptive behavior) were compared for those
with a final diagnosis of ASD or non-ASD using the t-test inferential statistic. This
provides an estimate of how well adaptive behavior differentiates those on the spectrum
from those not on the spectrum. A correlation coefficient was calculated for the PPVT-III
scores and the WISC-IV VCI scores as both largely measure lexical language. Thus,
these two measures administered at approximately the same time (i.e., within four weeks
of one another) are expected to result in similar scores as they measure a similar construct
of language (i.e., an estimate of concurrent validity).

In terms of degree of impairment in language functioning, comparisons of the
receptive language between children with a final diagnosis of ASD and those with a final
diagnosis of non-spectrum were evaluated and compared to previously published
findings. A comparison of expressive language (i.e., EVT) between the groups (i.e., autism and non-spectrum) was calculated. The t-test inferential statistic was used to assess whether group means on the PPVT-III and EVT were statistically different from one another. Similar t-tests were used to assess receptive and expressive language for those with an ASD and those with a final diagnosis that was not on the autism spectrum. In addition, the discrepancy (i.e., absolute difference) between receptive (PPVT-III) and expressive language (EVT) for each individual was calculated, aggregated into group means for the autism group and non-spectrum group, and compared using the t-test. Finally, individual differences between receptive and expressive language was assessed by comparing the mean absolute difference in PPVT-III and EVT scores for children with autism (excluding AS and PDD-NOS) compared to children not on the spectrum.

RESULTS

Data

Demographic Data

The average age at initiation of diagnostic assessment was 7 years, 9 months (94.4 months) with a range of 2 years (24 months) to 17 years, 1 month (205 months). Of the reviewed cases, 65 (84%) were male, and 12 (16%) were female (5.4 males: 1 female), roughly matching the gender estimates for the population with ASDs (4.3 males: 1 female; Fombonne, 2005).

The majority of children (n = 74; 96%) were in the custody of their biological parents, while two children (3%) were adopted and one child (1%) had a legal guardian. The primary caregiver was most often the mother (n = 69; 93%). In addition, two fathers
(3%), one grandmother (1%), and two legal guardians (3%) were primary caregivers. The primary caregiver was most often the primary interviewee (e.g., ADI-R, Vineland) and included 70 biological parents (94%), two adoptive parents (3%), and two legal guardians (3%). The majority (79%) of children had siblings (n = 50); however, 21% of children did not have any siblings (n = 13). In 11 cases, the presence and the number of siblings were unavailable. Of those with siblings, 42% had one sibling, 32% had two siblings, 10% had three siblings, 8% had four siblings, 0% had five siblings, 6% had six siblings, and 2% had seven siblings.

**Diagnostic Assessment Data**

Diagnostic assessments (n = 74) were conducted between February 2001 and June 2007 with the majority occurring in 2006 and 2007. As shown in Table 2, seven assessments were conducted in 2001 (10%), 11 in 2004 and in 2005 (15% each year), 27 (36%) in 2006, and 18 in 2007 (24%). No diagnostic assessments were conducted in 2002 and 2003 as the clinical psychologist supervising the diagnostic assessments was on maternity leave.

Table 2

<table>
<thead>
<tr>
<th>Year</th>
<th>n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2001</td>
<td>7 (10)</td>
</tr>
<tr>
<td>2002</td>
<td>—</td>
</tr>
<tr>
<td>2003</td>
<td>—</td>
</tr>
<tr>
<td>2004</td>
<td>11 (15)</td>
</tr>
<tr>
<td>2005</td>
<td>11 (15)</td>
</tr>
<tr>
<td>2006</td>
<td>27 (36)</td>
</tr>
<tr>
<td>2007</td>
<td>18 (24)</td>
</tr>
</tbody>
</table>
**Time 1/Time 2-assessment diagnosis.** Table 3 shows that the majority of children (n = 74) presented with no previous ASD diagnosis (n = 26; 35%), whereas seven (9.5%) presented with a previous diagnosis of autistic disorder, four (5.4%) presented with a previous diagnosis of AS, and five (6.8%) presented with a diagnosis of PDD-NOS. Additionally, one child (1.4%) presented with a previous diagnosis of mental retardation (MR), and 31 (42%) children presented with a non-mental retardation (non-MR) and non-autism spectrum disorder (non-ASD) diagnosis. Of those 31 children, 19 (25.7%) were previously diagnosed with Attention Deficit Hyperactivity Disorder (ADHD), and 12 (16.2%) were diagnosed with a variety of disorders (e.g., bipolar, mood, anxiety disorders) with no more than two with any particular one of those diagnoses.

As a result of the completed diagnostic assessment (n = 74), eight children (10.8%) were not assigned any DSM-IV diagnosis, 14 (18.9%) received a diagnosis of autistic disorder, almost a quarter (24.3%; n = 18) received a diagnosis of Asperger’s disorder, 21 children (28.4%) received a PDD-NOS diagnosis, and two children (2.7%) were assigned a diagnosis of MR with no other diagnosis given. In addition, the diagnostic evaluation resulted in four children (5.4%) receiving a diagnosis of ADHD, four children (5.4%) being assigned other mental health diagnoses (e.g., anxiety disorder-NOS, bipolar, mood disorder-NOS), one child (1.4%) being diagnosed with a language disorder, and two children (2.7%) being referred for additional evaluation (i.e., tic disorder, genetic anomaly). Excluding those children with no previous assessment, 18 (36%) of the final diagnoses confirmed the original diagnosis, and 32 (64%) of the final diagnoses differed from the original diagnosis of the child. Of the 32 children with final diagnoses that differed from the original, the majority (18) switched from a non-ASD
diagnosis to ASD (i.e., most often ADHD to AS), seven switched from non-ASD to non-ASD or no diagnosis for the final diagnosis (e.g., ADHD to “none”, bipolar to ADHD), four switched among the ASDs (e.g., AS to PDD) and two moved from an ASD (i.e., AS) to no diagnosis.

Table 3

**Time 1/Time 2-Assessment Diagnoses**

<table>
<thead>
<tr>
<th></th>
<th>Pre (%)</th>
<th>Post (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Diagnosis</td>
<td>26 (35.0)^a</td>
<td>8 (10.8)^b</td>
</tr>
<tr>
<td>ASD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autistic Disorder</td>
<td>7 (9.5)</td>
<td>14 (18.9)</td>
</tr>
<tr>
<td>Asperger’s Disorder</td>
<td>4 (5.4)</td>
<td>18 (24.3)</td>
</tr>
<tr>
<td>PDD-NOS</td>
<td>5 (6.8)</td>
<td>21 (28.4)</td>
</tr>
<tr>
<td>Non-ASD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MR</td>
<td>1 (1.4)</td>
<td>2 (2.7)</td>
</tr>
<tr>
<td>ADHD</td>
<td>19 (25.7)</td>
<td>3 (4.1)</td>
</tr>
<tr>
<td>Other</td>
<td>12 (16.2)</td>
<td>6 (8.1)</td>
</tr>
<tr>
<td>Referral for non-ASD assessment</td>
<td>2 (2.7)</td>
<td></td>
</tr>
</tbody>
</table>

^aNo previous assessment. ^bNo diagnosis assigned.

**Diagnostic instruments.** Table 4 provides a summary of the different instruments used during assessment and the number of respondents, where appropriate. Sixty-three ADI-Rs were conducted, and the vast majority of respondents (61 of 63) were mothers. Each of the four modules of the ADOS was utilized: (a) eight Module 1, (b) 15 Module 2, (c) 35 Module 3, and (d) seven Module 4, comprising 65 ADOS administrations. Seventy-nine total GARS questionnaires were completed for 34 children. Eight children had one GARS respondent, 11 children had two GARS respondents; 11 more children
had a second and a third informant, and four additional children had second, third and fourth GARS informants.

Table 4

*Diagnostic Instrument Descriptive Data*

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADI-R</td>
<td>63</td>
</tr>
<tr>
<td>ADOS</td>
<td>65</td>
</tr>
<tr>
<td>Module 1</td>
<td>8</td>
</tr>
<tr>
<td>Module 2</td>
<td>15</td>
</tr>
<tr>
<td>Module 3</td>
<td>35</td>
</tr>
<tr>
<td>Module 4</td>
<td>7</td>
</tr>
<tr>
<td>GARS&lt;sup&gt;a&lt;/sup&gt;</td>
<td>79</td>
</tr>
<tr>
<td>GADS&lt;sup&gt;b&lt;/sup&gt;</td>
<td>106</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Functioning</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Vineland Survey Interview</td>
<td>68</td>
</tr>
<tr>
<td>Wechsler</td>
<td></td>
</tr>
<tr>
<td>WPPSI-III</td>
<td>3</td>
</tr>
<tr>
<td>WISC-III</td>
<td>4</td>
</tr>
<tr>
<td>WISC-IV</td>
<td>17</td>
</tr>
<tr>
<td>WASI</td>
<td>9</td>
</tr>
<tr>
<td>PPVT-III</td>
<td>55</td>
</tr>
<tr>
<td>EVT</td>
<td>50</td>
</tr>
</tbody>
</table>

<sup>a</sup>34 cases had 1 or more respondents. <sup>b</sup>44 cases had 1 or more respondents

One hundred and two total GADS questionnaires were completed for 44 children.

Eight children had one GADS respondent, 14 children had two GADS respondents; 18 more children had a second and a third informant, and four additional children had
second, third and fourth GADS informants. The Wechsler scales were administered in 29 diagnostic assessments: (a) three children completed the WPPSI-III, (b) 13 completed the WISC-IV, (c) four completed the WISC-III, and (d) nine completed the WASI. Sixty-eight VABS were conducted with 64 mothers, three fathers and one grandmother being interviewed. Fifty-five children completed the PPVT-III and 50 completed the EVT.

Statistical Analyses

Psychometric properties of ASD instruments. Table 5 compares the published internal consistency for the ADI-R, ADOS, GADS, and GARS to the current sample's internal consistency estimates using Cronbach's alpha. Correlations for the current sample ranged from .20 to .88, below those correlations in the published psychometric studies (Gilliam, 1994, 1995; Le Couteur, Lord, & Rutter, 2003; Lord et al., 2000).

Table 6 compares the GARS and GADS published inter-rater reliability to the current sample inter-rater reliability on the GARS and the GADS. In the sample, inter-rater reliability on the GARS subscale stereotyped behavior and the Autism Quotient reached a similar level of reliability (p < .01) as the initial psychometric study published in the GARS Examiner's Manual. The index of inter-rater reliability on the GARS subscales of communication and social interaction did not achieve reliability at the published level but was statistically significant at the p < .05 level. Additionally, in the current study, there was a 48.0% agreement among pairs of GARS respondents (n = 25) on the GARS Interpretation Guide (i.e., same probability of autism category). In comparing the GADS published inter-rater reliability to the sample, the sample reached a significant reliability level (p < .05) among respondent pairs for the subscale restricted patterns of behavior only. In the current study, there was a 52.7% agreement among pairs
of GADS respondents (n = 36) on the GADS Interpretation Guide (i.e., same probability of Asperger’s category).

Table 5

Comparison of ADI-R, ADOS, GARS and GADS Published Internal Consistency Reliability, and Current Study Internal Consistency Reliability

<table>
<thead>
<tr>
<th></th>
<th>Published</th>
<th>Current Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ADI-R</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Reciprocity</td>
<td>.95*</td>
<td>.88*</td>
</tr>
<tr>
<td>Communication</td>
<td>.84*</td>
<td>.70</td>
</tr>
<tr>
<td>Restricted, Repetitive Behaviors</td>
<td>.69*</td>
<td>.20</td>
</tr>
<tr>
<td><strong>ADOS (all modules)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>.93b*</td>
<td>.55</td>
</tr>
<tr>
<td>Communication</td>
<td>.84b*</td>
<td>.75</td>
</tr>
<tr>
<td>Social + Communication</td>
<td>.92b*</td>
<td>.65</td>
</tr>
<tr>
<td>Restricted, Repetitive Behaviors</td>
<td>.84b*</td>
<td>.58</td>
</tr>
<tr>
<td><strong>GARS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Interaction</td>
<td>.84c*</td>
<td>.77</td>
</tr>
<tr>
<td>Communication</td>
<td>.86c*</td>
<td>.82*</td>
</tr>
<tr>
<td>Stereotyped Behavior</td>
<td>.88c*</td>
<td>.75</td>
</tr>
<tr>
<td>Autism Index</td>
<td>.84c*</td>
<td>.29</td>
</tr>
<tr>
<td><strong>GADS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Interaction</td>
<td>.88d*</td>
<td>.75</td>
</tr>
<tr>
<td>Restricted Patterns</td>
<td>.81d*</td>
<td>.85*</td>
</tr>
<tr>
<td>Cognitive Patterns</td>
<td>.86d*</td>
<td>.83*</td>
</tr>
<tr>
<td>Pragmatic Skills</td>
<td>.84d*</td>
<td>.23</td>
</tr>
<tr>
<td>Asperger’s Diagnostic Quotient</td>
<td>.94d*</td>
<td>.39</td>
</tr>
</tbody>
</table>

Table 6

*Comparison of GARS, GADS Published Inter-rater Reliability and Current Study Inter-rater Reliability*

<table>
<thead>
<tr>
<th></th>
<th>Published</th>
<th>Current Study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GARS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stereotyped Behavior</td>
<td>.82**</td>
<td>.63&lt;sup&gt;b&lt;/sup&gt;&lt;sup&gt;**&lt;/sup&gt;</td>
</tr>
<tr>
<td>Communication</td>
<td>.77**</td>
<td>.42&lt;sup&gt;b&lt;/sup&gt;&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>Social Interaction</td>
<td>.73**</td>
<td>.40&lt;sup&gt;b&lt;/sup&gt;&lt;sup&gt;*&lt;/sup&gt;</td>
</tr>
<tr>
<td>Autism Quotient</td>
<td>.88**</td>
<td>.59&lt;sup&gt;b&lt;/sup&gt;&lt;sup&gt;**&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>GADS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Interaction</td>
<td>.81**</td>
<td>.19&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Restricted Patterns</td>
<td>.84**</td>
<td>.48&lt;sup&gt;c&lt;/sup&gt;&lt;sup&gt;**&lt;/sup&gt;</td>
</tr>
<tr>
<td>Cognitive Patterns</td>
<td>.82**</td>
<td>.31&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Pragmatic Skills</td>
<td>.72**</td>
<td>.19&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Asperger's Diagnostic Quotient</td>
<td>.94**</td>
<td>.33&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>a</sup>n = 23. <sup>b</sup>n = 35. <sup>*</sup>p < .05. **p < .01

Table 7 shows the Analysis of Variance for those children with a final diagnosis of autism spectrum (i.e., AD, AS, PDD-NOS) and those children with a non-spectrum final diagnosis or no diagnosis (i.e., non-MR, MR only, ADHD, a diagnosis in the “other” category, referral for non-ASD) on the ADI-R. The current study findings do not reach statistical significance for differences in the ADI-R as in the published psychometric study (Lord et al., 1994) with only the social domain approaching significance at .08. However, similar to the published psychometric study it shows that among the three core areas of ASD abnormalities measured on the ADI-R, RRB is the least likely to discriminate between those on the spectrum and those not on the spectrum.

Table 8 shows the Analysis of Variance for those children with a final diagnosis of
autism spectrum (i.e., AD, AS, PDD-NOS) and those children with a non-spectrum final diagnosis or no diagnosis (i.e., non-MR, MR only, ADHD, Other, referral for non-ASD) on the ADOS. Findings in the current study show that the ADOS social, communication and imagination/creativity domains and the combined social and communication score each discriminate at a statistically significant level between those children with a final diagnosis of ASD and those with a non-ASD final diagnosis.

Table 7

*Analysis of Variance for Those with an ASD Final Diagnosis and Those with a non-ASD Final Diagnosis on the ADI-R*

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>F</th>
<th>η</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td>(1,55)</td>
<td>3.15</td>
<td>.23</td>
<td>.08</td>
</tr>
<tr>
<td>Communication</td>
<td>(1,55)</td>
<td>2.14</td>
<td>.19</td>
<td>.15</td>
</tr>
<tr>
<td>Restricted, Repetitive Behavior</td>
<td>(1,55)</td>
<td>1.63</td>
<td>.17</td>
<td>.21</td>
</tr>
<tr>
<td>Abnormality Evident in 36 mo.</td>
<td>(1,55)</td>
<td>.32</td>
<td>.08</td>
<td>.58</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01

Table 8

*Analysis of Variance for Those with an ASD Final Diagnosis and Those with a non-ASD Final Diagnosis on the ADOS*

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>F</th>
<th>η</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social</td>
<td>(1,56)</td>
<td>8.52</td>
<td>.36</td>
<td>.005**</td>
</tr>
<tr>
<td>Communication</td>
<td>(1,56)</td>
<td>6.93</td>
<td>.33</td>
<td>.01**</td>
</tr>
<tr>
<td>Social + Communication</td>
<td>(1,56)</td>
<td>10.2</td>
<td>.26</td>
<td>.002**</td>
</tr>
<tr>
<td>Imagination/Creativity</td>
<td>(1,54)</td>
<td>4.01</td>
<td>.39</td>
<td>.05*</td>
</tr>
<tr>
<td>Restricted, Repetitive Behaviors</td>
<td>(1,56)</td>
<td>2.35</td>
<td>.20</td>
<td>.13</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01
Analysis of ASD characteristics. Table 9 shows the mean GARS and GADS quotients by final diagnosis classification. The mean GARS Autism Quotient (AQ) was 84.7 for those children with a final diagnosis of autistic disorder (n = standardized mean of 100), designated as “below average” probability of having AD. The mean GARS AQ was 88.4 for those children with a final diagnosis of PDD-NOS, indicating a “below average” probability of having AD. The mean GARS Autism Quotient was 85.7 for those children with a final diagnosis of “none” or a non-ASD diagnosis (e.g., ADHD), which also results in a “below average” probability of having AD. These three averages for GARS scores fall right at or within one standard deviation (i.e., 15 pts) from the mean of 100 for the normative sample of participants who were diagnosed with autism. That is, regardless of final diagnostic categorization (i.e., AD, PDD-NOS or non-ASD) the average score of these three subgroups were below the mean and designated as “below average” probability of being on the spectrum.

The mean GADS Asperger’s disorder quotient was 91.9 for those children with a final diagnosis of AS (standardized mean of 100, standard deviation of 15). This mean falls in the “likely” probability of AS. The mean GADS Asperger’s disorder quotient for those children with a final diagnosis of “none” or a non-ASD diagnosis was 78.9, which falls in the “borderline” range for the probability of having AS.

Table 10 shows the ANOVA for ASD characteristics among those with a final diagnosis of Asperger’s disorder, while Table 10 shows the ANOVA for ASD characteristics among those with a final diagnosis of autism. Both Tables 9 and 10 illustrate no significant differences in the measurement of ASD characteristics across methods (e.g., interview, observation, and checklist).
Table 9

*GARS/GADS Quotients by Final Diagnostic Classification*

<table>
<thead>
<tr>
<th></th>
<th>Autism Quotient</th>
<th>Asperger’s Disorder Quotient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Autistic Disorder</td>
<td>84.7</td>
<td></td>
</tr>
<tr>
<td>PDD-NOS</td>
<td>88.4</td>
<td></td>
</tr>
<tr>
<td>No diagnosis or non-ASD</td>
<td>85.7</td>
<td></td>
</tr>
<tr>
<td>Asperger’s Disorder</td>
<td></td>
<td>91.9</td>
</tr>
<tr>
<td>No diagnosis or non-ASD</td>
<td></td>
<td>78.9</td>
</tr>
</tbody>
</table>

Table 10

*Analysis of Variance for ASD Characteristics among Those with a Final Diagnosis of Asperger’s*

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>F</th>
<th>η</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Reciprocity</td>
<td>(2, 45)</td>
<td>.36</td>
<td>.13</td>
<td>.70</td>
</tr>
<tr>
<td>Communication</td>
<td>(1, 30)</td>
<td>.74</td>
<td>.15</td>
<td>.40</td>
</tr>
<tr>
<td>Restricted, Repetitive Behavior</td>
<td>(2, 45)</td>
<td>.38</td>
<td>.13</td>
<td>.68</td>
</tr>
</tbody>
</table>

Table 11

*Analysis of Variance for ASD Characteristics among Those with a Final Diagnosis of Autism*

<table>
<thead>
<tr>
<th>Source</th>
<th>Df</th>
<th>F</th>
<th>η</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social Reciprocity</td>
<td>(2, 33)</td>
<td>.67</td>
<td>.20</td>
<td>.52</td>
</tr>
<tr>
<td>Communication</td>
<td>(2, 33)</td>
<td>1.26</td>
<td>.27</td>
<td>.30</td>
</tr>
<tr>
<td>Restricted, Repetitive Behavior</td>
<td>(1, 34)</td>
<td>1.46</td>
<td>.23</td>
<td>.23</td>
</tr>
</tbody>
</table>
The current study estimates of concurrent validity (r) for ASD characteristics are shown in Table 12. The ADI-R/ADOS correlations estimate the informant-examiner ratings of the child’s ASD characteristics utilizing interview and observation methods, respectively. The ADI-R/GARS and the ADI-R/GADS correlations estimate the informant-informant ratings of the child’s ASD characteristics utilizing interview and questionnaire methods. The ADOS/GARS and the ADOS/GADS correlations estimate the examiner-informant ratings of the child’s ASD characteristics utilizing observation and questionnaire methods, respectively. These correlations suggest that different measures of the same construct measured at the same time did not correlate highly, except for restricted and repetitive behaviors (ADI-R) and stereotyped behaviors (GADS) at the p < .01 level.

Table 12

*Current Study Estimates of Concurrent Validity (r) for ASD Characteristics*

<table>
<thead>
<tr>
<th></th>
<th>ADI-R/ADOS&lt;sup&gt;a&lt;/sup&gt;</th>
<th>ADI-R/GARS&lt;sup&gt;b&lt;/sup&gt;</th>
<th>ADI-R/GADS&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Social Reciprocity</td>
<td>Social Reciprocity</td>
<td>Social reciprocity/Social</td>
</tr>
<tr>
<td></td>
<td>Communication</td>
<td>Communication</td>
<td>Restricted, Repetitive Behaviors/Stereotyped Behavior</td>
</tr>
<tr>
<td></td>
<td>Restricted, Repetitive Behaviors</td>
<td>Restricted, Repetitive Behaviors/Stereotyped Behavior</td>
<td></td>
</tr>
</tbody>
</table>
Table 12 – Continued

<table>
<thead>
<tr>
<th>ADOS/GARS&lt;sup&gt;d&lt;/sup&gt;</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social reciprocal interaction/Social</td>
<td>.12</td>
</tr>
<tr>
<td>Communication</td>
<td>.11</td>
</tr>
<tr>
<td>Restricted, Repetitive Behaviors/Stereotyped Behavior</td>
<td>.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ADOS/GADS&lt;sup&gt;e&lt;/sup&gt;</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social reciprocal interaction/Social</td>
<td>.13</td>
</tr>
<tr>
<td>Restricted, Repetitive Behaviors/Stereotyped Behavior</td>
<td>-.03</td>
</tr>
</tbody>
</table>

<sup>*n = 62. *n = 23. *n = 38. *n = 27. *n = 40. *p < .01</sup>

**Analysis of functioning measures.** In terms of verbal comprehension, the correlation between the PPVT-III score and the WISC-IV VCI was .86 (n = 17). This score is considered “good” based on the Cicchetti (1994) criteria. For those with a final diagnosis of ASD and those without an ASD (excluding MR), the PPVT-III mean scores were compared. The ASD group mean (M = 100.9, SD = 23.6) and non-ASD group mean (M = 99.8, SD = 12.2) were not statistically different from one another: t(12) = -.16, p = .44, one-tailed, d = .16. For those with a final diagnosis of ASD and those without an ASD (excluding MR), the EVT mean scores were compared. The ASD group EVT mean (M = 91.3, SD = 21.3) and non-ASD EVT mean (M=91.5, SD =16.9) were not statistically different from one another: t(14) = .04, p = .51, one-tailed, d =.04. In terms of the degree of impairment in language between those with a final diagnosis of ASD (excluding MR) and without an ASD, the absolute difference between the PPVT-III and EVT scores for each individual were compared. The ASD group mean (M = 14.7, SD =10.6) and non-ASD (M=15.3, SD =11.6) group means of the absolute difference
(between the PPVT-III and EVT) were not statistically different from one another: t(14) = .25, p = .59, one-tailed, d = .13. Ten (41.6%) of the 24 children with a final autism or PDD-NOS diagnosis had higher receptive language than expressive language skills (range of absolute difference = 2 - 77). Nine (75%) of the 12 children with a final diagnosis of AS had higher receptive language than expressive language skills (range of absolute difference = 3-37). For those with a final diagnosis of ASD and those without an ASD (excluding MR), the ASD group Adaptive Behavior Composite mean (M = 79.3, SD = 31.1) and non-ASD (M=74.4, SD = 28.4) group means were not statistically different from one another: t(18) = -.63, p = .26, one-tailed, d = .20.

**Case examples.** Case examples (Tables 13 and 14) from the current study illustrate the differential diagnostic purpose of the original evaluation procedures and how these data may differ from that in a psychometric validation study. These cases demonstrate the importance of the MTMM and clinical judgment in assigning differential diagnosis. Table 13 provides the diagnostic assessment data for an 8 year, 1 month-old female. Her presenting symptoms included (a) occasional aggression towards her parents and others (especially during transitions from one activity to another), (b) rigid adherence to play behaviors when interacting with her peers, and (c) preference to play alone. Her scores on the ADI-R and ADOS exceeded cutoffs for an ASD. Her scores did not demonstrate any clinically significant delay in language, cognitive development or the use of age-appropriate self-help skills. However, there was a substantial discrepancy between her receptive language and her expressive language. In this case example, the ADI-R or ADOS scores, if taken alone, indicate an ASD. The GADS scores, if taken alone, provide equivocal direction with the teacher’s score indicating a *high-probable* of
Asperger's, one parent's scores indicating a low-not probable, and one parent's scores indicating borderline probability of Asperger's. The ADI-R and the ADOS scores, coupled with average cognitive functioning, minor impairment in adaptive behavior (due to below average socialization on the VABS-II), and high average receptive language juxtaposed to average but significantly discrepant expressive language strongly suggested an AS diagnosis. This child was assigned an AS final diagnosis.

Table 13

Case Example of 8-Year-Old Female: Diagnostic Assessment Data

<table>
<thead>
<tr>
<th>ASD Characteristics</th>
<th>Score (Cutoff for Autism/ASD)</th>
<th>Supports Asperger’s Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ADI-R</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormalities in Social Reciprocity</td>
<td>8(10)</td>
<td>Yes</td>
</tr>
<tr>
<td>Abnormalities in Communication</td>
<td>6(8)</td>
<td>Yes</td>
</tr>
<tr>
<td>Restricted, Repetitive Behaviors</td>
<td>3(3)</td>
<td>Yes</td>
</tr>
<tr>
<td>Abnormality Evident Before 36 Months</td>
<td>2(1)</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>ADOS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Interaction</td>
<td>2(3/2)</td>
<td>Yes</td>
</tr>
<tr>
<td>Communication</td>
<td>6(6/4)</td>
<td>Yes</td>
</tr>
<tr>
<td>Communication + Social</td>
<td>8(10/7)</td>
<td>Yes</td>
</tr>
<tr>
<td><strong>GADS</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent</td>
<td>63</td>
<td>Low-Not Probable</td>
</tr>
<tr>
<td>Parent</td>
<td>75</td>
<td>Borderline</td>
</tr>
<tr>
<td>Teacher</td>
<td>92</td>
<td>High-Probable</td>
</tr>
</tbody>
</table>
Table 13 – Continued

<table>
<thead>
<tr>
<th>Functioning</th>
<th>Score (Cutoff for Autism/ASD)</th>
<th>Supports Asperger’s Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>WASI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSIQ</td>
<td>113</td>
<td>Yes</td>
</tr>
<tr>
<td>PIQ</td>
<td>108</td>
<td></td>
</tr>
<tr>
<td>VIQ</td>
<td>116</td>
<td></td>
</tr>
<tr>
<td>VABS-II (Adaptive level)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Communication</td>
<td>84 (Average)</td>
<td>Yes</td>
</tr>
<tr>
<td>Daily Living Skills</td>
<td>90 (Average)</td>
<td></td>
</tr>
<tr>
<td>Socialization</td>
<td>71</td>
<td>(Below Average)</td>
</tr>
<tr>
<td>Composite</td>
<td>80</td>
<td>(Below Average)</td>
</tr>
<tr>
<td>PPVT-III</td>
<td>122</td>
<td>Yes</td>
</tr>
<tr>
<td>EVT</td>
<td>97</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Table 14 provides the diagnostic assessment data for a 5 year, 10 month-old male. His presenting symptoms involved delayed social skills and repetitive, restricted, and stereotyped behaviors. His parents reported many verbal and motor rituals, and insistence on sameness, which match the profile of autism. For example, he engaged in whole-body spinning, twirling his hair, and lining things up. In addition, his scores on the PPVT-III and EVT had a statistically significant difference, demonstrating a level of receptive language that was higher than his expressive language.

Early language delay is required for an autism diagnosis, but an absence of language delays is required for an Asperger’s diagnosis. In this case, the parent report
suggested he was inhibited to a point of qualitative impairment in social interactions with both adults and peers. However, during direct observations and assessments with the child during the ADOS, he demonstrated only mild impairment with markedly different behavior in the presence of his parents than when alone with the examiner(s). Additional instruments were used to assess his social skills and problem behaviors (Social Skills Improvement System [SSIS]) in which his former teacher endorsed substantial social problems. His current teacher indicated that many of these difficulties improved, such that his current social behavior was socially appropriate.

This pattern suggested two factors: his day care placement was facilitating social development, and his social anxiety in new situations and with new people, rather than a lack of social understanding or skill, was impacting his behavior. It also highlights the importance of multiple methods and multiple sources in a psychological assessment. In this case, social skills were assessed by interview, observation, and checklist (i.e., questionnaire) from three different sources—parents, examiner, and teacher, respectively.

Table 14

*Case Example of 5-Year-Old Male: Diagnostic Assessment Data*

<table>
<thead>
<tr>
<th>ASD Characteristics</th>
<th>Score</th>
<th>Supports Asperger’s Diagnosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADI-R (Cutoff for Autism)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abnormalities in Social Reciprocity</td>
<td>15(10)</td>
<td>Yes</td>
</tr>
<tr>
<td>Abnormalities in Communication</td>
<td>7(8)</td>
<td>Yes</td>
</tr>
<tr>
<td>Restricted, Repetitive Behaviors</td>
<td>11(3)</td>
<td>Yes</td>
</tr>
<tr>
<td>Table 14 – Continued</td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td><strong>ASD Characteristics</strong></td>
<td><strong>Score</strong></td>
<td><strong>Supports Asperger’s Diagnosis</strong></td>
</tr>
<tr>
<td>Abnormality Evident Before 36 Months</td>
<td>3(1)</td>
<td>Yes</td>
</tr>
<tr>
<td>ADOS (Cutoff for Autism/ASD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Interaction</td>
<td>1(3/2)</td>
<td>No</td>
</tr>
<tr>
<td>Communication</td>
<td>3(6/4)</td>
<td>No</td>
</tr>
<tr>
<td>Communication + Social</td>
<td>4(10/7)</td>
<td>No</td>
</tr>
<tr>
<td>GADS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent</td>
<td>92</td>
<td>High-Probable</td>
</tr>
<tr>
<td>Parent</td>
<td>98</td>
<td>High-Probable</td>
</tr>
<tr>
<td>Teacher</td>
<td>107</td>
<td>High-Probable</td>
</tr>
<tr>
<td>Functioning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WASI</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>FSIQ</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>PIQ</td>
<td>98</td>
<td></td>
</tr>
<tr>
<td>VIQ</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>VABS-II (Adaptive level)</td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Communication</td>
<td>108 (Average)</td>
<td></td>
</tr>
<tr>
<td>Daily Living Skills</td>
<td>100 (Average)</td>
<td></td>
</tr>
<tr>
<td>Socialization</td>
<td>85 (Below Average)</td>
<td></td>
</tr>
<tr>
<td>Motor Skills</td>
<td>81 (Below Average)</td>
<td></td>
</tr>
<tr>
<td>Composite</td>
<td>92 (Average)</td>
<td></td>
</tr>
<tr>
<td>PPVT-III</td>
<td>129</td>
<td>Yes</td>
</tr>
<tr>
<td>EVT</td>
<td>112</td>
<td>Yes</td>
</tr>
</tbody>
</table>
In the second case example, the ADI-R and GADS instrument scores were strongly suggestive of an autism spectrum disorder. The child’s cognitive functioning was average, and his adaptive functioning profile of average adaptive skills in all areas except socialization fit that of Asperger’s disorder. In addition, his language profile showed his receptive language to be significantly higher than his expressive language, with average to above average scores on both dimensions, suggesting Asperger’s disorder. However, his scores on the ADOS as well as informal observations during the WASI, PPVT-III, and EVT suggested that the examiner(s) consider alternative explanations for his delayed social skills and verbal and motor rituals. He was diagnosed with Anxiety Disorder-Not Otherwise Specified, and his parents were encouraged to re-evaluate him for AS in two years if treatment interventions did not resolve the presenting symptoms. This case example illustrates the importance of the MTMM assessment process and clinical judgment in differential diagnosis.

DISCUSSION

Significant progress has been made in the development and refinement of ASD diagnostic tools (Lord, 2010; Lord & Corsello, 2005), such that diagnoses of ASDs in early childhood can be reliable, valid, and stable over time when conducted in research settings (Coonrod & Stone, 2005; Cox et al., 1999; Lord & Corsello, 2005). However, findings about the psychometric properties of tools used in structured research settings may or may not be replicated in community-based clinical settings. Psychometric discrepancies between research and community-based settings are likely due to differing purposes of study designs, examiner training levels, differing measures of reliability and
validity, and sample and respondent characteristics. Several of the diagnostic tools for ASD (i.e., ADOS and ADI-R) prohibit their use for research purposes unless the researchers participate in a rigorous and intensive multi-day training followed by demonstrated performance to an explicit criterion (Ozonoff et al., 2005). These same requirements are not in place for clinicians who use the tools, though briefer trainings are available. These differential requirements suggest that use of the tool and its resulting psychometric properties might be expected to differ in a community sample in a clinical setting compared to the initial psychometric studies by the researchers who developed the tools (Lord et al., 2000). Similar discrepancies have been noted between published norms of the GARS and mean scores on the tool obtained by other researchers (Gilliam, 1995; South et al., 2002; Mazefsky & Oswald, 2006). Given the potential for discrepancies in psychometrics from an original research study to clinical use by others, the current study examined the psychometric properties of ASD instruments when used as part of a comprehensive multi-trait, multi-method (MTMM) assessment in a community sample.

The results from these archival data from two outpatient clinics over a span of several years indicated that the obtained psychometric studies of various ASD instruments were generally poorer than those in the initial development research studies. Internal consistency reliability was consistently lower and more variable across subdomains for the community sample (Table 5) on the ADI-R, ADOS, GARS and GADS compared to the published psychometric properties from research sites. One exception to this was on the GADS, restricted patterns subdomain (.81 vs. .85) in which the obtained correlation narrowly exceeded the published inter-item correlation. Overall, the obtained correlations ranged from “poor” (e.g., .20 on ADI-R, restricted and
repetitive behaviors subdomain) to “good” (e.g., .88 on ADI-R, social reciprocity) based on the Cicchetti (1994) criteria, suggesting that, when used in this sample, the psychometric properties of individual ASD instruments were quite variable with some correlation coefficients suggesting relatively inconsistent, inaccurate measures of ASD characteristics. None of the obtained correlations exceeded .90 to be deemed “excellent” (Cicchetti). A similar pattern of much lower ADI-R restricted repetitive behaviors published internal consistency was seen in the obtained internal consistency correlation compared to that for social reciprocity and communication; however, the obtained correlation was much lower than the published correlation (.69 versus .20). Moreover, the proportion of error variance that accounted for scores on the ASD instruments used in this sample was higher than in published psychometric studies.

Several factors likely contributed to an elevated error variance in this community-based sample including (a) a different purpose of the current evaluations (i.e., clinical versus research), (b) differing characteristics of the sample, (c) differing training practices for staff of the published studies and the current study, and (d) differing characteristics of the respondents. The first two factors are somewhat related so let us first consider the purpose of the evaluation compared to previous published psychometric research studies and the resulting makeup of the sample that may differ from published studies.

The purpose of a psychometric study is to establish the validity of an instrument for a particular purpose (Anastasi & Urbina, 1997) with the process of validation involving sequential stages of test construction to develop, revise and statistically test the appropriateness of items against external criteria (Anastasi & Urbina, 1997; American
Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). A common study analysis for psychometric validation is examination of the association between a test score and a criterion measure (e.g., the ADOS score for children diagnosed with AD and for children not diagnosed with AD). Thus, a single test is isolated from other tests and assessed for its strengths and limitations.

Alternatively, a comprehensive psychological assessment is undertaken to gain a “clear picture of the client’s mental health problems in clinical evaluations” (Butcher, 2006) or to understand the behavior of the individual child under evaluation. A single test used for this purpose is considered within the context of other scales and other information (Meyer et al., 2001). Ideographic assessment can be used to (a) describe current functioning, (b) identify therapeutic needs, (c) aid in differential diagnosis of cognitive, emotional, and behavioral disorders, and (d) monitor treatment over time (Meyer et al., 2001). The differing purposes of the psychometric studies and the current study result in different types of data in which “the validity coefficients from testing research may underestimate the validity of test findings when they are integrated into a systematic and individualized psychological assessment” (Meyer et al., 2001, p. 152). For example, the second case described above included measures (e.g., ADI-R, GADS) that suggested AS. However, his scores on the direct observation-based ADOS-G as well as informal observations during the WASI, PPVT-III, and EVT suggested alternative explanations for his delayed social skills, and verbal and motor rituals. He was diagnosed with Anxiety Disorder-Not Otherwise Specified within the context of a MTMM assessment. This same child included in a psychometric validation study of the ADI-R
would likely be excluded because he would not meet the overall ASD diagnostic criteria (e.g., expert consensus, previous diagnosis). Thus, differing purposes of a study can result in different participants being included or excluded from the sample. In addition, all of the current participants were children clinically referred to a specialty ASD clinic and were more likely to have complex or atypical presentations (i.e., primary medical and school providers felt uncertain of their results and sought additional assistance) and co-morbid language, learning, and psychiatric disorders (e.g., ADHD, anxiety, MR) (Hartley & Sikora, 2009; Sciutto & Cantwell, 2005) than those children who might be included in a psychometric study. Participants with test scores that are inconsistent with the child’s final diagnosis (e.g., scoring high on the ADI-R or ADOS with an anxiety disorder final diagnosis) might be excluded from a psychometric validation study (i.e., target sample has ASD and no co-morbid diagnosis or prior diagnosis). However, in the current study, this participant was retained and his scores likely introduced error variance for the ADI-R and GARS at a minimum.

If validation studies were expected to more closely approximate “real world” cases, highly individualized and contextually-specific tests would require a group of individuals with similar histories, behaviors and experiences, test scores, ages, sex, and various other factors. Validation studies of this type would be incredibly complicated, if not impossible (Meyer et al., 2001). Thus, the differing purposes of published psychometric studies and obtained psychological assessment inherently result in different data with different error variances.

The third potential factor that may account for the increased error variance is naturally occurring differences in the levels of training among those administering the
ADOS-G or ADI-R compared to initial psychometric studies. According to the ADOS developers (Lord et al., 2000), differing levels of training are required for those using the ADOS in research (i.e., research training) and those using the ADOS for clinical purposes (i.e., clinical training). Graduate students in the community-based study relied on the clinical training of one graduate student disseminating the training to other students at the WMU Center for Autism. This likely produced several minor differences in how scoring might occur between a priori research studies and clinical evaluations contributing to increased error variance among these student clinicians (Williams, Atkins, & Soles, 2009) as opposed to those administering the ADOS for the psychometric validation study. This particular source of increased variance was expected and was one of the primary motivating factors for conducting the current investigation.

As a fourth factor, the respondent (i.e., parents, guardians, teachers) characteristics in our sample may have also contributed to increased error variance (Meyer et al., 2001), including the potential for different motivations for the evaluation (e.g., personal concern, stress and fatigue, secondary gain associated with resources) and background of the sample (e.g., communication skills, lack of exposure to information about ASDs). As mentioned previously, many of the children referred to our service were referred from a widespread rural area by providers who wanted additional diagnostic resources or parents who were concerned about their child’s development or prior evaluations. Parents, guardians and teachers may have a preconceived idea of the “problem” of the individual child (e.g., “child has behavior problems,” “child is autistic”) that could lead them to respond to questionnaires (e.g., GARS, GADS) based on these notions by over or underreporting. Thus, in ongoing clinical evaluations the clinician
might weight direct observation more highly if there is reason to suspect a motive of secondary gain, and if multiple observation opportunities are available in different settings. The respondent’s ability to communicate (i.e., education level, culture) may also affect the accuracy of information on respondent report measures such as the ADI-R and the VABS-II. The current sample included many families that might or might not be recruited into or excluded from university-based research studies because of the location, reduced fees of the clinical, and lower reading abilities and educational backgrounds.

**Similarities with Other Published Psychometric Studies**

Three findings from the current study replicated findings from other published psychometric studies in three areas: (a) inter-rater reliabilities on subdomains of the GARS, (b) GARS underestimating the probability of an ASD and, (c) the ADOS-G’s ability to discriminate between those on the ASD spectrum and those not on the spectrum. In the current study, inter-rater reliability on the GARS subscale stereotyped behavior and the overall AQ reached a similar level of reliability (p < .01) as the initial psychometric study published in the GARS Examiner’s Manual. The index of inter-rater reliability on the GARS subscales communication and social interaction did not achieve reliability at the published level but was statistically significant at the p < .05 level. This suggests that in the current study, mothers and the next closest relative were relatively consistent in their ratings of the child such that the consistency approached that found in the GARS psychometric validation study (Gilliam, 1995).

The current study also replicated findings in the South et al., (2002) and Mazefsky and Oswald (2006) clinically-based samples. In these studies, the GARS underreported the probability of the child being on the ASD spectrum. In the South study, the mean
GARS AQ was 90.1 for those children with AD, “average” probability of having autism. In the Mazefsky and Oswald study, the mean GARS AQ was 87.87 for those with a final diagnosis of AD, rated as “below average” probability of autism. Consistent with these studies, the current study resulted in a mean GARS AQ for those with a final diagnosis of either AS or PDD-NOS of 85.7, rated as “below average” probability of ASD. For those with a final diagnosis of PDD-NOS, the mean GARS AQ was 88.7, and 84.7 for those with a final diagnosis of AS. Both of these means are designated as “below average” probability of autism. It is not clear why the current findings show that those with a final diagnosis of PDD-NOS had a slightly higher mean GARS AQ than those with AS, as one would expect children with PDD-NOS to have a lower GARS AQ than those with AS. Children diagnosed with PDD-NOS generally have fewer symptoms and/or subclinical levels of ASD symptoms. Nevertheless, these findings add to the extant literature showing that the GARS underreports ASD characteristics and underestimates the probability of being diagnosed with autism (Lecavalier, 2005; Mazefsky & Oswald, 2006; South et al., 2002) compared to other measures.

Although no single instrument is recommended for use in an autism spectrum assessment, it is likely that because of the ease of use and scoring of the GARS (Williams, Atkins, & Soles, 2009), it is used as a screening tool or to augment established screening tools such as the M-CHAT, particularly in general medical settings (South et al., 2002). The underestimates of ASD characteristics are likely to produce a high false negative rate during screening so the use of the GARS should be reconsidered. The GARS-II (Gilliam, 2006) purports new norms and more readable behavioral descriptions for respondents; however, publications addressing the under-identification of children
with autism using the GARS-II have not yet been reported in peer-reviewed journals.

Although no prior studies have reported any similar underestimation with the GADS, our data suggest that this possibility may be worth further investigation. In our current community-based sample, the mean GADS score for those with a final diagnosis of AS was 91.9, "likely" probable of AS. Based on the standardized mean of 100, one would expect the mean scores for those with a final diagnosis of ASD on the GARS or the GADS to be closer to 100. This finding adds to the existing literature on the GADS, in that this measure may be underreporting symptoms and the probability of being on the ASD similar to its sister measure, the GARS.

Lastly, similar to the published psychometric study, ADOS-G scores in the current study differentiated those children with an ASD from those with a non-ASD final diagnosis. The psychometric validation study showed that the social and communication domains of the ADOS-G differentiate well between those on the spectrum and those not on the spectrum; however, differences in algorithm scores for the restricted and repetitive behaviors domain were not significantly different for any of the four modules (Lord et al., 2000). The current study findings are similar, in that scores on the ADOS for social, communication, and social and communication were effective in differentiating between those children with a final diagnosis of ASD and those with a final non-ASD diagnosis. The subdomain of imagination/creativity differs by ADOS-G module, and is not incorporated into the diagnostic algorithm (Lord et al., 2000). Although the inferential statistic F was not as high as that reported in the published validation study (Lord et al., 2000), the findings in the current study were significant at the p < .05 and p < .01 levels. According to Cohen (1992), the effect sizes for these one-way analyses of variance were
medium effect sizes. This means that in the current study, approximately 12.9% of the variance (Eta squared) in the diagnosis of ASD/non-ASD is explained by the ADOS social subdomain when controlling for other predictor variables (Cohen, 1992). Similarly, approximately 10.9% and 6.8% of the variance in the diagnosis of ASD/non-ASD is explained by the ADOS communication and social and communication, respectively.

**Similar Findings for Descriptive Characteristics of ASD**

Two general findings in the descriptive literature on ASD were replicated in the current study’s community sample: (a) a strong correlation between the PPVT-III and VIQ, and (b) no significant differences in receptive and expressive lexical language for children with a final ASD diagnosis (Hundry et al., 2010; Kjelgaard & Tager-Flusberg, 2001). In the current study, the PPVT-III-VIQ correlation was “good” (r = .86) compared to “excellent” (r > .90) in the published literature (Dunn, 1997). However, the current correlation is still good, suggesting that the receptive language measure (PPVT-III) correlated relatively well with the Weschler scale measurement of language (i.e., verbal concept formation, verbal reasoning, and knowledge acquired from one’s environment).

Thus, for those 15 children who took both the PPVT-III and the Weschler measure of verbal aptitude (i.e., participants with more advanced language and academic development) in the current sample, the scores were relatively consistent. Only 15 (19.5%) of children in the current study were given the PPVT-III and the Weschler scales. Children with more substantial cognitive and language impairments did not complete these tests due to the inappropriateness of these tests for their functioning levels. Nevertheless, a significant correlation with a relatively small sample size suggests not only that the tools measure similar constructs, but also that they were consistently administered.
In terms of receptive and expressive language in the current study, no significant differences between those with a final diagnosis of ASD and those without an ASD (excluding MR) were found. Similarly, no differences were found in the degree of impairment in language (i.e., absolute difference between receptive and expressive language) between those with a final diagnosis of ASD (excluding MR) and those with a non-ASD or no diagnosis. These results are consistent with those in the published literature on language in children with ASDs (Hundry et al., 2010; Kjelgaard & Tager-Flusberg, 2001), suggesting expressive and receptive vocabulary do not differentiate children on the spectrum from those not on the autism spectrum. Although many children in the current study as well as those in the published literature have discrepancies between their expressive and receptive language, these discrepancies did not differentiate children on the spectrum from those not on the autism spectrum in the current study.

Isolated characteristics of children such as adaptive level (Matson, Mayville, et al., 2003), intelligence quotient (Loveland & Tunali-Kotoski, 2005) language (Tager-Flusberg, Paul, & Lord, 2005) and adult ratings of children on the GARS (Lecavalier, 2003; South et al., 2001; Mazefsky & Oswald, 2006) do not clearly indicate whether a child is on the autism spectrum or not. Similarly, although there are certain common presentations that often occur (e.g., advanced verbal skills in children with AD), isolated characteristics of functioning do not indicate which subcategory of ASD a child may fit (Klin, Pauls, Schultz, & Volkmar, 2005). The instruments specifically developed to assess ASD characteristics, the ADOS and the ADI-R, alone or in combination are relatively successful in identifying those children with autism spectrum features in the three core domains (Noterdaeme, Mikdenberger, Sitter, & Amorosa, 2002; Risi et al.,
However, within the published psychometric studies (e.g., Risi et al., 2006), the positive predictive value of a single instrument correlating to the final diagnosis is low since a single instrument is never used alone in the diagnosis of autism spectrum disorders or any other mental health diagnoses (Meyer, 2002).

Several authors of the psychometric studies discussed previously (e.g., Lord et al., 2000) explained the importance of combining information from (a) the examiner’s methods observation, (b) parent reports, and (c) the examiner’s clinical judgment based on knowledge about the specific problems and populations when integrating a meaningful and accurate assessment (Anastasi & Urbina, 1997; Ozonoff et al., 2005).

Meyer et al. (2001) explains that as researchers understand the limitations of any single method of measurement or definition of a construct and the enhancement of validity when variables are measured by multiple methods, the applied clinical practitioner understands that distinct assessment methods contribute unique information, and the quality of idiographic assessment can be enhanced by integrating the data from multiple methods of assessment. Thus, the safeguard against using instruments demonstrating poorer psychometric properties when used in a clinical setting similar to the current study lies in the use of the multi-trait, multi-method (MTMM) approach to diagnostics (Anastasi & Urbina, 1997; Campbell & Fiske, 1959). Anastasi and Urbina (1997) further explain, “Fundamentally, all activities connected with a psychological assessment . . . involve professional judgment based on knowledge about the specific problems and populations at hand” (p. 511).
SUMMARY OF CONTRIBUTIONS AND LIMITATIONS

The present study provides a unique contribution to the literature by examining findings from a specialty community clinic-based sample of children evaluated for autism spectrum disorders. This study is one of only a few studies (e.g., Mazefsky & Oswald, 2006; South et al., 2002) presenting ASD instrument psychometric properties in specialty clinic settings. The current study is unique in that the specialty clinic was also a training and service center rather than primarily a university research center (e.g., University of Michigan Autism and Communication Center, Yale Child Study Center). Students in training are less experienced, and therefore may be less likely to administer instruments correctly; however, the training process typically programs for substantially closer oversight than non-training clinics. Additionally, students in training may be particularly conscientious to perform in accordance with testing protocols due to the relative newness of the diagnostic evaluation process, limited prior familiarity with the particular instrument (e.g., ADOS, ADI-R, Weschler scale), and knowledge that they are being closely observed and evaluated as part of training. The obtained psychometric properties with student examiners were generally poorer than those in original published studies and it is unclear whether similar results would be obtained in a community-based sample that did not serve as a training clinic. Nevertheless, it is important to recognize that evaluations included in this data set were conducted by a unique sample of examiners.

Despite these contributions to the literature, some important limitations to the present study must be noted and some caution should be used in interpreting and extrapolating these findings. First, the cases assessed in this study represented families
who willingly sought services for their children. That is, these children and their parents were seeking explanations to their children’s challenges and behaviors which can cause parents to overreport or underreport on interviews and questionnaires (Meyer et al., 2001). Second, although the clinical services were not for the purposes of a specific study, a more direct comparison of reliability between the published studies and the current study would have used multiple raters for the ADI-R and the ADOS (inter-rater reliability) to provide a more comprehensive picture of the psychometric properties of these scales in community use.

As noted above, several different students with varying levels of experience and training conducted the assessments. This unique group of examiners is a limitation of the study as it likely introduced error variance into the use of the functioning and ASD-specific instruments. In particular, students were not trained on the ADI-R or ADOS through either the clinical (with one exception) or research training offered by the instrument developers/manufacturer.

Another limitation of the study is that it did not employ a true experimental research design. This study was purely archival in nature which means that no particular variables were held constant (e.g., all children diagnosed with AD) or manipulated by the researcher (e.g., treatment intervention). The limitation of such a research design is that the data must generally be limited to primarily descriptive and correlational results with a more limited use of inferential statistical analyses and no opportunity to draw conclusions about causality.

**Future Research and Clinical Recommendations**

Two areas for future research include (a) defining the process, parameters, and
accuracy of clinical judgment in ASD assessments, and (b) limitations of the MTMM in ASD assessments. As Dixon et al. (2009) pointed out, clinical judgment as a core research question is largely absent from ASD research; rather, research focuses on the plethora of ASD (a) instruments, (b) differential diagnostics, (c) characteristics, and (d) functioning. More than ever before, families are seeking differential diagnostic assessments from ASD specialty clinics that, in the best of circumstances, employ clinicians with extensive ASD expertise and experience. However, it is unclear who actually qualifies as having “ASD expertise and experience.” Currently, ASD diagnostic assessments are conducted by (a) clinical psychologists, (b) school psychologists, (c) neurologists, (d) neuropsychologists, and (e) psychiatrists (Matson & Kozlowski, 2011). Many authors discuss the importance of ASD expertise and experience, as well as “clinical judgment” (Meyer, 2002; Meyer et al., 2001) in ASD differential diagnostics (Lord, 2010); however, little systematic research exists regarding clinical judgment in MTMM psychological assessments despite development of such research methodologies (Meyer et al., 2001). This issue may never be fully addressed, in part because the process of research informing clinical practice is commonplace, while clinic-based research questions informing research is far less common (Meyer et al., 2001).

Moreover, clinicians in training learn that clinical judgment leads to “good diagnostic practices” (Meyer et al., 2001). The limitations of such judgments are not defined in the empirical literature. Specifically, lapses in clinical judgment have been identified but it is unclear whether such lapses in reasoning can be corrected by using a multi-trait, multi-method assessment battery, as Meyer et al. (2001) suggest. Sciutto and Cantwell (2005) utilized analogue methodology (i.e., presentation of clinical vignettes) to
assess clinicians' differential diagnoses of Asperger's and high functioning autism. In these hypothetical cases clinicians consistently favored an Asperger's diagnosis even in the presence of an early language delay citing a child's desire to engage in social interactions and overall IQ as supportive of an Asperger's diagnosis. According to the current DSM-IV-TR criteria, a language delay precludes a diagnosis of Asperger's disorder suggesting that clinicians are making decisions that are not fully in line with the diagnostic system. However, the proposed DSM-V system will make no distinction between any of the different ASDs, making this a moot point if the new system is adopted as proposed. Nonetheless, questions regarding clinician decision-making in complex, interactive, and detailed diagnostic assessments require the use of analogue research methodologies as well as clinical observation methodologies to further define the amorphous "clinical judgment" in psychological assessment of children presenting at ASD specialty clinics. Specifically, future research in clinic or community-based settings using qualitative research methods should focus on the process of clinical judgment as briefly described in the case examples above, including the need for and use of additional instruments, diagnosis hypothesis-testing as well as accuracy based on panel consensus, vignettes or other methods to judge "accuracy" of a diagnosis.

Theoretically, the MTMM enhances convergent validity (Campbell & Fiske, 1959, Meyer et al., 2001) which guides the clinician in differential diagnosis (Meyer et al., 2001). Several ASD scientist practitioners have defined the instruments to be used in an ASD assessment (e.g., Klin, Saulnier, Tsatsanis, & Volkmar, 2005; Lord & Corsello, 2005; Ozonoff et al., 2005; Matson & Boisjoli, 2007; Risi et al., 2006), and provided empirical evidence to support their use (Ozonoff et al., 2005; Risi et al., 2006). However,
research in clinical process has yet to combine such psychometric expertise with that of statisticians and methodological theorists to further understand the limitations of the MTMM ASD assessment and more fully define the benefit in community or clinic-based settings. Specifically, a research agenda examining the number of methods and type of instruments optimally useful in clinic settings is warranted. For example, deleting questionnaire data from the current study may have had a negligible effect on final diagnosis. In addition, a research design which parallels the dismantling study design in which effective psychological interventions (e.g., cognitive behavior therapy [CBT]) are broken into their component parts (e.g., cognitive therapy and behavioral activation) might be applied to the core versus extended battery of ASD assessments (Ozonoff et al., 2005; Matson & Boisjoli, 2007; Risi et al., 2006). As the number of children diagnosed with ASDs increases and their access to quality services increases, the professional environment will be ripe to examine these research questions.
REFERENCES


Appendix A

Data Coding Sheet
<table>
<thead>
<tr>
<th>Abnormality Evident At or Before 36 Mo.</th>
<th>Restricted, repetitive bxs</th>
<th>Communication</th>
<th>Social reciprocity</th>
<th>Informants 1 = mother, 2 = father, 3 = grandmother, 4 = grandfather, 5 = sitter, 6 = other</th>
</tr>
</thead>
</table>

**Social reciprocity**

**Abnormality Evident At or Before 36 Mo.**

- 0 = None
- 1 = Autism
- 2 = Asperger
- 3 = PDDNS
- 4 = MR
- 5 = Non-ASD
- 6 = Referral for non-ASD, 7 = ADHD/ODD

**Presenting Dx**

- 0 = None
- 1 = Autism
- 2 = Asperger
- 3 = PDDNS
- 4 = MR
- 5 = Non-ASD
- 6 = Referral for non-ASD

**Number of siblings (per 5 ADL-R)**

- Father age
- Mother age
- Guardian
- Primary caregiver
- Relation to child

**Sex**

- Male
- Female

**Age (in total months)**

**Year of assessment**

**Age (in years)***

#ID
<table>
<thead>
<tr>
<th>Module</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>Play</td>
</tr>
<tr>
<td>Imagination/Creativity</td>
<td>Communication + Social</td>
</tr>
<tr>
<td>Restricted, repetitive bxs</td>
<td></td>
</tr>
</tbody>
</table>

**GARS (Informants 1 & 2)**

- **GARS Autism Quotient Score**
- **GARS %ile**
- **GARS Age Equiv**
- **GARS Social Interact Standard Score**
- **GARS Commun Standard Score**
- **GARS Stereotyped Bxs**

- **GARS Autism Quotient Score**
- **GARS %ile**
- **GARS Age Equiv**
- **GARS Social Interact Standard Score**
- **GARS Commun Standard Score**
- **GARS Stereotyped Bxs**

---

**Informants 1=mother, 2=father, 3=grandmother, 4=grandfather, 5=sitter, 6=teacher, 7=aunt, 8=other**

**GADS Asperger Quotient Score**
- **GADS %ile**
- **GADS Social Interaction**
- **GADS Restricted Bxs SS**
- **GADS Cognitive Patterns**
- **GADS Pragmatic Skills**

---

**Module**

- **Social**
- **Communication**
- **Play**
- **Imagination/Creativity**
- **Communication + Social**
- **Restricted, repetitive bxs**
<table>
<thead>
<tr>
<th>Language Measures</th>
<th>Language Measures</th>
<th>Language Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPVT-III Score</td>
<td>PPVT-III Score</td>
<td>PPVT-III Score</td>
</tr>
<tr>
<td>PPVT-III Raw</td>
<td>PPVT-III Raw</td>
<td>PPVT-III Raw</td>
</tr>
<tr>
<td>PPVT-III Age Equivalent: Round up, Round down</td>
<td>PPVT-III Age Equivalent: Round up, Round down</td>
<td>PPVT-III Age Equivalent: Round up, Round down</td>
</tr>
<tr>
<td>EVT Score</td>
<td>EVT Score</td>
<td>EVT Score</td>
</tr>
<tr>
<td>EVT Raw</td>
<td>EVT Raw</td>
<td>EVT Raw</td>
</tr>
<tr>
<td>EVT Age Equivalent: Round up, Round down</td>
<td>EVT Age Equivalent: Round up, Round down</td>
<td>EVT Age Equivalent: Round up, Round down</td>
</tr>
</tbody>
</table>

**Vineland (VABS-II)**

- Informants: 1 = mother, 2 = father, 3 = grandmother, 4 = grandfather, 5 = sister, 6 = other
- VABS Type: 1 = Survey Interview Form, 2 = Parent/Caregiver Rating Form
- Communication Domain Standard Score
- Daily Living Skills Domain Standard Score
- Socialization Domain Standard Score
- Adaptive Behavior Composite
- Percentile Rank

**WPPSI-III**

- Full Scale (FSIQ)
- Verbal IQ (VIQ)
- Performance IQ (PIQ)
- Processing Speed Quotient (PSQ)
- Composite Language Score
- Full Scale (FSIQ)
- Verbal Comprehension Index (VCI)
- Processing Speed Index (PSI)
- Working Memory Index (WMI)
- Full Scale (FSIQ)
- Verbal IQ (VIQ)
- Performance IQ (PIQ)
- Processing Speed Quotient (PSQ)
Appendix B

HSIRB Approval Letter
Date: September 10, 2008

To: Linda LeBlanc, Principal Investigator
    Sheryl Lozowski-Sullivan, Student Investigator for dissertation

From: Amy Naugle, Ph.D., Chair

Re: HSIRB Project Number: 08-09-14

This letter will serve as confirmation that your research project entitled “Diagnostic Assessment Measures of Children with Suspected Autism Spectrum Disorders: A Review of 80 Cases” has been approved under the exempt category of review by the Human Subjects Institutional Review Board. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the application.

Please note that you may only conduct this research exactly in the form it was approved. You must seek specific board approval for any changes in this project. You must also seek reapproval if the project extends beyond the termination date noted below. In addition if there are any unanticipated adverse reactions or unanticipated events associated with the conduct of this research, you should immediately suspend the project and contact the Chair of the HSIRB for consultation.

The Board wishes you success in the pursuit of your research goals.

Approval Termination: September 10, 2009