2012

Examination of the reliability and validity of a new observation measure for autism spectrum disorders: the autism spectrum disorder observation for children

Daniene Neal
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: http://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Psychology Commons

EXAMINATION OF THE RELIABILITY AND VALIDITY OF A NEW OBSERVATION
MEASURE FOR AUTISM SPECTRUM DISORDERS: THE AUTISM SPECTRUM
DISORDER OBSERVATION FOR CHILDREN

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctorate of Philosophy

in

The Department of Psychology

by
Daniene Neal
B.A., Nicholls State University, 2003
M.S., Mississippi State University, 2005
December 2012

TABLE OF CONTENTS

Abstract

With increasing attention being drawn to autism spectrum disorders (ASD), specifically with regard to early and accurate diagnosis and treatment, researchers and clinicians alike have placed emphasis on finding assessment tools that can aid in this goal. The purpose of this study is to examine the psychometric properties of a new observation measure for ASD, the Autism Spectrum Disorders-Observation for Children (ASD-OC). The ASD-OC was found to have good to excellent interrater reliability, and excellent internal consistency ($\alpha = .96$). As a result of these initial reliability analyses, nine items were removed from the scale. The resulting 45 item ASD-OC was found to converge strongly with the Childhood Autism Rating Scale (CARS) and had a moderate inverse relationship with the Daily Living Domain of the Vineland Adaptive Behavior Scales, Second Edition (VABS-II). Based on analyses of criterion validity, the ASD-OC was able to differentiate children with ASD from children without ASD. Discriminant analysis confirmed the ability of the 45 item measure to accurately predict diagnostic group membership. Finally, an exploration of the effect of the changing DSM criteria based on the ASD-OC was conducted resulting in support for concerns regarding children who may no longer meet criteria but continue to exhibit significant impairments. Implications of these findings and future directions are discussed.

Introduction

Autism spectrum disorders (ASD) have become a hot topic in both the popular media and the clinical and research communities over the last several years. As such, a greater demand for services and early identification has resulted. Research into the development of assessment tools that can aid in early, reliable diagnosis has become a priority in the field. The goal of the present study was to examine the psychometric properties of a new observation scale, the Autism Spectrum Disorder Observation for Children (ASD-OC). First, a brief history of ASD will be provided with definitions for specific ASD, followed by the assessment of ASD in general, and finally a review of existing observation measures.

History of ASD

Leo Kanner (1943) was the first to describe autism in his observations of 11 children whom he observed to have a number of similar atypical characteristics: social deficits, impairments in language, and insistence on sameness. These same symptoms, or triad of autism, are still considered to be the core features of the disorder today. However, there are features that Kanner described that are no longer considered accurate. Based on his observations, Kanner asserted that children with autism exhibited normal if not advanced intellectual functioning. Researchers now support the view that while there may be some cases of normal intellectual functioning in children with autism, termed high-functioning autism, as many as 75% of children with autism may also have an intellectual disability (ID; Joseph, Tager-Flusberg, & Lord, 2002). Kanner also noted that the children he observed all came from parents with above average intelligence. He hypothesized that these parents may have been too consumed by their own lifestyles and careers to provide the requisite nurturing and affection to their children, resulting in the disorder. Kanner acknowledged that symptoms of autism appeared to be present from

birth and that biological factors likely played a role in etiology. Nonetheless, he maintained that cold, emotionless parenting also played a key role. A similar theory would later be popularized by Bettleheim (i.e., refrigerator mothers; 1959).

This psychogenic approach placed a great degree of blame on caregivers and was not popular among parent groups. Also, there was a lack of research to support the theory. With the zeitgeist of the time changing from psychodynamic to behaviorism and biological psychiatry, researchers began to look toward biological, neurological, and genetic theories of etiology. Rimland (1964) published a book, *Infantile Autism: The Syndrome and Its Implications for a Neural Theory of Behavior,* arguing for a neurological basis for autism. He cited empirical research to support his claims and helped shape the direction of autism research. Folstein and Rutter (1977) were among the first researchers to examine the genetic etiology of autism. They studied pairs of monozygotic (MZ) and dizygotic (DZ) twins and found a 36% pair wise concordance rate for autism in the MZ twins and 0% for the DZ twins. These findings have been replicated and support a genetic contribution for autism, involving interactions among multiple genes (Bailey et al., 1995; LeCouteur et al., 1996; Ritvo et al., 1985).

Distinguishing autism from schizophrenia. Kanner (1943) was not the first to use the term autism, but he was the first to define the term as it is understood today. Bleuler (1950) was, in fact, the first person to coin the term autism, referring to the social withdrawal often seen in patients with schizophrenia. General consensus at Bleuler's time was that children observed with symptoms described by Kanner (1943), were exhibiting precursors to schizophrenia. In fact, in the first *Diagnostic and Statistical Manual of Mental Disorders* (DSM-I; American Psychiatric Association [APA], 1952) children with symptoms of autism instead would have been diagnosed as exhibiting Schizophrenic Reaction, Childhood Type. However, as time went

on researchers began to suggest that autism was a distinct disorder, leading to confusion among the interchangeable use of terms such as autistic and schizophrenic. In 1961, Creak provided nine criteria he argued depicted childhood manifestations of adult schizophrenia. The nine criteria included (a) emotional aloofness; (b) lack of awareness of personal identity; (c) preoccupation with objects and parts of objects ignoring function; (d) resistance to change/insistence on sameness; (e) abnormal sensory reactions; (f) anxiety prompted by environmental changes; (g) loss or lack of language or abnormal language use (e.g., echolalia or pronoun reversal); (h) distorted pattern of motility, including abnormal gait, unusual body posturing, rocking or spinning; and i) intellectual impairment, although some intellectual skills may be normal or exceptional. Interestingly, while Creak intended these criteria to clarify symptoms of schizophrenia, they helped shape the definition of autism as we currently use it today.

Contrary to Blueler (1950) and Creak's (1961) definition of autism, Kanner (1943) argued that autism was a distinct disorder from schizophrenia. He noted that the children he observed exhibited social deficits, termed "autistic aloneness" (p.242), from the very beginning rather than exhibiting a withdrawal later in life as observed in schizophrenia. Rutter (1968) also contrasted autism from schizophrenia, citing a higher male to female ratio for autism which did not exist for schizophrenia. In addition, age of onset tended to be much younger for children with autism versus childhood onset of schizophrenia. Children with autism also lacked delusions or hallucinations typical in schizophrenia. Finally, Rutter noted the chronic course of autism which remained stable for most of the lifetime versus the waxing and waning of symptoms often observed in schizophrenia.

Defining autism. Throughout history, many researchers developed their own definitions and criteria for autism. Following his initial observations in 1943, Kanner and a colleague, Eisenberg, outlined criteria for what they referred to as "early infantile autism" (1956). They identified two core symptoms of autism: self-isolation and insistence on sameness. Kanner attributed the communication deficits that he observed in 1943 to the lack of social interactions with others. As such, Eisenberg and Kanner (1956) did not consider communication deficits to be uniquely indicative of autism. In addition, Kanner noted that the deficits observed in autism appeared to be present early on, distinguishing autism from many other childhood disorders. Therefore, an age cut-off of 30 months was also included in Kanner and Eisenberg's criteria.

Following his research to distinguish autism from schizophrenia, Rutter (1978) also developed his own criteria for autism. Like Eisenberg and Kanner's (1956) criteria, Rutter cited impaired social relationships and insistence on sameness as core features. Unlike Kanner, however, Rutter included abnormal communication as a core feature. Rutter noted several similarities in language deficits exhibited by children with autism. Even in children where language developed, abnormal patterns of speech often emerged such as echolalia and pronoun reversal. He also observed that children with autism often used language as a means of having their wants and needs met, rather than as attempts at interaction. In addition to verbal communication, deficits in nonverbal communication (e.g., gesturing, eye contact) were also present. As such, Rutter (1978) believed these deficits were a key feature to the disorder of autism and warranted being part of the diagnostic criteria. Rutter agreed with Eisenberg and Kanner (1956) in setting an age cut-off and included the criteria of symptoms present by age 30 months in his definition of autism as well.

As the concept of autism became more widespread, many political and parenting groups became involved in advocating for services for families of children with autism. One such group was the National Society for Autistic Children (NSAC). In 1978, NSAC enlisted the help of Edward Ritvo, a psychiatrist, to develop their own set of criteria for autism. As with Rutter's (1978) definition, Ritvo listed social impairments, communication deficits, and 30 month age of onset. However, Ritvo (1978) also included additional criteria. He noted that children with autism often have abnormal reactions to sensory stimuli. While some children exhibited oversensitivity to sights and sounds, others appeared unaffected. Differing from Rutter's definition, Ritvo noted that children with autism often exhibited delays in meeting normal developmental milestones including motor and language skills. Despite the overlap between Rutter and Ritvo's criteria for autism, the two definitions were created for entirely different purposes. While Rutter's criteria was developed to clarify and drive future scientific research, Ritvo's criteria was developed to make autism more understandable to the general public and increase the services available to families at that time (Schopler, 1978).

Asperger's disorder. Hans Asperger, an Austrian pediatrician, published his own paper describing a group of children observed with deficits similar to those described by Kanner in 1943 (1944, as translated by Frith, 1991). Asperger also acknowledged the similarities between the symptoms exhibited by these children and those exhibited in persons with schizophrenia, terming the disorder *autistic psychopathy*. However, he noted some of the same differences between the two disorders as proposed by Kanner (1943) and Rutter (1968), particularly that symptoms were present since birth rather than developing later in childhood (Asperger, 1944, as translated by Frith, 1991).

Asperger described the children he observed as having poor interpersonal relationships and exhibiting a restricted set of specialized interests. He also noted that the children tended to evince high intelligence, with no delay in language and, in some cases, having advanced linguistic skills (Asperger, 1944, as translated by Frith, 1991). Asperger, however, did not formally test IQ in these children and Wing (1981) later argued that the appearance of high intelligence was likely a result of memorization skills without comprehension.

Asperger's work was originally published in German. As a result, almost four decades passed before his writing was translated into English. While it was translated in 1991 by Frith, Wing wrote about Asperger's work in 1981, referring to the cluster of symptoms as Asperger's syndrome. Wing (1981) did agree that Asperger's syndrome could be defined as having symptoms of autism with the intact language and less severe social impairments. However, Wing argued that language development was not necessarily normal in these children. The appearance of advanced linguistic skills was likely the product of rote memory which was simply repeated, often in the wrong context.

It is often argued that the primary difference between Asperger's disorder and autism is that cognitive and language deficits should not exist in Asperger's disorder (APA, 2000). However, Frith (2004) argues that there are several problems with using this distinction alone. Although language deficits may not exist in Asperger's disorder, it does not mean that the development of language skills is the same as that in typically developing children. Many parents of children with Asperger's disorder reported that their child exhibited language that was adult-like, containing vocabulary rarely used by children. In addition, the ability to exhibit normal expressive language does not imply normal receptive language skills. Finally, as this information is often gathered by retrospective reports, it can be difficult to get a reliable picture

of early cognitive and language development. The issues just discussed make it difficult to decide on a differential diagnosis based on language development alone, prompting the elimination of a separate diagnosis of Asperger's disorder from the DSM-V (discussed below; APA, 2010).

   Pervasive developmental disorder, not otherwise specified (PDD-NOS). As early as Kanner's initial report (1943) there were cases of children exhibiting autistic-like social impairments with little or no difficulties with language and repetitive behaviors (Bender, 1946; Despert & Sherwin, 1958). The diagnostic category of PDD-NOS was first added to the DSM in 1980 (APA). The diagnosis of PDD-NOS is often better defined by what it is not rather than what it is (Matson, Nebel-Schwalm, & Matson, 2007). Many regard the category as a diagnosis for children who exhibit some symptoms of autism, but not enough to meet criteria for autism or Asperger's disorder (Heflin & Alaimo, 2006). Others view it as a diagnosis for children with a milder version of autistic symptoms. The vagueness of the definition for PDD-NOS has been under scrutiny for some time and researchers have attempted to study whether PDD-NOS is truly a separate disorder or a different form of autism (Matson & Boisjoli, 2007). However, these efforts are fairly new and the ability of assessment tools to differentiate among various ASD is not well established. Unfortunately, that leaves PDD-NOS as the most frequently diagnosed ASD with the least of amount of research.

   Rett's disorder. Rett's disorder was first described by Andreas Rett, an Austrian pediatrician, in 1966 (as cited in Hagberg, 1993), referring to a group of 31 females exhibiting cognitive impairment, deficits in communication and socialization, and abnormal motor movements and gait. Nearly 20 years later, Hagberg, a Swedish researcher, published a description of his observations of females exhibiting similar behaviors and developed a set of

diagnostic criteria (Hagberg, Aicardi, Dias, & Ramos, 1983; Hagberg, Goutières, Hanefeld, Rett, & Wilson, 1985). Rett's disorder is the only ASD in which a specific genetic etiology has been found, a mutation of the $MECP_2$ gene (Amir et al, 1999). Symptoms of Rett's disorder differ from other ASD in that the disorder is primarily found in females, is characterized by slowed head growth, and includes a loss of previously acquired hand skills (APA, 2000). Given these distinctions, the APA has proposed to remove Rett's disorder from the DSM-V (2010)

Childhood disintegrative disorder (CDD). Heller (1908), a Viennese special educator, was the first to describe CDD, referring to a group of six children exhibiting a regression of skills between the ages of 3-4 years (as cited in Volkmar, 1997). The regression observed in CDD is quite broad, affecting communication, self-care skills, and social skills. While a specific neurobiological etiology of CDD has not been identified, CDD is more readily differentiated from the other ASD based on age of onset, course, and prognosis (Irwin, MacSween, & Kerns, in press). CDD is characterized by a period of normal development through at least age 2 years, with a mean onset of symptoms at age 3.36 years without later recovery of skills (Volkmar, 1992). The prevalence of CDD is much rarer than autistic disorder, Asperger's disorder, and PDD-NOS, with estimates of less than 2/100,000 (Fombonne, 2002).

Evolution of ASD Diagnosis in the DSM

DSM-III. Autism was first acknowledged as a distinct disorder in the DSM-III, labeled "infantile autism" (APA, 1980). The DSM-III differed from the previous versions of the DSM in several ways. While the DSM I and II were strongly influenced by the psychodynamic perspective popular at the time, the DSM-III marked a shift to biological and medical perspectives (Mayes & Horowitz, 2005). As such, the DSM-III made use of specific symptom criteria for diagnosis. The DSM-III also introduced the use of the five axis system still currently

in use which encouraged a biopsychosocial model for mental illness. Axis I is where the majority of clinical disorders were coded with the exception of intellectual disability, personality disorders, and pervasive developmental disorders (PDD), which were coded on Axis II. Axis III was for medical conditions, Axis IV was for environmental and social issues, and Axis V was for the Global Assessment of Functioning rating indicating a person's current level of functioning from 0-100.

A new diagnostic category, PDD, was added to the DSM-III and included two disorders: infantile autism and PDD-NOS (APA, 1980). In order to warrant a diagnosis of infantile autism according to the DSM-III, the child must have: a) an onset before 30 months of age; b) pervasive lack of responsiveness to other people; c) gross deficits in language development; d) peculiar speech patterns such as immediate and delayed echolalia, metaphorical language, pronoun reversal; e) bizarre responses to various aspects of the environment (e.g., resistance to change, peculiar interest in, or attachments to, animate or inanimate objects); and f) absence of delusions, hallucinations, loosening of associations, and incoherence as in Schizophrenia.

The criteria for PDD-NOS was more general than that of infantile autism: a) gross and sustained impairment in social relationships; b) three features from a list including excessive anxiety, constricted or inappropriate affect, resistance to change, oddities of motor movement, abnormalities of speech, hypo- or hypersensitivities, or self-mutilation; and c) an onset of the full syndrome after 30 months of age and before 12 years of age with the absence of delusions, hallucinations, incoherence, or marked loosening of associations (APA, 1980).

DSM-III-R. In 1987, a revised version of the DSM-III was published, the DSM-III-R (APA). The diagnostic term for autism changed from "infantile autism" to "autistic disorder." The broad category of PDD was still included in the DSM-III-R, with autistic disorder being the

primary diagnosis in the group. The diagnostic criteria were revised to state that individuals had to exhibit at least two specific behaviors from each of three core symptoms: (a) qualitative impairments in reciprocal social interaction; (b) qualitative impairments in verbal and nonverbal communication and in imaginative activity; and (c) markedly restricted repertoire of activities and interests. In addition, an age of onset was specified as symptoms appearing in infancy or early childhood. As in the DSM-III (APA, 1980), PDD-NOS remained under the category of PDD and was meant for children exhibiting social impairments similar to those of children with autism, but did not meet the criteria for either deficits in communication or restricted interests (APA, 1987).

DSM-IV and IV-TR. By the time the DSM-IV (APA, 1994) was released, the idea of a spectrum of autistic disorders had become popular, referring to this group of disorders with common symptoms that varied in severity (Constantino, Przybeck & Friesen, 2000; Posserud, Lundervold & Gillberg, 2006). The DSM-IV marked a change from coding PDD on Axis II to Axis I. Another major change included the addition of more disorders to the PDD category: Asperger's disorder, Rett's disorder, and CDD. The DSM-IV-TR was published in 2000 (APA) making few changes to diagnoses or criteria, but instead providing updates to the coding system.

In order to warrant a diagnosis of Asperger's disorder, the presence of at least two of the following four criteria is required: (a) impairment in the use of multiple nonverbal behaviors; (b) failure to develop social relationships appropriate for one's developmental level; (c) a lack of spontaneous seeking to share enjoyment, interests, or achievements with other people; and (d) a lack of social or emotional reciprocity (APA, 2000).

Rett's disorder is defined as the following impairments subsequent to a period of normal development: (a) deceleration of head growth between 5 and 48 months; (b) loss of purposeful

hand skills between 5 and 30 months, followed by the development of stereotyped hand movements; (c) loss of social engagement early in the course of the disorder; (d) poorly coordinated gait or trunk movements; and (e) severe impairment in the development of expressive and receptive language and severe psychomotor retardation (APA, 2000).

In order to meet the DSM-IV-TR criteria for CDD, a period of normal development until at least age 2 years must be followed by a significant loss of skills by no later than age 10 years in at least two of the following areas: (a) expressive and receptive language, (b) social skills or adaptive behavior, (c) bladder or bowel control, (d) play skills, and (e) motor skills (APA, 2000). In addition to regression, impairments in at least two areas including social interactions, communication, and restricted interests/repetitive behaviors must be present.

DSM-V. The newest revision of the DSM is set for publication in 2013 (APA, 2010). Overall, major changes in the diagnostic system are proposed for the DSM-V. The multiaxial system in use since the DSM-III (APA, 1980) is under consideration for changes. Proposed changes include collapsing axes I, II, and III, providing a coding system for axis IV, and revising the rating system used for axis V. An emphasis on dimensional assessment and diagnosis versus the traditional classification methods is another proposed change for the DSM-V.

Major changes to the diagnosis of ASD have been proposed as well (APA, 2010). To begin, the label for the diagnostic category will be changed from PDD to ASD, to match the terminology often used in literature. However, the DSM-V will eliminate the separate diagnoses within ASD (i.e., autism, Asperger's disorder, PDD-NOS, CDD, and Rett's disorder). In addition, Rett's disorder will no longer be included as an ASD. Despite having similar symptoms to the other ASD, Rett's disorder has frequently been regarded as a separate disorder given a better understanding of its etiology and the rareness with which it occurs (Ozonoff,

Goodlin-Jones & Solomon, 2007; Rutter, 2005). APA rationalizes that while research has supported the distinction between ASD and other disorders, there is less certainty regarding differentiating among the ASDs (2010). Therefore, the new diagnosis will only identify children as having an ASD.

The specific criteria for ASD has also been changed from previous versions of the DSM (APA, 2010). The criteria for communication and social skill deficits were merged rather than being separate core symptom areas. According to the APA (2010), this change was rationalized through research suggesting that communication impairments observed in autism may be related to the lack of socialization and vice versa. In addition, there is a wide range of severity regarding communication impairments observed in children with ASD, with some children exhibiting little to no impairment in this domain.

The second criteria for a diagnosis of ASD, according to the DSM-V (APA, 2010), is the presence of restricted interested and/or repetitive behaviors. In addition to criteria used in previous versions of the DSM (e.g., stereotypies, insistence on sameness), sensory peculiarities were added to the restrictive interests repetitive behaviors criteria. This came following research that many children with autism exhibit hypo- or hyper sensitivity to sensory stimuli (e.g., loud noises, bright lights) and that this particular symptom may be observed in younger children before other symptoms of restricted interests/repetitive behaviors emerge. The age of onset criteria was clarified with a statement indicating that while symptoms should be present in infancy or early childhood, for some children, symptoms may not be apparent until a later age when more advanced social demands are placed on the child (e.g., preschool).

Finally, specifiers for level of severity have been added (APA, 2010). The DSM-V has three severity specifiers: Level 1 (requiring support), Level 2 (requiring substantial support), and

Level 3 (requiring very substantial support). These specifiers are based on the impairment

caused by both the socialization/communication and repetitive behaviors/restricted interests

domains.

The changes proposed by the DSM-V (APA, 2010) have been met with criticism.

Szatmari (2000) noted that specifying severity would pose a problem given the lack of a single

identifier of autism severity. That is, would a child who is nonverbal and therefore does not

exhibit problems such as echolalia be considered more or less severe than a child who is verbal

with speech abnormalities such as echolalia? In addition, symptoms of ASD have been shown to

subside in middle childhood and increase again in adolescence, which may affect the severity

indicator given at a particular time. Wing, Gould, and Gillberg (2011) also expressed concerns

regarding the proposed DSM-V changes. Specifically, the combining of the social and

communication criteria and lack of descriptive criteria have made it so that only very

experienced clinicians can use the criteria appropriately. In addition, removal of the labels

Asperger's disorder and PDD-NOS have important clinical implications with regard to social

stigma and availability of services, particularly for children already given these diagnoses. With

the release of the new DSM-V will come much research into how these changes effect

assessment, diagnosis, and treatment availability.

Prevalence

Popular media today has made an issue over increases in the prevalence of ASD observed

over the last several years (Rutter, 2005). While many parents and advocacy groups argue that

autism is becoming an epidemic with accusations of vaccines, diet, and other environmental

hazards as the culprits, researchers have found little to no support for these factors. Rather,

results from research suggest that increases in prevalence are more likely due to changes in the

diagnostic criteria over time and increased awareness and availability of services for the disorder. Before reviewing the research available in this debate, current estimates of prevalence of ASD are summarized below.

Fombonne (2005) conducted a review of epidemiological studies of autism spectrum disorders. Regarding autism, there was a statistically significant difference between prevalence rates published in 1966-1993 and 1994-2004. The median prevalence in the early studies was 4.7/10,000 whereas the median prevalence in the more recent studies was 12.7/10,000, indicating an increase in prevalence of autism over time. Based on the results of studies published since 1987, Fombonne developed a conservative estimate of autism prevalence at 13/10,000. Regarding PDD-NOS, prevalence was estimated at 20.8/10,000. It is not surprising that this particular diagnostic group had a higher prevalence than autism given that it is a broad category including individuals not meeting full criteria for autism. Studies of prevalence for Asperger's disorder are less common given the debate over whether it is a separate disorder from autism. However, based on the studies that did investigate Asperger's disorder and estimates of the ratio of individuals with autism versus Asperger's disorder, Fombonne calculated a prevalence rate of 2.6/10,000 for Asperger's disorder. Grouping these three most common ASD together, Fombonne calculated the current prevalence of ASD to be 36.4/10,000, an increase from 27.5/10,000 in 2003 (Fombonne) and 18.7/10,000 in 1999 (Fombonne).

As evident from epidemiological studies described above, there has been a steady increase in the prevalence of ASD over the last 15-20 years. While some argue that environmental variants are playing a causal role in this increase (e.g., diet, vaccines), there is much more evidence in support of factors including changing diagnostic criteria, increased public awareness and availability of services, varied assessment tools, and methodological

14

differences among epidemiological studies (Fombonne, 2005; Nicholas et al., 2008; Rutter, 2005). While the vaccine debate has garnered much attention in the media over the last 10-15 years, largely fueled by now debunked information provided by Wakefield in 1998, multiple studies have failed to demonstrate a causal relationship between the measles, mumps, and rubella vaccine and ASD (Chen and DeStefano, 1998; Dales, Hammer, & Smith, 2001; Peltola et al., 1998; Taylor et al., 1999). Similarly, studies regarding the effects of diet and supplements in the etiology and treatment of ASD have failed to provide consistent or promising results (Buie et al., 2010; Millward, Ferriter, Calver, & Connell-Jones, 2008). As such, researchers have turned to simpler explanations for increases in the prevalence of ASD.

As reviewed above, the diagnostic criteria for ASD has been revised and broadened through each iteration of the DSM. In 2002, Wing and Potter examined the effects of the changing diagnostic criteria by reviewing epidemiological studies conducted using DSM-III-R versus DSM-IV or International Classification of Diseases, 10th revision criteria (ICD-10; World Health Organization [WHO], 1992). Comparison of the mean prevalence of ASD based on each set of criteria indicated a significantly higher prevalence when using the broader DSM-IV or ICD-10 criteria rather than the previous DSM-III-R criteria. Beyond the diagnostic criteria included in the DSM, the general concept of ASD has broadened over the years to include a spectrum of disorders including PDD-NOS and Asperger's disorder, as well as labels such as high functioning autism to describe children with average and above average intelligence with autism (Rutter, 2005). Thus, children who decades ago would not have been classified as ASD now fall into this broad category.

There is no denying that ASD has taken the spotlight in the popular media and, as a result, public awareness has greatly increased (Rutter, 2005). Parents have become aware of the

early signs and symptoms and are seeking assessment and diagnosis at younger ages (Matson, Rieske, & Tureck, 2011). With increased demand, has come an increased availability of services including state-wide early intervention programs intended to screen large numbers of at-risk children. A recent push for early identification and early intervention for children with ASD has also led to assessment and diagnosis in children as young as 18 months, with more reliable diagnosis at 2 years (Matson, Wilkins, & Gonzalez, 2008). Therefore, it is likely that this decreasing age of diagnosis has contributed to the increased prevalence observed in epidemiological studies.

Differences in the methodologies used in epidemiological studies may also contribute to the perceived increase in the prevalence of ASD. Differences in the diagnostic criteria used, specific ASD included in the study, and assessment tools and methods have all varied among studies (Fombonne, 2005). As time has passed, there have been multiple changes in the assessment tools available and diagnostic practices for ASD. Therefore, the populations sampled in each study may vary significantly, especially over time. This makes comparisons of studies from 1980 to studies from 2000 very difficult in terms of identifying a true cause for the increased prevalence. Thus far, researchers tend to agree that the available evidence does not suggest an increase in the incidence of autism, but rather that the increase in prevalence is a product of the variables discussed above (Fombonne, 2005).

Assessment of ASD

The assessment of ASD currently continues to focus on the core symptom areas identified by Kanner (1943): socialization, communication, and repetitive behaviors/restricted interests. However, a comprehensive assessment will also consider additional features such as feeding problems, sensory issues, and comorbid psychopathology (Matson et al., 2011). In addition, a clinician experienced in assessment and diagnosis of ASD will be familiar with these symptom deficits and how to directly observe and identify them throughout the lifespan (Volkmar, Charwarska, & Klin, 2005).

Core Features of ASD

Socialization. Early developmental milestones include looking at the face of a caregiver, eye contact, smiling, and babbling by age 3 months (CDC, 2010). By the end of age 7 months, most children engage in and enjoy social play, respond to their name, and imitate sounds. By age 1 year, most children begin to show a preference for specific people and toys, use sounds and gestures to communicate, and are able to say some single words (e.g., dada). Many children with ASD fail to meet these milestones. In addition, children with ASD may show little interest in the activities of others, isolate themselves, and choose to interact only when a want or need must be met (Matson, Stabinski-Compton & Sevin, 1991). Although Kanner (1943) noted that children with autism failed to form adequate attachments with their parents, it is now accepted that children with ASD may demonstrate some typical attachment behaviors (e.g., becoming upset when caregiver is not around, clinging to caregiver in a new situation). Other skills, such as joint attention and sharing interest tend to develop around age 6-9 months in typically developing infants (CDC). However, impairments in eye gaze and gestures typically prevent children with ASD from developing these skills (Osterling & Dawson, 1994).

The severity of social skill impairments differs within the various ASD. Njardvik, Matson, and Cherry (1999) studied the differences in social skills betweem participants with autism, PDD-NOS, and ID only. Social skills deficits were most severe in people with autism, followed by those with PDD-NOS, and then ID. These results are consistent with current literature, characterizing ASD as a disorder of social skills distinct from just ID, with more severe deficits in social skills in people with autism and less severe deficits in individuals with PDD-NOS.

Communication. Impairments in communication are another core symptom of ASD, although the type of impairment varies by specific ASD. Even with acquired language, many children with ASD exhibit abnormalities in their speech (Lord, et al., 2000; Rutter & Bartak, 1971). Abnormalities such as an unusual rhythm (e.g., "sing-song" speech), stress, intonation, or volume, echolalia (i.e., repeating words and phrases previously heard) and pronoun reversal (e.g., stating "you" when meaning "I") may be present and significantly impair an individual's ability to communicate. Therefore, communication impairments in ASD can appear as delayed or nonexistent language development, or as an abnormal use of language.

The overlap between social and communication skills and the impairments caused by deficits in these areas cannot be ignored and is a reason that the DSM-V (APA, 2010) proposes to combine these two symptom areas. Deficits in non-verbal forms of communication such as eye contact, social smiling, and gestures (e.g., nodding) interfere with an individual's ability to successfully form and maintain social interactions (Sasso, Garrison-Harrell, & Rogers, 1994). A lack of understanding of social cues interferes with the ability to understand non-verbal communication from others such as facial expressions and body language (Volkmar, Carter,

Grossman, & Klin, 1997). Therefore, even with verbal ability, many children with ASD experience deficits in communication and socialization.

Restricted/repetitive behavior. Restricted/repetitive behaviors may include an insistence on sameness, lining up of objects such as toys, circumscribed interests, and stereotypies such as hand flapping, body rocking, spinning, or staring at moving objects or lights (Lewis & Boucher, 1988; Turner, 1999). The DSM-V (APA, 2010) has also proposed to add sensory impairments to the category of restricted/repetitive behaviors. Sensory impairments may include an over or under sensitivity to lights, sounds, or touch, and/or an excessive desire for sensory stimulation (Miller, Anzalone, Lane, Cermak, & Osten, 2007). Abnormal restricted/repetitive behaviors can be hard to identify, especially at young ages, as many typically developing children exhibit similar behaviors (Bodfish, Symons, Parker, & Lewis, 2000). For example, Berkson, Tupa, & Sherman (2001) found that 90% of typically developing children exhibit body rocking. Another issue with identifying abnormal repetitive behaviors is that many of these behaviors do not appear until after age 3 years (Lord, 1995; Moore & Goodson, 2003) and tend to continue increasing through age 4 years (Cox et al., 1999; Moore & Goodson, 2003). It is when these behaviors continue beyond a certain developmental period that they become more apparent and are considered abnormal (Baumeister & Forehand, 1973; Berkson, 1983).

Components of a Diagnostic Battery

In general, best practice for a comprehensive assessment is to use multiple sources of information (i.e., interview, observation, rating scales) and involve multiple informants whenever possible (e.g., both parents, grandparents, alternate caregivers, teachers; Haynes & O'Brien, 2000). Specific to ASD, there are a number of assessment tools available as screeners, structured interviews, rating scales, and structured observations. However, this makes it difficult

when deciding which measures to use as the psychometric properties of many of these measures are still being investigated. As a result, there is no "gold standard" when it comes to assessment of ASD (Kleinman et al., 2008). There are however, some recommendations for the types of assessments that should be included in a diagnostic evaluation for ASD.

Zwaigenbaum and colleagues (2009) suggest the use of (a) standardized observation which provides a set of tasks to observe behavior at least within the core symptoms areas of ASD and a scoring system; (b) an interview with the parent(s) or primary caregiver regarding developmental history; (c) the core symptoms of ASD (i.e., socialization, communication, repetitive/restricted interests or behaviors) as well as any problems with feeding, sensory issues, or challenging behaviors; (d) a standardized assessment of cognitive, language, and adaptive skills; and (e) clinical judgment of a person experienced with assessment and diagnosis of ASD. Tidmarsh and Volkmar (2003) provide a similar list with the addition of a physical exam, given the high comorbidity of ASD with other medical conditions (e.g., seizures, gastrointestinal problems) and the potential confounding of an ASD diagnosis by a medical condition (e.g., hearing loss interfering with communication skills). Matson et al. (2011) also recommend that when selecting assessment tools, that clinicians use measures that go beyond the core symptoms of ASD to assess for challenging behaviors and additional psychopathology that is highly comorbid with ASD.

Without a standard battery for the assessment of ASD, many clinicians vary in the specific instruments they use and the comprehensiveness of their assessment. With so many methods of assessment, the question of diagnostic stability is raised. Kleinman et al. (2008) examined the diagnostic stability of ASD in 77 children initially assessed between the ages of 1-3 years and assessed again between ages of 3-7 years. The assessment tools utilized were the

Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994), *Autism Diagnostic Observation Schedule* (ADOS; Lord et al., 1989), *Childhood Autism Rating Scale* (CARS; Schopler, Reichler, & Rocher-Renner, 1988), and clinical judgment with the DSM-IV criteria. They found that the diagnostic stability of clinical judgment and DSM-IV criteria alone was 80%. Using the ADI-R, diagnostic stability was at 67%. The ADOS had a higher diagnostic stability of 83%. Finally, the CARS had a diagnostic stability of 76%. Kleinman et al. reported that the general trend over time was toward milder symptoms (i.e., moving from a diagnosis of autism to PDD-NOS) with the exception of a few cases where younger children only exhibited signs of impairment in social and communication skills at the initial assessment and then began displaying restricted/repetitive interests or behaviors by the second assessment. Based on a review of studies of diagnostic stability, diagnoses relying solely on the ADI-R, a structured interview, had a significantly greater number of children moving from no diagnosis to an ASD diagnosis at a later assessment. As a result, Kleinman et al., recommend supplementing the ADI-R with direct observation and clinical judgment for a more reliable diagnosis.

Observation Measures

In addition to being an important component of a comprehensive assessment, observation measures are also key in early screenings to identify at risk children (Kleinman et al., 2008). Recently, there has been a push for widespread screening by pediatricians and other healthcare professionals of infants to identify those at risk for ASD (Matson et al., 2011). However, a lack of brief, reliable observation measures have impeded this movement. What follows is a description of the standardized observation measures for ASD available today and a description of the psychometric properties, strengths, and weaknesses.

Behavior Rating Instrument for Autistic and Atypical Children, Second Edition (BRIAAC; Ruttenberg, Wolf-Schein, & Wenar, 1991)

The BRIAAC was originally published in 1977 by Ruttenberg, Kalish, Wenar, and Wolf as a means of assessing behaviors consistent with atypical development, including autism, and to serve as a tool for measuring developmental changes in behaviors over time. Originally, the BRIAAC contained eight scales including relationship to adults, communication, drive for mastery, vocalization and expressive speech, sound and speech reception, social responsiveness, body movement, and psychobiological development (Ruttenberg, Dratman, Fraknoi, & Wenar, 1966). The second edition of the BRIAAC (Ruttenberg et al., 1991) eliminated the body movement scale and added two supplemental scales for nonverbal children: expressive gesture and sign language, and receptive gesture and sign language. Observations are conducted in a natural setting (e.g., child's home) in the presence of a parent or caregiver. The observer watches the child for a minimum of 2 hours in order to complete the measure. Each scale of the BRIAAC is based on severity, duration, and frequency of the behavior and provides 10 levels ranging from severely disturbed to typical behavior for a child ages 3-4 years. Scores from each

of the scales are summed to get an overall weighted severity score from 10 to 100. In addition, the second version of the BRIAAC includes a behavior checklist which is rated as 0 (*not observed/no opportunity*), 1 (*occasional/intermittent*), or 2 (*characteristic/frequent*; Ruttenburg et al., 1991).

Development. The behaviors measured by the BRIAAC are based on observations of children diagnosed with autism who were enrolled in psychoanalytic treatment programs (Ruttenberg et al., 1977). Diagnostic criteria for children at that time were largely based on Kanner's (1943) description of autism. The revision of the BRIAAC was completed in an effort to update the scales and provide a behavioral checklist in the event that a full observation using the weighted scores could not be completed (Ruttenberg et al., 1991).

Psychometric properties. The psychometric studies reported for the BRIAAC to date have all involved the original version of the BRIAAC (Ruttenberg et al., 1977; 1991). Regarding interrater reliability, average Spearman rank correlation for the eight original scales was .88, with a range of .84-.93 (Wenar & Ruttenberg, 1976). Test-retest has demonstrated an average correlation of .81 across the eight original scales, with a range of .61-.91. Regarding validity, concurrent validity was tested by comparing independent diagnosis and ratings on the BRIAAC (Wenar & Ruttenburg, 1976). Pearson correlation between the BRIAAC total score and clinical judgment was significant (r=.64) along with three of the scales: relationship to adult (r=.43), vocalization and expressive speech (r=.64), and sound and speech reception (r=.65). Discriminate validity was investigated by Cohen and colleagues (1978), indicating that the BRIAAC total score was unable to distinguish between children with autism and low functioning children with other developmental disabilities (e.g., aphasia, mental retardation).

Limitations. The BRIACC has been shown to be able to discriminate between typically developing children and children with autism; however, it is not able to differentiate between children with autism and children with other developmental disabilities (Cohen et al., 1978). In order to administer the BRIAAC, one must undergo training with a certified BRIAAC trainer for approximately 40-80 hours (Ruttenberg et al., 1991). In addition, the 2-hour observation period in a natural setting adds significant time to the assessment process.

Behavior Observation Scale for Autism **(BOS; Freeman, Ritvo, & Schroth, 1984)**

The original BOS was comprised of 67 items describing behaviors associated with autism (e.g., looking at hands, covering ears, eye contact, echolalia). However, following reliability studies (Freeman & Schroth, 1984), several items were eliminated resulting in a modified version of the BOS containing 24 behaviors. The 24 items comprise four groups: solitary, relationship to objects, relationship to people, and language (Freeman et al., 1984). To administer the BOS, the child is directed to a room with age-appropriate toys and allowed to engage in free play. The BOS is comprised of nine 3-minute sessions, with two baseline periods at the beginning and end of the session and one period where the examiner actively tries to engage the child in interactive play. The observer watches for the presence of the 24 target behaviors and rates them as 0 (*did not occur at all*), 1 (*occurred once*), 2 (*occurred twice*), or 3 (*occurred regularly*).

Development. The target behaviors for the BOS were identified based on Ritvo's (1978) definition of autism, literature review, and clinical judgment (Freeman et al., 1984). The authors stated that the goals of the BOS were to differentiate children with autism from those without, as well as from children with ID and other developmental disabilities. In addition, the authors

24

hoped that the BOS could be used to differentiate among types of ASD as well as track severity and symptom course over the lifespan (Freeman et al., 1984).

Psychometric properties. For the original 67 item BOS, interrater reliability was greater than .84 for 55 of the items (Freeman, Ritvo, Guthrie, Schroth, & Ball, 1978). Regarding discriminant validity, items on the BOS were compared between children with autism and children with ID (Freeman et al., 1979). Eleven of the 67 items on the BOS were significantly different between the two groups. Freeman et al. (1984) investigated the psychometric properties of the revised 24 item version of the BOS. Interrater reliability was grater than .70 for 16 of the 24 items. Discriminant validity indicated that four of the behaviors were significantly different between the autism group and the control group: purposeful use of objects, non-purposeful use of objects, talks to examiner, and repetitive vocalizations.

Limitations. Up to 2 months of training on the use of the BOS is needed in order to reliably use the observation measure, including memorization of the coding system (Freeman et al., 1984). The BOS also lacks a diagnostic cutoff score. In addition, it has only been shown to discriminate between autism and ID (Freeman et al., 1984), but not between other forms of ASD or other developmental disabilities, making it less useful for differential diagnosis. Finally, the BOS is lacking in studies of psychometric properties (i.e., test-retest, internal consistency) and is in need of updated research.

CARS and CARS-2 (Schopler et al., 1988; Schopler, Van Bourgondien, Wellman, & Love, 2010)

The CARS is a clinician completed observation measure comprised of 15 independently rated subscales: Relating to people; Imitation; Emotional response; Body use; Object use; Adaptation to change; Visual response; Listening response; Taste, smell, and touch response and

use; Fear or nervousness; Verbal communication; Nonverbal communication; Activity level; Level and consistency of intellectual response; and General impressions. The clinician assigns a rating to each scale based on observations of the child supplemented with parent report and record reviews. The ratings range from 1 (*within normal limits for child's age*) to 4 (*severely abnormal*), with midpoints available (e.g., 1.5). The ratings for each scale are then summed and yield a total score which indicate non-autistic range (below 30), mild to moderate autistic range (score between 30 and 36.5), and moderate to severe autistic range (score between 37 and 60).

A second version of the CARS, the CARS-2, was recently published (Schopler et al., 2010). This revised version includes separate scales for high-functioning versus low-functioning individuals. The standard version of the form is the same as the original CARS and is indicated for use with children under age 6 years and/or those with below average cognitive abilities. The high-functioning version of the form is a new addition for individuals age 6 years and older and/or those with IQ scores above 80. There is also the addition of a parental questionnaire which provides background information that may useful to the clinician in making ratings. Both versions of the form contain 15 item areas.

Development. To develop the CARS, Schopler et al. (1988) used Kanner (1943) and Creak's (1964) definition of the disorder, in addition to observed autism symptoms. The CARS was originally used in North Carolina's statewide autism program, Treatment and Education of Autistic and related Communication handiCapped cHildren (TEACCH; Schopler et al., 1988) to differentiate between ASD and other developmental disabilities. As published studies began to demonstrate the utility of the measure, use of the CARS became widespread and it is currently one of the most commonly used autism diagnostic instruments (Morgan, 1988; Sturmey & Sevin, 1994; Teal & Wiebe, 1986). For the CARS-2, items on the high-functioning form were modified

based on current research on symptoms of high functioning autism and Asperger's disorder (Schopler et al., 2010).

Psychometric properties. The original CARS has been studied extensively, showing strong psychometric properties (Schopler et al., 1988). In the original study of the CARS, tests of internal consistency yielded a coefficient alpha of .94 and interrater reliability ranged from .55-.93, with an average of .71 (Schopler, Reichler, DeVellis, & Dalay, 1980). Criterion-related validity was assessed by comparing CARS diagnoses to those made based on clinical judgment. Agreement between diagnoses was high at r=.80. Schopler et al. (1988) assessed the validity of the CARS across settings and information sources. Validity was $r = .73$ for classroom observations, r=.82 for parent interview, and r=.82 for review of behavioral information from history. Volkmar et al. (1988) assessed the discriminant validity of the *CARS* and found that the measure was able to discriminate between those diagnosed with and without autism. Studies using the CARS-2 are needed to demonstrate the utility of separating out the forms based on level of functioning.

Limitations. The original CARS was designed prior to the currently accepted DSM-IV-TR criteria for autism and combines different definitions for the disorder. There is also no differentiation between autism and other ASD (e.g., PDD-NOS) on the CARS (Klinger & Renner, 2000). While the CARS-2 (Schopler et al., 2010) has added forms for different levels of functioning, it is unclear at this time how this will add to the utility of the measure.

ADOS (Lord et al., 1989)

The original version of the ADOS is a structured observation measure used to assess the social and communication behaviors in individuals ages 5 years through adulthood as part of a diagnostic autism assessment. The ADOS outlines a specific set of eight standardized

interactions for the examiner to complete as prompts for social and communicative behaviors. For most tasks, there are two sets of stimuli to choose from based on chronological age and developmental level. The eight tasks and corresponding behaviors of interest are a construction task (asking for help), unstructured presentation of toys (symbolic play, reciprocal play, and giving help), drawing game (turn taking), demonstration task (gestures and imitation), poster task (describing agents and actions), book task (storytelling), conversation (reciprocal communication), and socioemotional questions (language use). The examiner then rates each behavior in the context of the scenario as either 0 (*within normal limits*), 1 (*infrequent or possible abnormality*), or 2 (*definite abnormality*). In addition, there are some items with an option of rating the behavior as abnormal, but in a way not described by the other ratings. There are also overall ratings at the end of the assessment based on behavior throughout the observation period in the domains of reciprocal social interaction, communication/language, stereotyped/restricted behaviors, and mood and nonspecific abnormal behaviors. This scoring system allows for not only rating the presence or absence of specific behaviors, but also the quality of the behaviors. Overall, the measure takes approximately 20-30 minutes to administer.

Development. Development of the measure first involved identifying behaviors of interest that are commonly part of criteria for a diagnosis of autism via diagnostic systems (i.e., DSM and ICD), research, and clinical experience (Lord et. al., 1989). Next, a set of scenarios designed to elicit the behaviors of interest were developed. Finally, operational definitions of the behaviors of interest were created and a standardized system for rating the behaviors was created.

Psychometric properties. Both interrater and test-retest reliability for the ADOS is adequate (Lord et. al., 1989). Weighted kappas for interrater reliability ranged from .61-.92 for

the tasks and from .58-.87 for the general ratings. Weighted kappas for test-retest reliability ranged from .57-.84 for the tasks and from .58-.92 for the general ratings. Regarding discriminant validity, Lord et. al (1989) found that more than half of the rating items and all but three of the tasks were able to differentiate between autistic and non-autistic children at the $p < .01$ level.

Lord et. al. (1989) created an algorithm based on the ICD-10 criteria for autism using items from the ADOS measuring reciprocal social interaction, communication/language, and restricted/stereotyped behaviors. Discriminant validity for each domain, as well as for a total score, was examined. Significant differences were found between autistic and non-autistic groups for the individual domains as well as for the total algorithm score. However, the ADOS diagnostic classifications differed from clinical judgments for some of the children in the sample. Interrater and test-retest reliability for the domains and the total algorithm score was adequate, ranging from .70-.96 with restricted/stereotyped behaviors having the lowest reliability ranges.

Limitations. Administration and scoring of the ADOS requires that the examiner have previous experience with children with ASD in addition to substantial training and supervised practice with the measure (Lord et al., 1989). In addition, the psychometric properties and norms for the ADOS have only been developed for individuals with a mental age greater than 3 years and require that the individual have some language abilities. Finally, although an overall rating for repetitive behaviors and restricted interests is available at the end of the assessment, there are no specific prompts and ratings to address this core symptom of autism in the ADOS (Lord et al., 1989).

***Pre-Linguistic Autism Diagnostic Observation Schedule* (PL-ADOS) – DiLavore, Lord, &
Rutter, 1995)**

   The PL-ADOS was developed in order to address the limitation of the ADOS that

required children have a mental age of 3 years or greater and have some verbal ability (DiLavore

et al., 1995). Like the ADOS, the PL-ADOS is a semi-structured observation measure; however,

it provides more flexibility than the ADOS. For example, rather than remaining seated across

from the examiner, the prompts can take place as the child wanders around the room. In

addition, a parent can be involved in the prompts if the child does not respond to the examiner.

The PL-ADOS provides a specific set of 12 standardized interactions for the examiner to

complete as prompts for social and communicative behaviors (DiLavore, 1995). The 12 tasks

and corresponding behaviors of interest are a free play period (independent use of toys,

engagement with parents), imitation of child (repeats own action when imitated), mechanical

animal or car (responding to joint attention), bubble gun and balloon play (anticipates routine

with objects, initiates joint attention), social routines (anticipates a social routine, requesting),

simple actions with objects (functional/symbolic imitation), toy drum (turn taking), birthday

party (imitation during party), snack (request during snack, response to name, social smile),

dropping papers (response to another's distress), and adapted strange situation (separation from

mother, reunion with mother). As with the ADOS, the examiner rates each behavior in the

context of the scenario as either 0 (*no abnormality*), 1 (*neither clearly typical nor clearly

indicative of autism*), or 2 (*definite abnormality*). There are also overall ratings at the end of the

assessment based on behavior throughout the observation period in the domains of reciprocal

social interaction, communication, play, stereotyped/restricted behaviors, other abnormal

behaviors, and an overall autism clinical rating. Overall, the measure takes approximately 30 minutes to administer (DiLavore et al., 1995).

Development. In creating the PL-ADOS, the authors originally set out to adapt the ADOS scenarios for younger and non-verbal children (DiLavore et al., 1995). However, given the developmental differences between children ages 6 years and older and children under age 3 years, it became apparent that more flexibility would be needed. As a result, the activities were based on nonverbal symptoms of autism (e.g., eye gaze, joint attention, imitation, pretend play) identified through research and diagnostic systems criteria (i.e., DSM-IV and ICD-10) and were shortened.

Psychometric properties. Regarding reliability of the PL-ADOS, weighted kappas for interrater reliability ranged from .63-.95 for the activity codes, .60-.94 for the overall summary ratings, and .86 for the overall autism rating (DiLavore et al., 1995). Restricted/repetitive behaviors had the lowest interrater reliability, ranging from .18-.60. Regarding discriminant validity, 9 of 17 of the task ratings were found to significantly differ between children with autism and children without autism. In addition, 21 of 27 summary scores yielded significant differences between children with autism and children without autism.

As with the ADOS, the authors of the PL-ADOS developed an algorithm based on diagnostic systems criteria for autism (i.e., ICD-10 and DSM-IV) using only those items that significantly discriminated between groups (DiLavore et al., 1995). The resulting algorithm yielded two domains: social interaction/communication and restricted, repetitive behaviors. Cutoff scores for the algorithm were determined by using the scores that correctly differentiated the most children. The resulting cutoff criteria for a diagnosis of autism was a score of 12 or higher on the social/communication domain *and* a score of 2 or higher on the restricted/repetitive

31

behaviors domain. In the sample that these cutoffs were based on, this criteria correctly identified 40 of 42 children with developmental disability (i.e., non-autistic) and 16 of 18 nonverbal children with autism. However, all verbal children with autism were misdiagnosed, highlighting a weakness of the PL-ADOS in assessing children with autism with verbal abilities (DiLavore et al., 1995).

Limitations. As with the ADOS, the PL-ADOS requires substantial training and practice in order to reliably administer. In addition, while the PL-ADOS algorithm is able to differentiate nonverbal children with autism from other developmentally disabilities, it is not yet able to differentiate within verbal children (DiLavore et. al., 1995). This may limit its diagnostic usefulness in very young children with verbal abilities.

Autism Diagnostic Observation Schedule – Generic (**ADOS-G; Lord et al., 2000**)

The ADOS-G is the result of a combination of the original ADOS and the PL-ADOS. It was created in an attempt to address some of the weaknesses of its predecessors by including algorithms for a broader range of ages and for children with and without language abilities (Lord et al, 2000). In order to achieve this, the ADOS-G has four modules to choose from: Module 1 for little to no language abilities, Module 2 for short sentences and phrases, and Modules 3 and 4 for fluent speech. The examiner then chooses the appropriate version of the module based on the child's age. As with the previous versions of the ADOS (i.e., ADOS and PL-ADOS), the examiner engages in a structured set of activities and rates the individual's responses on a scale of 0-2. However, contrary to the previous versions, scoring for the ADOS-G takes place after the observation period is over. The ADOS-G provides cutoffs for both autism and an ASD classification (i.e., referring to possible diagnoses of autism, PDD-NOS, and Asperger's disorder) on the basis of severity (i.e., lower thresholds for ASD versus autism; Lord et al, 2000).

In either case, the individual must meet or exceed thresholds on social, communication, and social-communication domains. Overall, the measure takes approximately 30-45 minutes to administer.

Development. As research on the assessment of autism increased, it became apparent that identification of children under the age of 3 years was critical for early intervention and better prognosis (Lord et al., 2010). In response, the authors of the ADOS developed a measure that could assess young children through adulthood with various levels of language skills. The activities included in the PL-ADOS and the original ADOS were combined and adapted to create four modules. While modules for younger children allowed for more flexibility (i.e., movement around a room, play activities and toys), modules for adults involved more structured sessions (i.e., sitting at a table, age appropriate stimuli). The authors of the ADOS-G also changed the scoring so that ratings were based on the overall observations rather than specific responses to specific tasks. Finally, anywhere from one to three scoring algorithms were developed for each module to account for various combinations of age and language ability (Lord et al., 2000).

Additional revisions to the ADOS-G are currently underway as well (Lord, 2010). A toddler module for children between the ages of 12-30 months has been developed to improve sensitivity and specificity of the ADOS-G for this young age group (Luyster et al., 2009). Adaptations to Modules 1 and 2 are also being developed for adolescents and adults with little to no language skills (Lord, 2010).

Psychometric properties. Interrater reliability was examined for the ADOS-G items within the four modules and diagnostic classification (Lord et. al, 2000). Modules 1-4 had mean weighted kappas of .78, .70, .65, and .66, respectively. Using the diagnostic algorithms for each module, the ADOS-G correctly diagnosed 100% of the participants as either ASD or non-ASD

for Modules 1 and 3, 91% for Module 2, and 90% for Module 4. However, when considering diagnoses of autistic disorder, PDD-NOS, or non-autistic the percent agreements fell to 93%, 87%, 81%, and 84% for the four modules, respectively. Test-retest intraclass correlations ranged from .59-.82 for the domains (i.e., Social, Communication, Social communication, and Restricted/repetitive), with the Communication and Social domains having excellent stability and the Stereotyped Behaviors and Restrictive Interests having the lowest stability.

In regards to internal consistency across modules, Cronbach's alpha ranged from .86-.91 for the social domain, .74-.84 for the communication domain, .47-.65 for the Stereotyped Behaviors and Restricted Interests domain, and .91-.94 for the Social-Communication total (Lord et al., 2000). The ADOS-G was able to correctly classify 95% of the participants as having autism and 92% of the participants as not having autism. However, when PDD-NOS was considered, only 33% of the children in this diagnostic group were correctly classified by the ADOS-G. Therefore, while the ADOS-G is useful for distinguishing between ASD and non-ASD, it has a weakness in discriminating between autism and PDD-NOS.

Research has also looked at the validity of using the ADI-R and ADOS-G in combination for making an autism diagnosis (de Bildt et al., 2004). Agreement between the ADIR and ADOS was considered to be fair (de Bildt et al., 2004). That is, for the total sample, the percent agreement between the ADIR and ADOS-G was at 63.6%. Agreement was higher for children ages 5-8 years versus children above 8 years. de Bildt et al. (2004) also add that the diagnostic value of the two measures used together does not justify the extensive training and administration time needed for both the ADIR and the ADOS-G.

Limitations. According to the authors of the ADOS-G, in order to reliably use the ADOS-G in clinical practice, extensive training and practice including measures of interrater

reliability should be conducted (Lord et al., 2000). The ADOS-G also does not include the domain of restricted interests/repetitive behaviors in any of the diagnostic algorithms nor does it assess for age of onset. As a result, some children not meeting criteria in all three domains (i.e., PDD-NOS) may not meet criteria for ASD on the ADOS-G.

Screening Test for Autism in Two-year-olds **(STAT; Stone & Ousley, 1997)**

The STAT is an observational screening measure designed to be used with children between the ages of 24-35 months. It is comprised of 12 items including 2 play items, 4 imitation items, 4 directing attention items, and 2 requesting items. The session is set up as an interactive play period lasting approximately 20 minutes where the observer performs certain tasks to elicit a response from the child (e.g., observer blows up a balloon and then allows it to deflate, flying around the room). Domains consisting of 2 items (i.e., play and requesting items) are scored 0, .50, or 1.0 while domains consisting of 4 items (i.e., directing attention and imitation items) are scored 0, .25, .50, .75, or 1.0. Higher scores are indicative of greater levels of impairment. The scores for each domain are summed, with the exception of the two requesting items. A child with a total score of 2 or greater is considered in the at risk range for autism (Stone, Coonrod, Turner, & Pozdol, 2004).

Development. Items for the STAT were empirically derived using a database of 2-year olds containing scores on measures of play (Play Assessment Scale; Fewell, 1991), imitation (Motor Imitation Scale; Stone, Ousley, & Littleford, 1997), and communication (Prelinguistic Communication Assessment; Stone, Ousley, Yoder, Hogan, & Hepburn, 1997). Items from each measure were selected for inclusion in the STAT if they were able to significantly differentiate between autism and other developmental disabilities (Stone, Coonrod, & Ousley, 2000).

Psychometric properties. Interrater reliability for the STAT was at 1.00 using Cohen's kappa and test-retest reliability was $K = .90$ (Stone et al., 2004). Concurrent validity was examined by comparing the STAT and ADOS-G. Cohen's kappa agreement was .95 for diagnosis between the two measures. Using the cut-off score of 2, Stone et al. (2004) found that the STAT correctly identified 92% of the autism group and 85% of the developmental disabilities group.

Limitations. The psychometric studies for the STAT have included small sample sizes and replications with community based samples are needed (Stone et al., 2000; 2004). In addition, the STAT was designed to screen for autism alone. It does not assess for the risk of other ASD, such as PDD-NOS (Stone et al., 2004) Stone et al. (2004) also found that children with milder autism symptoms (e.g., PDD-NOS) were less likely to be identified as at risk by the STAT than children with more severe autism symptoms.

Autism Observation Scale for Infants **(AOSI; Bryson, Zwaigenbaum, McDermott, Rombough, & Brian, 2008)**

The AOSI was developed as a semi-structured observation for children between the ages of 6-18 months (Bryson et al., 2008). It is comprised of 18 items rated from 0 (*typical function*) to 3 (*deviant functioning*). A set of standard activities are provided to outline the interaction between the infant and the examiner. During these interactions, the examiner administers specific "presses" and observes the infants reaction to these presses. Presses are administered a set number of times; however, the order for the presses is flexible. Domains assessed include visual tracking, disengagement of attention, orientation to name, differential response to facial emotion, anticipatory social response, imitation, social babbling, eye contact, reciprocal social smile coordination of eye gaze and action, behavioral reactivity, cuddliness, soothability, social

interest and shared affect, transitions, motor control, atypical motor behavior, and atypical sensory behavior (Bryson et al., 2008). The total assessment time ranges from 15-20 minutes depending on the developmental level and cooperation of the infant.

Development. The AOSI was developed in order to address a need for observation measures appropriate for infants (Bryson et al., 2008). To design the scale, the authors identified target behaviors, developed activities during which the target behavior could be elicited, operationally defined the target behaviors and designed the rating system, and conducted pilot testing resulting in revision of the measure. The target behaviors were selected based on early signs of autism as provided by retrospective parental report, home videos, case studies of children with autism, and clinical experience. Pilot testing included testing younger siblings of children already diagnosed with ASD (Bryson et al., 2007; Zwaigenbaum et al., 2005).

Psychometric properties. Bryson et al. (2008) conducted a study to assess the interrater and test-retest reliability of the AOSI. Interrater reliability was examined for ages 6, 12, and 18 months with correlations for total score being .74, .93, and .94, respectively. Individual items' interrater reliability ranged from .33-1.0 agreement, with generally lower rates of agreement for younger ages. Items with particularly low levels of agreement included eye contact, behavioral reactivity, soothability, social interest and shared affect, and motor control. Intraclass correlations indicated fair test-retest reliability for total scores (.61) at 12 months of age (Bryson et al, 2008).

Limitations. The AOSI was originally developed as a research measure for detecting and monitoring early signs of autism (Bryson et al., 2008). As such, it has not been studied as a tool for clinical use in screening or diagnosing ASD. Studies regarding the validity of the AOSI have not yet been conducted. In addition, studies determining the specificity of the measure, that is, the ability of the AOSI to differentiate between symptoms of ASD versus other developmental disabilities in infants, are also non-existent.

37

Purpose

As described above, there are a number of observation measures used for assessment of autism in children. However, there are also several limitations of the existing measures. Some of the measures are lacking research beyond preliminary analyses (i.e., STAT, AOSI) and/or have outdated norms (i.e., BOS, BRIAAC). Even those measures considered to be widely used and more heavily researched (i.e., ADOS-G, CARS) have concerns for clinical use. To begin, extensive training and administration time makes the ADOS-G inconvenient and, at times, impractical to use. It also does not incorporate repetitive behaviors and restricted interests, one of the core symptoms of autism according to DSM-IV-TR criteria, into the scoring algorithm. While the CARS is a briefer measure, it is not based on current DSM-IV-TR criteria. Importantly, none of the observation measures described above have been able to reliably provide cutoff scores for differentiating among ASD (e.g., autism vs. PDD-NOS). As such, many children with PDD-NOS are not identified as being at risk on these observation measures despite possessing significant impairments warranting an ASD diagnosis. Taken together, the limitations of the current observation measures available for use in diagnostic assessments of ASD suggest that a brief, easy to administer observation measure that can identify children with ASD is warranted.

The ASD-OC has the potential to address many of the limitations of existing measures noted above. The ASD-OC is a brief clinician rated measure requiring no additional formal training other than knowledge of and experience with symptoms of ASD. The observation is semi-structured, with sample prompts provided for specific items allowing the clinician the freedom to tailor the session to the child, including length of the observation. The items of the ASD-OC were developed based on current diagnostic criteria for ASD in the literature and the

38

DSM-IV-TR. It includes items addressing all of the core symptoms of ASD: socialization, communication, and repetitive/restricted interests and activities, including sensory problems. The purpose of the current study is to assess the preliminary psychometric properties of a new observation measure, the ASD-OC. Future research with the measure is intended and can address factor structure, score cut-offs, differentiating within subtypes of ASD, and assessment with various age populations. Given the potential of the ASD-OC to become a useful component of a comprehensive assessment for ASD, the current study is a relevant first step in establishing the strength of this measure.

Method

Participants

Participants included 114 children ages 1–15 years ($M = 6.91$, $SD = 3.62$) assessed at a university outpatient clinic in Louisiana. Primary referrals included ASD, anxiety, learning disorders, and behavior problems. There were 82 boys (72%) and 29 (25%) girls in the sample (3 participants did not indicate gender). The ethnic breakdown was 68% Caucasian, 15% African American, 2% Hispanic, and 15% of other/unidentified ethnicity.

Based on diagnosis, assigned as described under Procedure, children were placed into one of 3 groups: ASD, typical development or atypical development. The ASD group consisted of children diagnosed with autism, PDD-NOS, or Asperger's disorder. The typical development group included children who did not meet criteria for any Axis I diagnosis and were not developmentally delayed. The atypical development group consisted of children who met criteria for one or more Axis I diagnoses (e.g., attention-deficit/hyperactivity disorder [ADHD], depression, generalized anxiety disorder, separation anxiety disorder, enuresis, obsessive compulsive disorder [OCD], selective mutism, nonverbal learning disorder, social phobia, and specific phobia) and/or had reported developmental delays or genetic conditions (e.g., Down's syndrome, Turner's syndrome) but did not meet criteria for ASD. Based on the above criteria, there were 43 children in the ASD group, 19 in the typical development group, and 52 in the atypical development group.

Measures

ASD-OC. The ASD-OC is a new observation measure for ASD comprised of 39 items. Scale items for the ASD-OC were generated through a series of steps suggested by Crocker and Algina (1986) and DeVellis (1991). The initial pool of items for the ASD-OC were developed

following a comprehensive research review of the ASD literature and current diagnostic guidelines (i.e., DSM-IV-IV and ICD-10), as well as items on other assessment measures of ASD including the ASD-DC, ADOS, and CARS, and observations noted by a clinical psychologist with more than 30 years of experience with this population. The drafted items were then reviewed by a child clinical psychologist who had extensive experience with this population for expert review. This expert review generated additional items, as well as suggestions for minor revisions to the original set of items. Subsequently, the assessment instrument was piloted by administering the items to several children referred for an ASD evaluation at an outpatient university clinic. Additional revisions were made to the items (i.e., deletion and clarification) as needed.

The ASD-OC is completed by the clinician following a play period with activities varying based on the age and developmental level of the child, cooperation of the child, and clinical judgment of the observer. A set of example probes for observing each item area is provided. Sample items include: looks when name is called, imitates pretend play, asks for help, imitates facial expressions, eye contact, peculiar body posture, and abnormal fascination with moving objects. Each item is rated as 0 (*no impairment*), 1 (*mild impairment*) or 2 (*severe impairment*). For the purposes of the current study, the ASD-OC was completed as part of a comprehensive assessment battery; however, it was not considered in the diagnostic formulation of the child. Assessments were administered by doctoral level graduate students trained on completing the ASD-OC prior to administration. As part of this training, new students were supervised by a senior student familiar with the ASD-OC to complete the observation measure.

CARS (Schopler et al., 1988). See Observation Measures section above for a complete description of the CARS including psychometric properties. The original CARS was used as

41

part the diagnostic battery for the current sample. While the CARS-2 was published prior to the completion of data collection for the current study, the original CARS was used to investigate convergent validity of the ASD-OC given the extensive support for its psychometric properties versus the CARS-2 which has not yet been independently studied to determine the effect that the revision had on reliability and validity.

Vineland Adaptive Behavior Scales, Second Edition **(VABS-II; Sparrow, Cicchetti, & Balla, 2005).** The VABS-II is a semi-structured interview designed to assess adaptive behavior in the domains of communication, socialization, daily living, and motor skills for ages birth through 90 years (Sparrow et al., 2005). Each domain consists of subdomains that are summed to compute an overall domain score. Items are ordered according to development and are rated as 0 (*Never*), 1 (*Sometimes*), or 2 (*Usually*). Each domain is then summed to calculate a total score of adaptive functioning, the Adaptive Behavior Composite (ABC). The VABS-II has demonstrated strong reliability and validity across all domains and for the ABC (Sparrow et al., 2005). Regarding interrater reliability, correlation averages ranged from .75-.85 for the domains and .67-.80 for the subdomains. Internal consistency coefficients for the domains across ages ranged in the low .80's to mid .90's, indicating good to excellent internal consistency. Regarding validity, the VABS-II has undergone a confirmatory factor analysis to support the theoretical domains proposed by the authors. In addition, several clinical groups have been tested using the VABS-II resulting in norms for populations including individuals with mental retardation, autism, ADHD, learning disability, emotional/behavioral disturbance, and visual and hearing impairments. For the purpose of the current study, only the Daily Living domain was used. The daily living domain is comprised of three subdomains: personal (e.g., hygiene, eating, dressing), domestic (e.g., cooking, cleaning), and community (e.g., money, safety skills, time).

DSM-IV-TR/ICD-10 Checklist. The DSM-IV-TR/ICD-10 checklist is a parent/caregiver report measure that consists of 19 items, assessing diagnostic criteria for ASD based on the DSM-IV-TR and ICD-10. Respondents mark a "yes" if the specific criteria is applicable to their child or "no" if it is not. The checklist includes symptoms from the three core areas: socialization, communication, and restricted, repetitive, and stereotyped patterns of behavior (see Appendix A). An additional item regarding whether delays or abnormalities in one of the three core areas were present before the age of 3 years is also included. The DSM-IV-TR/ICD-10 Checklist has been shown to have excellent interrater reliability ($r = .90$), test-retest reliability ($r = .97$), and internal consistency ($\alpha = .95$; Matson, González, Wilkins, & Rivet, 2008). For the purposes of the current study, the DSM-IV-TR/ICD-10 Checklist was used to assign children to groups for the comparison of ASD-OC total score to children meeting DSM-IV-TR versus DSM-V diagnostic criteria.

Procedure

All evaluations were conducted by doctoral level clinical psychology students supervised by a licensed clinical psychologist who confirmed the diagnoses through clinical consensus. Informed consent was obtained from the child's primary caregiver and assent was obtained from the child whenever possible. This study was approved by the Louisiana State University Institutional Review Board. For children referred for ASD or who screened at risk for ASD based on a brief rating measure (i.e., *Baby and Infant Screen for Children with aUtIsm Traits* [BISCUIT], ASD-DC), a clinical history and parent interview, observation, adaptive skills measure (i.e., VABS-II), broadband rating scale (i.e., *Behavior Assessment Scale for Children, Second Edition* [BASC-2], *Child Behavior Checklist* [CBCL]), and other standardized measures of ASD (i.e., ADI-R, CARS, CHAT) comprised the comprehensive assessment. Children

referred for problems other than ASD (e.g., anxiety, ADHD, learning disability) received a psychoeducational assessment which consisted of a structured diagnostic interview (i.e., *Anxiety Disorders Interview Schedule* [ADIS], adaptive skills measure (i.e., VABS-II), broadband parent rating scales [i.e., BASC-2, CBCL], measures of intelligence and achievement (*Wechsler Intelligence Scale for Children* [WISC-IV], *Wechsler Preschool and Primary Scale of Intelligence* [WPPSI], *Stanford Binet, Fifth Edition* [SB-V], *Woodcock Johnson, Third Edition* [WJ-III], *Wechsler Individual Achievement Test* [WIAT])and a school observation. It is important to note that all children received a screening to rule out possible ASD. Clinical consensus for diagnosis was reached by reviewing the referral reason, previous records, and assessment results with a licensed psychologist.

Data Analyses

Power

An *a priori* power analysis using GPOWER (Faul & Erdfelder, 1992) was conducted to determine the sample size required for the reliability and validity analyses. The sample size necessary to detect a large effect size of $r = 0.5$, with alpha set at a significance level of .05, and power set at .80 for a two-tailed correlation was calculated. Results of the GPOWER power analysis indicated that a total sample size of 29 participants was necessary.

Another *a priori* power analysis was conducted with GPOWER (Faul & Erdfelder, 1992) for an ANOVA for analyses involving diagnostic group and DSM version. The sample size necessary to detect a large effect size of $f = .40$, with alpha set at a significance level of .05 and power set .80 for an ANOVA with 3 groups was calculated. Results of the GPOWER analysis indicated that a total sample size of 66 participants was necessary. Finally, an *a priori* power analysis was conducted for an independent samples *t*-test comparing ASD-OC total score for 2 age groups. The sample size necessary to detect a large effect size of $d = .80$, with alpha set at a significance level of .05 and power set at .80 for an independent samples *t*-test was calculated. Results of the GPOWER analysis indicated that a total sample size of 52 participants was necessary.

Reliability

Interrater reliability. Interrater reliability was calculated for each item using intraclass correlations (ICC). This technique was chosen to address the continuous, ordinal values used on the ASD-OC (i.e., 0 (*no impairment*), 1 (*mild impairment*), and 2 (*severe impairment*; Cicchetti, 1994). Two clinicians, trained on how to complete the observation measure, each completed the ASD-OC independently for a subset of the sample following a simultaneous observation of a

child. Based on guidelines of clinical significance by Cicchetti and Sparrow (1981), item coefficients below .40 are poor, .40-.59 are fair, .60-.74 are good, and .75-1.00 are excellent. In order to ensure that items on the ASD-OC provided good to excellent reliability, any item with a coefficient less than .60 was considered for removal from the scale.

Item analysis and internal consistency. An item analysis was conducted by calculating the variance of each item and removing items with zero or near zero variance. Inter-item correlations were examined using a correlation matrix, noting any items with a weak correlation for possible removal from the scale. Next, item-scale correlations were run to examine the association of each item with the scale as a whole. Each item with a coefficient value less than .30 was considered for removal if data analysis indicated that the total scale alpha was increased by its deletion (Leech, Barrett, & Morgan, 2008). Cronbach's alpha was computed for the scale with the retained items to give an estimate of the internal consistency of the scale. Clark and Watson (1995) recommend using alpha levels of 0.8 for adequate internal consistency of a new scale.

Validity

Convergent validity. To determine convergent validity, Pearson product moment correlation coefficients were calculated between the total score of the ASD-OC and total score on the CARS. As both the ASD-OC and the CARS are diagnostic observation measures for ASD, it was hypothesized that the correlation would be a significant positive correlation indicating strong convergent validity.

Divergent validity. To determine divergent validity, Pearson product moment correlations coefficients were calculated between the total score of the ASD-OC and the VABS-II total score for the daily living domain. As the ASD-OC targets diagnostic symptoms of autism

46

in the areas of communication, socialization, and repetitive/restricted behaviors and the VABS-II daily living domain targets skills such as feeding, dressing, and bathing, a low or non-significant correlation, indicating little to no relationship between the measures, was expected.

Criterion validity. In order to investigate criterion validity, ASD-OC total scores were compared by diagnostic group: ASD, atypical development, and typical development using an analysis of variance (ANOVA). Individuals diagnosed with ASD and evincing more severe symptoms were assumed to score significantly higher (i.e., more impaired) on the ASD-OC than children with atypical development. The factor of age (i.e., less than 3 years vs. 3 years or greater) was also examined via an independent samples *t*-test. These age criteria were chosen as current DSM-IV-TR diagnostic criteria indicates that symptoms of autism should be present before age 36 to warrant a diagnosis (APA, 2000). As a result, many assessment tools for ASD have used this 36 month age cut-off. However, as early diagnosis and intervention have become more of a focus, children as young as 18 months have started receiving diagnoses of ASD (Matson, Wilkins, & Gonzalez, 2008). Therefore, an examination of score differences for this age cut-off of 36 months is relevant for a measure of ASD symptoms. It was hypothesized that younger children would have lower total scores on the ASD-OC as a result of the difficulty observing some of the symptoms in this young population and the development of additional symptoms as the child ages (Wing et al., 2011). Finally, an examination of the mean total scores on the ASD-OC by diagnostic group and age was conducted to detect trends in scoring patterns.

Discriminant analysis (DA). To determine if both the individual items and total scale of the ASD-OC significantly discriminated between groups, a DA was conducted. DA is useful for predicting group membership with a dichotomous categorical dependent variable and several continuous independent variables (Leech et al., 2008). This method also creates a model to

47

predict variables that discriminate best between groups. Diagnostic group was entered as the grouping variable (i.e., ASD vs. atypical development) and all retained items of the ASD-OC were entered as predictors. It was hypothesized that each retained item on the scale and the scale as a whole would significantly discriminate between children with ASD and children without.

DSM-IV vs. DSM-V. As described in the introduction, there are significant changes to the diagnostic criteria for ASD being proposed for the DSM-V. In order to examine how the ASD-OC would perform with this new diagnostic criteria, the ASD-OC total score was compared based on DSM-IV-TR and DSM-V criteria. Three groups were created based on DSM version: participants meeting criteria for ASD on both the DSM-IV-TR and DSM-V, participants meeting criteria on only the DSM-IV-TR, and participants not meeting criteria on either DSM version. An ANOVA was run with DSM version as the independent variable and total score on the ASD-OC as the dependent variable. It was hypothesized that given the more stringent criteria in the DSM-V, that children only meeting criteria based on the DSM-IV-TR would score significantly lower than children meeting DSM-V criteria. Both DSM groups were expected to score significantly higher than children not meeting criteria according to either version of the DSM.

Results

Interrater Reliability

Interrater reliability was calculated for a subset of the total sample (n=26). Participants

ranged in age from 1–14 years ($M = 5.25$, $SD = 3.46$). There were 23 boys (89%) and 2 (8%)

girls in the sample (1 participant did not indicate gender). Ethnic breakdown was 62%

Caucasian, 12% African American, and 26% of other/unidentified ethnicity. Breakdown

regarding diagnostic group was 39% ASD, 46% atypical, and 15% typical.

Total scale ICC with all items was .96. ICC for each of the 54 items on the ASD-OC was

also calculated. The mean inter-item ICC was .66 (range = -.19 – 1.00). Seven items had ICC

less than .60, with peculiar body posture, fascination with spinning objects, intonation, phrasing

to change meaning, and repeating words/phrases producing ICC less than .40. Two items (i.e.,

abnormal motor movement and pedantic speech) had ICC's between .40 and .60 and were

retained for further analysis. However, the weak correlations were noted and considered along

with additional analyses when deciding whether to retain these items in the final measure. The

original 54 items of the ASD-OC can be seen in Appendix B. ICC for each item and the total

scale are listed in Table 1.

Table 1
Interrater reliability coefficients using ICC (n = 26)

Item	ICC	Item	ICC	Item	ICC
1	.64**	22	.78**	38f	.85**
2	.85*	23	.79**	38g	--
3	.96**	24	.93**	38h	.83**
4	.77**	25	.92**	38i	--
5	.86**	26	.71**	38j	.48
6	.80**	27	.71**	38k	--
7	.84**	28	.33	38l	-.07
8	.84**	29	.64*	38m	.66*
9	.92**	30	.85**	38n	.88**

Table 1 cont.

10	.91**	**31**	.43	**38o**	.93**
11	.86**	**32**	.76**	**38p**	.83**
12	.76**	**33**	--	**39**	.72**
13	.80**	**34**	.29	**Total**	.96**
14	.83**	**35**	.65*		
15	.79**	**36**	.66*		
16	.95**	**37**	1.00**		
17	.81**	**38a**	.68*		
18	.85**	**38b**	.26		
19	.84**	**38c**	.72**		
20	.82**	**38d**	-.19		
21	.86**	**38e**	.71**		

Note. *p < .05; **p < .01

Item Analysis and Internal Consistency

Item analysis for the remaining items (i.e., after removal of items with poor interrater reliability as described above) was conducted for the total sample size of 114. Four additional items were removed due to having little to no variance (i.e., staring at hands, pronoun reversal, speaks overly precise, and repeatedly provides facts about a specific topic). The mean inter-item correlation was .33 (range = -.15 – .86). The mean item-scale correlation was .56 (range = .08 – .83). Five items had an item-scale correlation of less than .30 (respect for others' personal space; repetitive sniffing, touching, feeling, licking, mouthing, tapping of objects or surfaces; rate of speech; pedantic speech; responds with on topic comments). However, these items were not removed from the scale as the total scale alpha would not have increased with their removal. Cronbach's alpha for the remaining 45 items (i.e., following removal of poor interrater reliability and low variance items) was .96. The item-scale correlations can be found in Table 2.

Table 2
Item-scale correlations for the ASD-OC

Item	*r*	Item	*r*	Item	*r*
1	.75	**22**	.41	**38o**	.43
2	.57	**23**	.60	**38p**	.48

Table 2 cont.

3	.83	**24**	.68	**39**	.23
4	.79	**25**	.76		
5	.79	**26**	.24		
6	.78	**27**	.42		
7	.82	**29**	.55		
8	.76	**30**	.45		
9	.69	**31**	.48		
10	.71	**32**	.48		
11	.77	**35**	.42		
12	.79	**36**	.22		
13	.78	**37**	.38		
14	.71	**38a**	.32		
15	.62	**38c**	.31		
16	.81	**38e**	.24		
17	.79	**38f**	.35		
18	.68	**38h**	.35		
19	.64	**38j**	.08		
20	.77	**38m**	.43		
21	.74	**38n**	.50		

Convergent and Divergent Validity

For convergent validity, Pearson's product moment correlation was calculated for the total scores for a subset of the total sample (n = 76) on the ASD-OC and CARS. Participants ages ranged from 1–15 years ($M = 6.53$, $SD = 3.57$). There were 57 boys (75%) and 16 (21%) girls in the sample (3 participants did not indicate gender). The ethnic breakdown was 68% Caucasian, 12% African American, 1% Hispanic, and 19% of other/unidentified ethnicity. Breakdown regarding diagnostic group was 36% ASD, 51% atypical, and 13% typical. Analyses yielded a significant large correlation of $r = .83$.

Regarding divergent validity, Pearson correlations were again calculated for total ASD-OC score and the subdomain score on the VABS-II Daily Living Skills subscale for a subset of the total sample (n = 59). Participants ages ranged from 1–15 years ($M = 5.38$, $SD = 3.27$). There were 47 boys (80%) and 9 (15%) girls in the sample (3 participants did not indicate

gender). The ethnic breakdown was 64% Caucasian, 15% African American, 2% Hispanic, and 19% of other/unidentified ethnicity. Breakdown regarding diagnostic group was 54% ASD, 39% atypical, and 7% typical. Analyses yielded a significant moderate correlation of $r = -.43$.

Criterion Validity

Total ASD-OC scores were compared by diagnostic group (i.e., ASD, atypical development, and typical development) to determine if scores on the ASD-OC differed based on diagnosis. Based on the assumptions of an ANOVA, the sample size of each group could not be more than 1.5 times larger than the smallest group (Field, 2005). Therefore, 15 participants were randomly removed from the ASD group and 24 from the atypical group, resulting in a total of 75 participants for the analysis. Demographic characteristics by diagnostic group can be seen in Table 3. The groups were compared based on ethnicity, gender, and age to assess for group differences. Chi square analyses indicated that the groups were not significantly different in regards to ethnicity or gender. Results from an ANOVA revealed that the mean age of the groups were not significantly different from one another.

Table 3
Demographic information for comparing ASD-OC total score by diagnostic group

	Total Sample	ASD:	Atypical:	Typical:
	N = 75	n = 28	n = 28	n = 19
Age: Years				
Mean (SD)	6.80 (3.76)	5.11 (3.02)	8.41 (3.52)	7.71 (3.85)
Range	1-15	1-15	1-14	2-14
Gender				
Male	78%	85%	71%	74%
Female	22%	15%	29%	26%
Ethnicity				
Caucasian	69%	57%	68%	79%
African American	13%	21%	21%	5%
Hispanic	1%	4%	0%	0%
Other/Unknown	17%	18%	11%	16%

Next, an ANOVA was conducted with diagnostic group as the independent variable and total score on the ASD-OC as the dependent variable. There was a statistically significant difference between diagnostic groups on ASD-OC total score, $F(2,72) = 24.47$, $p < .001$, indicating that the diagnostic groups significantly differed from each other on clinician rated ASD symptoms. Tukey post hoc tests were used to investigate group differences. Children in the ASD group ($M = 33.68$) scored significantly higher (i.e., indicating more symptoms of ASD) than children in the atypical group ($M = 7.93$) and typical group ($M = 9.05$). However, no significant difference emerged between participants in the atypical development and typical groups.

It was hypothesized that children under 36 months of age would score significantly lower than older children on the ASD-OC. To test this hypothesis, the sample was divided into two groups: children under 36 months (n = 20) and children 36 months and above (n = 93). To keep sample sizes within 1.5 times each other (Field, 2005), 63 participants were randomly removed from the older age group, leaving a total of 50 participants for the analysis. Demographic characteristics by age group can be seen in Table 4. The groups were compared based on ethnicity and gender, and diagnostic group to assess for group differences.

Table 4
Demographic characteristics by age group

	Total Sample	< 36 mos:	>= 36mos:
	N = 50	n = 20	n = 30
Diagnostic Group:			
ASD	38%	55%	27%
Atypical	40%	25%	50%
Typical	22%	20%	23%
Gender			
Male	74%	90%	63%
Female	24%	5%	37%
Ethnicity			

Table 4 cont.

Caucasian	64%	50%	73%
African American	14%	10%	17%
Hispanic	0%	0%	0%
Other/Unknown	12%	40%	10%

Chi square analyses indicated that the groups were not significantly different in regards to ethnicity or diagnostic group. There was a significant difference on gender between the two age groups, $\chi^2 = 6.20$, $p < .05$. This gender difference is not surprising given a higher male to female ratio of ASD in general (Fombonne, 2005; Kanner, 1971). That said, researchers have demonstrated that the core symptoms of ASD do not significantly differ between males and females (Rivet, 2010). In addition, in further analyses gender was not significantly related to symptoms of autism for either age group. Therefore, it was not deemed necessary to control for gender in subsequent analyses.

Next, a *t*-test was conducted with age group as the independent variable and total score on the ASD-OC as the dependent variable. There was a statistically significant difference between age groups on ASD-OC total score, $t(48) = 4.94$, $p < .001$, indicating that the children under 36 months of age ($M = 34.25$) scored significantly higher than children ages 36 months and older ($M = 12.53$), regardless of diagnostic group. Given that this finding was contrary to what was hypothesized, group means were visually examined for the two age groups by diagnosis. Due to the small sample size within cells (e.g., only 4 children under age 3 in the control group), a factorial ANOVA was unable to be conducted to test for interactions with age and diagnosis. However, an inspection of group means suggested that all children under the age of 36 months regardless of diagnosis were rated high on the ASD-OC (see Table 5). This result was not the case for children ages 36 months and above. These children showed a clearer pattern of high scores for the ASD group and similarly low scores for the typical and atypical groups.

54

Given this lack of variability in ratings for children under 36 months of age, this age group was

excluded from further analyses.

Table 5.
ASD-OC means by age group and diagnosis

	< 36 mos:	>= 36mos:
	n = 20	n = 30
Diagnostic Group:		
ASD	(n = 11)	(n = 8)
	36.18	38.38
Atypical	(n = 5)	(n = 15)
	26.40	3.80
Typical	(n = 4)	(n = 7)
	38.75	1.71

DA

To determine which items of the ASD-OC discriminated between diagnostic groups best,

a DA was conducted. In order to have adequate sample size for a DA, the sample size of the

smallest group must exceed the number of predictors. In this case, 45 items of the ASD-OC

were entered as predictors. Therefore, only the ASD (n = 31) and atypical groups (n = 47) were

compared. Demographic characteristics of the groups used in the DA appear in Table 6.

Table 6
Demographic characteristics by diagnostic group for DA

	Total Sample	ASD:	Atypical:
	N = 78	n = 31	n = 47
Age: Years			
Mean (SD)	7.71 (3.15)	6.61 (3.08)	8.43 (3.01)
Range	3-15	3-15	3-14
Gender			
Male	70%	77%	65%
Female	30%	23%	35%
Ethnicity			
Caucasian	70%	74%	68%
African American	18%	20%	17%
Hispanic	3%	3%	2%
Other/Unknown	9%	3%	13%

All of the items, excluding imitates facial expressions, repetitive touching of objects, walks or runs on toes/balls of feet, verbally labels facial expressions, and shows empathy were significant predictors by themselves. Wilks' lambda was significant, $\lambda = .16$, $\chi^2 = 115.91$, $p <$ .001 for the function, indicating that all of the items together significantly discriminated between diagnostic groups. A canonical correlation of .94 indicated that the model accounted for 88.55% of the variation between groups. When the DA was run again with the non-significant items listed above removed, variability accounted for by all of the items and prediction of group membership decreased. As a result, all 45 items were retained for inclusion in the final version of the ASD-OC (see Appendix C).

DSM-IV vs. DSM-V

A subset of the total sample who completed the DSM-IV-TR/ICD-10 checklist (n = 78) was analyzed. Participants were divided into groups based on DSM-IV-TR and DSM-V diagnostic criteria for ASD. The DSM-IV-TR/ICD-10 checklist was used to assess the DSM criteria for each version. In order to meet DSM-IV-TR criteria for a diagnosis of autistic disorder or PDD-NOS, at least three items had to be endorsed on the DSM-IV-TR/ICD-10 checklist: two impairments in social interaction and one in either communication or repetitive, stereotyped, or restricted patterns. Based on DSM-IV-TR criteria, 38 participants qualified for a diagnosis of ASD and 40 did not.

In order to meet DSM-V criteria, three impairments in socialization and two in restricted interests and repetitive behaviors needed to be endorsed, as outlined in the proposed DSM-V diagnostic criteria (APA, 2010). While the DSM-IV-TR/ICD-10 checklist includes all three of the social communication and social interaction symptoms listed as criteria in the DSM-V, it is missing the proposed addition of hyper or hypo-reactivity to sensory input under the restricted

interests/repetitive behaviors symptoms. In order to conservatively control for this, any participants who met three impairments in socialization and one criterion for restricted interests/repetitive behaviors were excluded from the analysis (n = 3). Based on DSM-V criteria, 21 participants qualified for a diagnosis of ASD and 57 did not. All of the participants meeting DSM-V diagnostic criteria also met DSM-IV-TR diagnostic criteria. However, 17 participants no longer qualified for a diagnosis of ASD when switching from DSM-IV-TR to DSM-V criteria.

Three groups were created based on DSM version: participants meeting criteria for ASD on both the DSM-IV-TR and DSM-V (DSM-V group; n = 21), participants meeting criteria on only the DSM-IV-TR (DSM-IV-TR group; n = 17), and participants not meeting criteria on either DSM version (control group; n = 40). To keep sample sizes within 1.5 times each other (Field, 2005), 15 participants were randomly removed from the control group, leaving a total of 63 participants for the analysis. No significant differences were found based on gender, ethnicity, and age. Refer to Table 7 for the demographic information of participants included in the DSM-IV-TR and DSM-V analysis.

Table 7
Demographic information by DSM version

	Total Sample	DSM-V:	DSM-IV-TR:	Control:
	N = 63	n = 21	n = 17	n = 25
Age: Years				
Mean (SD)	8.35 (3.12)	7.95 (3.09)	7.88 (3.26)	9.00 (3.06)
Range	3-15	3-15	4-14	3-14
Gender				
Male	73%	81%	82%	60%
Female	27%	19%	18%	40%
Ethnicity				
Caucasian	79%	66%	82%	88%
African American	13%	24%	6%	8%
Hispanic	3%	5%	6%	0%
Other/Unknown	5%	5%	6%	4%

Next, an ANOVA was conducted with DSM version as the independent variable and total score on the ASD-OC as the dependent variable. The main effect of DSM version was significant, $F(2,205) = 8.13$, $p = .001$, indicating that the groups significantly differed from each other on clinician rated autism symptoms. As the assumption of homogeneity of variances was violated, $F(2,60) = 9.40$, $p < .001$, Games-Howell post hoc analyses were interpreted. Children in the DSM-V group ($M = 15.41$) and children in the DSM-IV-TR group ($M = 24.05$) scored significantly higher (i.e., indicating more symptoms of ASD) than children in the control group ($M = 4.68$). However, no significant difference emerged between participants in the DSM-V and DSM-IV-TR groups.

Discussion

The purpose of the current study was to examine the preliminary psychometric properties of a new observation measure for ASD, the ASD-OC. Interrater reliability for each item ranged from -.19 – 1.00, with the majority of items having a coefficient larger than .60, indicating good to excellent interrater reliability. The interrater reliability for the scale as a whole was excellent, with a coefficient of .96. Five items were removed after this analysis given poor interrater reliability coefficients of less than .40 (i.e., peculiar body posture, fascination with spinning objects, intonation, phrasing to change meaning, and repeating words/phrases). Two additional items had fair interrater reliability (i.e., abnormal motor movement and pedantic speech) between .40 and .59, but were retained for further analyses. Mean interrater reliability for the 49 retained items was good $\alpha = .73$, with coefficients ranging from .43 to 1.00.

Item analysis resulted in the removal of an additional four items given little to no variance in scores: staring at hands, pronoun reversal, speaks overly precise, and repeatedly provides facts about a specific topic. The mean inter-item correlation for the remaining 45 items was .33 (range = -.15 – .86) and the mean item-scale correlation was .56 (range = .08 – .83). Only five items had item-scale correlations less than .30; however, as their removal did not increase alpha, those items were retained for further analyses. Internal consistency of the ASD-OC as measured by Cronbach's alpha was excellent at .96.

Regarding convergent validity, the ASD-OC showed a strong positive correlation ($r = .83$ with the CARS total score as predicted. This finding was expected given that the CARS and ASD-OC both measure observable symptoms of autism as rated by the clinician. The CARS has been well researched over the years and has strong psychometric properties with regard to diagnosing autistic disorder (Schopler et al., 2010). Therefore, strong convergent validity with

the CARS demonstrates the potential of the ASD-OC as a valid measure of symptoms of autistic disorder. To examine divergent validity, the ASD-OC was correlated with the Daily Living domain of the VABS-II. Despite predictions of a low or non-significant correlation, there was a significant, moderate negative correlation between the ASD-OC and Daily Living domain total scores ($r = -.43$). This finding is not surprising given that researchers have found that symptom severity in atypically developing children often negatively impacts adaptive skills (Matson, Dempsey, & Fodstad, 2009; Matson, Rivet, Fodstad, Dempsey, & Boisjoli, 2009). Developmental delays observed in children with both ASD and other disorders are often broad in nature and impact a variety of behaviors, including daily living skills such as eating, dressing, and hygiene. Worth noting, however, is that the correlation between the ASD-OC and CARS was stronger than the correlation between the ASD-OC and Daily Living subdomain. That is, while symptom severity as measured by ASD-OC is moderately related to poorer adaptive skills, it is much more strongly related to another established measure of core symptoms of autism.

As a measure of criterion validity, mean total scores for the ASD-OC were compared by diagnostic group to determine if total score differed significantly based on diagnosis. Based on the results of an ANOVA, children with a diagnosis of ASD scored significantly higher than both typically and atypically developing children, as predicted. However, there was no significant difference on ASD-OC total scores between typically and atypically developing children. Thus, it appears that the ASD-OC is a valid measure of ASD symptoms and is able to discriminate between children with ASD and children without ASD. While it may not discriminate between atypically and typically developing children based on the current sample, this is not the intent of the measure.

The diagnosis of ASD in children under the age of 3 years is difficult due to great variability in the frequency and severity of symptoms at this young age (Lord, 1995; Moore & Goodson, 2003; Wing et al., 2011), it was hypothesized that children under 36 months would score significantly lower on the ASD-OC than children older than 36 months. However, the opposite was found with a *t*-test yielding a significantly higher total ASD-OC score for children less than 36 months of age. Based on visual inspection, regardless of diagnosis, children under the age of 36 months all scored similarly high on the ASD-OC. The most plausible explanation for these findings involves the nature of the sample used for this study. The majority of children less than 36 months seen in the clinic sampled were referred from a state early intervention program after having screened in the at-risk for ASD. As a result, these children were likely all exhibiting symptoms of developmental delay. Therefore, even children who did not receive a diagnosis of ASD may have been recommended to return for re-evaluation in a year, when symptoms could be better assessed. Given the confounds of this particular age sample, children under 36 months were excluded from further analyses.

To further evaluate the ability of the ASD-OC to discriminate between diagnostic groups, a DA was run to establish which items of the ASD-OC could predict group membership between children with ASD and atypically developing children. A control group of typically developing children was unable to be used in this analysis due to small sample size. However, given that an ANOVA demonstrated no significant differences between typical and atypically developing children on the ASD-OC total score, the results of the DA are still useful for providing preliminary psychometric properties of the ASD-OC. The 45 items retained after the initial reliability and validity analyses were significantly able to predict group membership. They accounted for nearly 89% of the variance between the two diagnostic groups. The DA further

confirmed the validity of the ASD-OC for assessing symptoms of ASD and its ability to discriminate between children with ASD and children who are atypically developing.

As the proposed changes to the DSM-V for ASD have been quite controversial, it appeared relevant to examine how children meeting criteria for differing versions compared on a measure of symptoms of ASD. All children included in the sample who met criteria according to the DSM-V, met criteria according to DSM-IV-TR. However, 17 of the children no longer met criteria for an ASD when moving from DSM-IV-TR to DSM-V criteria. Based on the results of an ANOVA, children meeting either version of DSM criteria scored significantly higher on the ASD-OC than children not meeting criteria on either version. There were no significant differences on ASD-OC scores between children meeting DSM-IV-TR criteria only and children meeting DSM-V criteria. This indicates that children identified as high in ASD symptoms on the ASD-OC would coordinate with the current DSM-IV-TR criteria, as well as the proposed new DSM-V criteria. That said, these findings highlight some critical issues with the proposed DSM-V criteria that have been previously brought up by clinicians and researchers. Based on the sample used in the current study, 17 children meeting criteria according to the DSM-IV-TR did not meet the DSM-V proposed criteria. However, according to the ASD-OC, they exhibited symptoms of a severity consistent with children who did meet DSM-V criteria for ASD. These findings are consistent with research regarding DSM-IV-TR and DSM-V criteria and children's scores on the ASD-DC (Worley, 2010). Therefore, with the DSM-V, children with symptoms of ASD severe enough to previously warrant a diagnosis and treatment may no longer receive a diagnosis making service availability for these children questionable.

As previously described, many observation measures available for ASD are limited in their use due to lack of research regarding psychometric properties, outdated norms, difficulty

differentiating between ASD and atypically developing children, and intensive time and money investments for training. Of those measures, the CARS appears to be have the strongest psychometric properties and is most similar to the ASD-OC. Both are brief, unstructured observation measure, requiring little formal training to conduct. For a new measure to be truly useful, it must not only perform as well as the CARS, but also surpass it in areas where the CARS may be lacking. While the ASD-OC currently appears to be on par with the CARS regarding initial reliability and validity, future research, as described later, may provide support for the use of the ASD-OC in areas not addressed by the CARS including specialized populations within ASD (e.g., very young children and adults), cut-off scores within the spectrum, and correspondence with proposed DSM-V criteria. In addition, the ASD-OC has companion measures available for a comprehensive, multi-informant based assessment (i.e., ASD-DC, BISCUIT). These measures not only have diagnostic scales but also subscales measuring comorbidity and problem behaviors. The preliminary psychometric properties presented in the current study make the case for continued research with the ASD-OC to investigate its potential use in areas where other observation measures are lacking.

In addition to use in diagnostic assessment, an observation measure for ASD may be beneficial for identification of treatment targets and progress monitoring. Livanis and Mouzakitis (2010) indicated that even the most commonly used ASD screeners, including the CARS and now CARS-2 have not been studied regarding treatment validity. They note that the CARS-2 does not identify key areas that could serve as targets for treatment intervention. Livanis and Mouzakitis argue that previous researchers have found that language and social interaction items on measures may be the best indicators for treatment targets. However, when examining the CARS-2, they found that only 33% of the items assess language and/or social

interaction domains (Livanis & Mouzakitis, 2010). On the other hand, 82% of the items on the ASD-OC assess domains of language or social interaction, making it a potentially more useful measure with regard to treatment selection and progress monitoring.

The current study supports the reliability and validity of the ASD-OC as a new observation measure for ASD. However, there are limitations to consider and address in future research. Due to the nature of the sample used, different subsets of the total sample were used for each analysis. This was a result of assessment protocols varying based on individual needs and referral reasons (e.g., not all children received the VABS-II, two clinicians were only assigned to some cases). From a clinical standpoint, individualized assessment is considered best practice; however, it can make comparisons for research purposes somewhat complicated. As subgroups of the total sample were used, the sample size available was less than optimal in some analyses. However, small samples sizes are not an uncommon feature in research on ASD. Many studies on ASD report sample sizes of 18 to 50, due to difficulties finding large groups of participants within this population (Bishop & Norbury, 2002; Garfin, McCallon, & Cox, 1988; Moore & Goodson, 2003; Yirmiya, Sigman, & Freeman, 1994). Given that this is a preliminary study regarding the psychometric properties of the ASD-OC, the current findings certainly support continued data collection and research to further evaluate the use of this measure.

Using larger sample sizes to replicate the reliability and validity found in the current study would be essential for future research. In addition, a larger sample size would allow for a factor analysis of the measure leading to cut off scores for diagnosis of ASD. It would also be interesting to further examine the issue of using the ASD-OC with children under 36 months as little variability in scores was observed in the current study regardless of diagnosis. A larger, more diverse sample of children under 36 months would better establish the psychometric

properties of the ASD-OC for this young age group. Few observation measures have been developed to address this younger age group and revisions to the existing ASD-OC that lead to accurate early diagnosis would differentiate the ASD-OC from existing observation measures. On the other extreme, little research has been conducted with a population of older adolescents and adults with ASD. Data collection and analyses of the ASD-OC with this older population may also provide support for use of the ASD-OC with yet another subset of the ASD population.

Parent report measures of ASD that coordinate with the ASD-OC are also available (i.e., ASD-DC, BISCUIT). Investigation into the incremental validity of adding the ASD-OC to these measures may support the use of this package of assessments for a comprehensive diagnostic assessment. Finally, to further set the ASD-OC apart from other measures of its kind, cut-off scores that differentiate between the various types of ASD (i.e., autism, PDD-NOS, Asperger's) would be progressive for the field. That said, proposed changes in the DSM-V call for the use of a severity indicator in lieu of differentiating types of ASD. In either case, future research regarding score differences on the ASD-OC within the spectrum should be pursued. Overall, the ASD-OC appears to be a reliable and valid observation measure of ASD worthy of continued research efforts that may set it apart from existing observation measures, providing clinicians with a valuable diagnostic tool.

References

American Psychiatric Association. (1952). *Diagnostic and statistical manual of mental disorders* (1st ed.). Washington, D.C.: Author.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, D.C.: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., revised).Washington, D.C.: Author.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, D.C.: Author.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed, Text Revision.). Washington, D.C.: Author.

American Psychiatric Association. (2010). Autism Spectrum Disorder. *DSM-V development.* Retrieved September 9, 2011 from http://www.dsm5.org/ProposedRevision/Pages/proposedrevision.aspx?rid=94

Amir, R. E., Van den Veyver, I. B., Wan, M., Tran, C. Q., Francke, U., & Zoghbi, H. Y. (1999). RS is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nature Genetics, 23,* 185-188.

Bailey, A., LeCouteur, A., Gottesman, I., Bolton, P., Simonoff, E., Yuzda, F. Y., & Rutter, M. (1995). Autism is a strongly genetic disorder: Evidence from a British twin study. *Psychological Medicine, 25,* 63-77.

Baumeister, A. A., & Forehand, R. (1973). Stereotyped acts. In N. R. Ellis (Ed.), *International review of research in mental retardation* (pp. 55-96). New York: Academic Press.

Berkson, G. (1983). Repetitive stereotyped behaviours. *American Journal of Mental Deficiency, 88*, 239-246.

Berkson, G., Tupa, M., & Sherman, L. (2001). Early development of stereotyped and self-injurious behaviours: I. Incidence. *American Journal on Mental Retardation, 106*, 539-547.

Bettelheim, B. (1959). Feral children and autistic children. *The American Journal of Sociology, 64*, 455-467.

Bishop, D. V. M., & Norbury, C. F. (2002). Exploring the borderlands of autistic disorder and specific language impairment: a study using standardized diagnostic instruments. *Journal of Child Psychology and Psychiatry, 43*, 917-929.

Bleuler, E. (1950). Dementia praecox or the group of schizophrenias. Oxford: International University Press.

Bodfish, J. W., Symons, F. J., Parker, D. E., & Lewis, M. H. (2000). Varieties of repetitive behaviour in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders, 30*, 237-243.

Bryson, S. E., Zwaigenbaum, L., Brian, J., Roberts, W., Szatmari, P, Rombough, V., & McDermott, C. (2007). A prospective case series of high-risk infants who developed autism. *Journal of Autism and Developmental Disorders, 37,* 12-24.

Bryson, S. E., Zwaigenbaum, L., McDermott, C., Rombough, V., Brian, J. (2008). The Autism Observation Scale for Infants: Scale development and reliability data. *Journal of Autism and Developmental Disabilities, 38,* 731-738.

Buie, T., Campbell,, D. B., Fuchs III, G. J., Furuta, G. T., Levy, J., VandeWater, J., . . .Winter, H. (2010). Evaluation, diagnosis, and treatment of gastrointestinal disorders in individuals with ASDs: A consensus report. *Pediatrics*, *125,* S1-S18.

Centers for Disease Control and Prevention. (2010). Developmental milestones. *Learn the signs. Act early.* Retrieved September 12, 2011 from http://www.cdc.gov/ncbddd/actearly/milestones/index.html

Chen, R. T., & DeStefano, F. (1998). Vaccine adverse events: Causal or coincidental? *The Lancet, 351,* 611-612.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284-290.

Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86,* 127-137.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7,* 309-319.

Cohen, D. J., Caparulo, B. K., Gold, J. R., Waldo, M. C., Shaywitz, B. A., Ruttenberg, B. A., & Rimland, B. (1978). Agreement in diagnosis: Clinical assessment and behavior rating scales for pervasively disturbed children. *Journal of the American Academy of Child Psychiatry, 17,* 589-603.

Constantino, J. N., Przybeck. T., & Friesen. D, (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *Journal of Developmental Behavior Pediatrics*, *21*, 2-11.

Cox, A., Klein, K., Charman, T., Baird, G., Baron-Cohen, S., Swettenham, J., ... Wheelwright, S. (1999). Autism spectrum disorders at 20 and 42 months of age: Stability of clinical and ADI-R diagnosis. *The Journal of Child Psychology and Psychiatry, 40*, 719-732.

Creak, M. (1961). Schizophrenia syndrome in childhood: Progress report of a working party. *Cerebral Palsy Bulletin, 3,* 501-504.

Crocker, L., & Algina, J. (1986). *Introduction to Modern and Classical Test Theory*. Harcourt Fort Worth, TX: Brace Jovanovich.

Dales, L., Hammer, S. J., & Smith, N. J. (2001). Time trends in autism and in MMR immunization coverage in California. *Journal of the American Medical Association, 285,* 1183-1185.

de Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Mulder, E., Volkmar, F., & Minderaa, R. (2004). Interrelationship between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with mental retardation. *Journal of Autism and Developmental Disorders, 34,* 129-137.

Despert, J. L., & Sherwin, A. C. (1958). Further examination of diagnostic criteria in schizophrenic illness and psychoses of infancy and early childhood. *The American Journal of Psychiatry, 114,* 784-790.

Devillis, R.F. (1991). *Scale Development: Theory and applications.* Newbury Park, CA: Sage Publications.

DiLavore, P. C., Lord, C., & Rutter, M. (1995). The Pre-Linguistic Autism Diagnostic Observation Schedule. *Journal of Autism and Developmental Disorders, 25,* 355-379.

Eisenberg, L., & Kanner, L. (1956). Early infantile autism 1943-55. *American Journal of Orthopsychiatry, 26,* 556-566.

Faul, F., & Erdfelder, E. (1992). GPOWER: A priori, post-hoc, and compromise power analyses for MS-DOS [Computer software] Bonn, FRG: Bonn University, Department of Psychology.

Fewell, R. R. (1991). *Play Assessment Scale (5th Revision).* Unpublished manuscript, University of Miami School of Medicine.

Field, A. (2005). *Discovering statistics using SPSS*. London: Sage Publications.

Folstein, S., & Rutter, M. (1977). Genetic influences and infantile autism. *Nature, 265,* 726-728.

Fombonne, E. (1999).The epidemiology of autism: A review. *Psychological Medicine, 29,* 769-786.

Frombonne E. (2002). Prevalence of childhood disintegrative disorder. *Autism*, *6*, 149-157.

Fombonne, E. (2003). Epidemiological surveys of autism and other pervasive developmental disorders: An update. *Journal of Autism and Developmental Disorders, 33,* 365-382.

Fombonne, E. (2005). The changing epidemiology of autism. *Journal of Applied Research in Intellectual Disabilities, 18,* 281-294.

Freeman, B. J., Guthrie, D., Ritvo, E. R., Schroth, R., Glass, R., & Frankel, F. (1979). Behavior Observation Scale: Assessing behavioral similarities and differences between autistic and mentally retarded children. *Psychological Reports, 44,* 519-524.

Freeman, B. J., Ritvo, E. R., Guthrie, D., Schroth, P., & Ball, J. (1978). The Behavior Observation Scale for Autism: Initial methodology, data analysis and preliminary findings on 89 children. *Journal of the American Academy of Child Psychiatry, 17,* 576-588.

Freeman, B. J., Ritvo, E. R., & Schroth, P. C. (1984). Behavior assessment of the syndrome of autism: Behavior observation system. *Journal of the American Academy of Child Psychiatry, 23,* 588-594.

Freeman, B. J., & Schroth, P. C. (1984). The development of the behavior observation system. *Behavioral Assessment, 6,* 177-187.

Frith, U. (1991). *Autism and Asperger Syndrome*. Cambridge, England: Cambridge University Press.

Frith, U. (2004). Emmanuel Miller lecture: Confusions and controversies about Asperger syndrome. *Journal of Child Psychology and Psychiatry, 45,* 672–686.

Garfin, D., McCallon, D., & Cox, R. (1988). Validity and reliability of the Childhood Autism Rating Scale with autistic adolescents. *Journal of Autism and Developmental Disorders, 18,* 367-368

Hagberg, B., Aicardi, J., Dias, K., & Ramos, O. (1983). A progressive syndrome of autism, dementia, ataxia, and loss of purposeful hand use in girls: Rett's syndrome: Report of 35 cases. *Annals of Neurology, 14*, 471-479.

Hagberg, B., Goutières, F., Hanefeld, F., Rett, A., & Wilson, J. (1985). Rett syndrome: Criteria for inclusion and exclusion. *Brain and Development, 7*, 372-373.

Haynes, S. N., & O'Brien, W. H. (2000). *Principles and practice of behavioral assessment.* New York: Plenum Publishing Corporation.

Heflin, L. J., & Alaimo, D. F. (2006). *Students with autism spectrum disorders.* Columbus, Ohio: Pearson Prentice Hall.

Irwin, J., MacSween, J., & Kerns, K. (in press). History and evolution of the autism spectrum disorders. In J. L. Matson & P. Sturmey (Eds.), International handbook of autism and pervasive developmental disorders. New York: Springer.

Joseph, R. M., Tager-Flusberg, H., & Lord, C. (2002). Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *Journal of Child Psychology and Psychiatry, 43,* 807-821.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2,* 217-250.

Kanner, L. (1971). Follow-up study of eleven autistic children originally reported in 1943. *Journal of Autism and Childhood Schizophrenia, 1,* 119-145.

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S. … Fein, D. (2008). Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38,* 606-615.

Klinger, L. G., & Renner, P. (2000). Performance-based measures in autism: Implications for diagnosis, early detection, and identification of cognitive profiles. *Journal of Clinical Child Psychology, 29,* 479-492.

Le Couteur, A., Bailey, A. J., Goode, S., Pickles, A., Robertson, S., Gottesman, I., … Rutter, M. (1996). A broader phenotype of autism: The clinical spectrum in twins. *Journal of Child Psychology and Psychiatry, 37,* 785-801.

Leech, N. L., Barrett, K. C., & Morgan, G. A. (2008). *SPSS for intermediate statistics: Use and interpretation (3rd ed.).* New York: Taylor & Francis Group.

Lewis, V., & Boucher, J. (1988). Spontaneous, instructed and elicited play in relatively able autistic children. *British Journal of Developmental Psychology, 6*, 325-339.

Livanis, A., & Mouzakitis, A. (2010). The treatment validity of autism screening instruments. *Assessment for Effective Intervention, 35,* 206-217.

Lord, C. (1995). Follow-up of two-year-olds referred for possible autism. *Journal of Child Psychology and Psychiatry, 36*, 1365-1382.

Lord, C. (2010). Autism: From research to practice. *American Psychologist, 65,* 815-826.

Lord, C., Risi, S., Lambrecht, Cook, E. H., Leventhal, B. L., DiLavore, P. C., … Rutter, M. (2000). The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficit associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30,* 205-223.

Lord, C., Rutter, M., Goode, S., Heemsbergen, J., Jordan, H., Mawhood, L., & Schopler, E. (1989). Autism Diagnostic Observation Schedule: A standardized observation of

communicative and social behavior. *Journal of Autism and Developmental Disorders, 19,* 185-212.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 24*, 659-685.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., … Lord, C. (2009). The Autism Diagnostic Observation Schedule – Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 39,* 1305-1320.

Matson, J. L., & Boisjoli, J. A. (2007). Differential diagnosis of PDDNOS in children. *Research in Autism Spectrum Disorders, 1,* 75-84.

Matson, J. L., Dempsey, T., Fodstad, J. C. (2009). The effect of autism spectrum disorders on adaptive independent living skills in adults with severe intellectual disability. *Research in Developmental Disabilities, 30,* 1203-1211.

Matson, J. L., González, M. L., Wilkins, J., & Rivet, T. T. (2008). Reliability of the Autism Spectrum Disorders-Diagnostic for Children (ASD-DC). *Research in Autism Spectrum Disorders, 2, 696-706.*

Matson, J. L., Nebel-Schwalm, M., & Matson, M. L. (2007). A review of methodological issues in the differential diagnosis with autism spectrum disorders in children. *Research in Autism Spectrum Disorders, 1,* 38-54.

Matson, J. L., Rieske, R. D., & Tureck, K. (2011). Additional considerations for the early detection and diagnosis of autism: Review of available instruments. *Research in Autism Spectrum Disorders, 5,* 1319-1326.

Matson, J. L., Rivet, T. T., Fodstad, J. C, Dempsey, T., & Boisjoli, J. A. (2009). Examination of adaptive behavior differences in adults with autism spectrum disorders and intellectual disability. *Research in Developmental Disabilities, 30,* 1317-1325.

Matson, J. L., Stabinsky-Compton, L., & Sevin, J. A. (1991). Comparison and item analysis of the MESSY for autistic and normal children. *Research in Developmental Disabilities, 12,* 361-369.

Matson, J. L., Wilkins, J., & Gonzalez, M. (2008). Early identification and diagnosis of autism spectrum disorders in young children and infants: How early is too early? *Research in Autism Spectrum Disorders, 2,* 75-84.

Mayes, R. & Horowitz, A.V. (2005). DSM-III and the revolution in the classification of mental illness. *Journal of the History of the Behavioral Sciences*, *41*, 249-267.

Miller, L. J., Anzalone, M. E., Lane, S. J., Cermak, S. A., & Osten, E. T. (2007). Concept evolution in sensory integration: A proposed nosology for diagnosis. *American Journal of Occupational Therapy, 61*(2), 135-140.

Millward, C., Ferriter, M., Calver, S. J., & Connell-Jones, G. G. (2008). Gluten- and casein-free diets for autistic spectrum disorder. *Cochrane Database of Systematic Reviews,* 2. doi: 10.1002/14651858.CD003498.pub3

Moore, V., & Goodson, S. (2003). How well does early diagnosis of autism stand the test of time? *Autism, 7*, 47-63.

Morgan, S. (1988). Diagnostic assessment of autism: A review of objective scales. *Psychoeducational Assessment, 6,* 139-151.

Nicholas, J. S., Charles, J. M., Carpenter, L. A., King, L. B., Jenner, W., & Spratt, E. G. (2008). Prevalence and characteristics of children with autism spectrum disorders. *Annals of Epidemiology, 18,* 130-136.

Njardvik, U., Matson, J. L., Cherry, K. E. (1999). A comparison of social skills in adults with autistic disorder, pervasive developmental disorder not otherwise specified, and mental retardation. *Journal of Autism and Developmental Disorders, 29,* 287-295.

Osterling, J., & Dawson, G. (1994). Early recognition of children with autism: A study of first birthday home videotapes. *Journal of Autism and Developmental Disorders, 24,* 247–257.

Ozonoff, S. Goodlin-Jones, B.L. & Solomon, M. (2007). Autism spectrum disorder In E. J. Mash & R.A. Barkley (Eds.), *Assessment of Childhood Disorder - Fourth Edition* (PP. 478-508). New York: Guilford Press.

Peltola, H., Patja, A., Leinikki, P., Valle, M. Davidkin, I., & Paunio, M. (1998). No evidence for measles, mumps, and rubella vaccine-associated inflammatory bowel disease or autism in a 14-year prospective study. *The Lancet, 351,* 1327-1328.

Posserud, M. B., Lundervold, A. J. & Gillberg, C. (2006). Autistic features in a total population of 7–9-year-old children assessed by the ASSQ (Autism Spectrum Screening Questionnaire). *Journal of Child Psychology & Psychiatry*, *47*, 167-175.

Rimland, B. (1964). *Infantile autism.* London: Methuen.

Ritvo, E. R. (1978). National Society for Autistic Children definition of the syndrome of autism. *Journal of Autism and Childhood Schizophrenia*, *8,*162-169.

Ritvo, E. R., Freeman, B. J., Mason-Brothers, A., Mo, A., & Ritvo, A. M. (1985). Evidence for autosomal recessive inheritance in 46 families with multiple incidence of autism. *American Journal of Psychiatry, 142,* 187-192.

Rivet, T. T. (2010). *Gender differences in core symptomatology in autism spectrum disorders across the lifespan.* Unpublished doctoral dissertation, Louisiana State University, Baton Rouge.

Ruttenberg, B. A., Dratman, M. L., Fraknoi, J., & Wenar, C. (1966). An instrument for evaluating autistic children. *Journal of the American Academy of Child Psychiatry, 5,*453-478.

Ruttenberg, B. A., Kalish, B. I., Wenar, C., & Wolf, E. G. (1977). The behavior rating instrument for autistic and other atypical children. Philadelphia: Developmental Center for Autistic Children.

Ruttenberg, B. A., Wolf-Schein, E. G., & Wenar, C. (1991). The behavior rating instrument for autistic and other atypical children: Instruction manual. Wood Dale, IL: Stoelting.

Rutter, M. (1968). Concepts of autism: A review of research. *Journal of Child Psychology and Psychiatry, 9,* 1-25.

Rutter, M. (1978).Diagnosis and definition. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment.* New York: Plenum Press.

Rutter, M. (2005). Incidence of autism spectrum disorders: Changes over time and their meaning. *Acata Pediatrica*, *94*, 2-15.

Rutter, M., & Bartak, L. (1971). Causes of infantile autism: Some considerations from recent research. *Journal of Autism and Developmental Disorders, 1,* 20-32.

Sasso, G. M., Garrison-Harrell, L., & Rogers, L. (1994). The conceptualization of socialization and autism. In T. Scruggs & M. Mastropieri (Eds.), *Relevant research issues in developmental disabilities* (pp. 161-175). New York: Plenum Press.

Schopler, E. (1978). On confusion in the diagnosis of autism. *Journal of Autism and Childhood Schizophrenia*, *8,* 137-138.

Schopler, E., Reichler, R. J., DeVellis, R. F., & Dalay, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders, 10,* 91-103.

Schopler, E., Reichler, R. J., & Rocher-Renner, B. (1988). *The Childhood Autism Rating Scale (CARS).* Los Angeles: Western Psychological Services.

Schopler, E., Van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *The Childhood Autism Rating Scale, (2nd ed.).* Los Angeles: Western Psychological Services.

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

Stone, W. L., Coonrod, E. E., Turner, L. M., & Pozdol, S. L. (2004). Psychometric properties of the STAT for early autism screening. *Journal of Autism and Developmental Disorders, 34,* 691-701.

Stone, W. L., Coonrod, E. E., & Ousley, O. Y. (2000). Brief report: Screening tool for autism in 2-year-olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders, 30,* 607-612.

Stone, W. L. & Ousley, O. Y. (1997). *STAT Manual: Screening Tool for Autism in Two-Year-Olds.* Unpublished manuscript, Vanderbilt University.

Stone, W. L., Ousley, O. Y., & Littleford, C. (1997). Motor imitation in young children with autism: What's the object? *Journal of Abnormal Child Psychology, 25,* 475–485.

Stone, W. L., Ousley, O. Y., Yoder, P. J., Hogan, K. L., & Hepburn, S. L. (1997). Nonverbal communication in 2- and 3-year old children with autism. *Journal of Autism and Developmental Disorders, 27,* 677–696.

Sturmey, P., & Sevin, J. A. (1994). Defining and assessing autism. In J. L. Matson (Ed.). *Autism in children and adults: Etiology, assessment, and intervention* (pp. 13-16). Pacific Grove, CA: Brooks/Cole.

Szatmari, P. (2000). The classification of autism, Asperger's syndrome, and pervasive developmental disorder. *Canadian Journal of Psychiatry, 45,* 731-738.

Taylor, B., Miller, E., Farrington, C. P., Petropoulos, M., Favot-Mayaud, I., Li, J., & Waight, P. A. (1999). Autism and measles, mumps, and rubella vaccine: No epidemiological evidence for a causal association. *The Lancet, 353,* 2026-2029.

Teal, M. B., & Weibe, M. J. (1986). A validity analysis of selected instruments used to assess autism. *Journal of Autism and Developmental Disorders, 16,* 485-494.

Tidmarsh, L., & Volkmar, F. R. (2003). Diagnosis and epidemiology of autism spectrum disorders. *The Canadian Journal of Psychiatry, 48,* 517-525.

Turner, M.A. (1999). Annotation: Repetitive behaviour in autism: A review of psychological research. *The Journal of Child Psychology and Psychiatry, 40*, 839-849.

Volkmar, V. (1997). Childhood Disintegrative Disorder. In Cohen, D.J., & Volkmar, F.R. (Eds*) Handbook of Autism and Pervasive Developmental Disorders*. New York: John Wiley & Sons.

Volkmar, F. R., Carter, A., Grossman, J., & Klin, A. (1997). Social development in autism. In D. J. Cohen & F. R. Volkmar (Eds.), *Handbook of autism and pervasive developmental disorders, second edition* (pp. 173-194). New York: John Wiley & Sons, Inc.

Volkmar, F., Chawarska, K., & Klin, A. (2005). Autism in infancy and early childhood. *Annual Review of Psychology, 56,* 315-336.

Volkmar, F. R., Cicchetti, D. V., Dykens, E., Sparrow, S. S., Leekman, J. F., & Cohen, D. J. (1988). An evaluation of the Autism Behavior Checklist. *Journal of Autism and Developmental Disorders, 18,* 81–97.

Wakefield, A. (1998). Ilial-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *The Lancet, 351,* 637-641.

Wenar, C., & Ruttenberg, B. A. (1976). The use of BRIAAC for evaluating therapeutic effectiveness. *Journal of Autism and Childhood Schizophrenia, 6,* 175-191.

Wing, L. (1981). Asperger's syndrome: A clinical account. *Psychological Bulletin, 11*, 115-129.

Wing, L., Gould, J. & Gillberg, C. (2011). Autism spectrum disorder in the DSM-V: Better or worse than the DSM-IV? *Research in Developmental Disabilities, 32,* 768-773.

Wing, L., & Potter, D. (2002). The epidemiology of autistic spectrum disorders: Is the prevalence rising? *Mental Retardation and Developmental Disabilities, 8*, 151-161.

World Health Organization. (1992). *International classification of diseases (10th ed.)*. Geneva, Switzerland: World Health Organization.

Worley, J. A. (2011). Comparing symptoms of autism spectrum disorders using the current *DSM-IV-TR* diagnostic criteria and the proposed *DSM-V* diagnostic criteria. Unpublished doctoral dissertation, Louisiana State University, Baton Rouge.

Yirmiya, N., Sigman, M., & Freeman, B. J. (1994). Comparison between diagnostic instruments for identifying high-functioning children with autism. *Journal of Autism and Developmental Disorders, 24,* 281-291.

Zwaigenbaum, L., Bryson, S., Lord, C., Rogers, S., Carter, A., Carver, L. … Yirmiya, N. (2009). Clinical assessment and management of toddlers with suspected autism spectrum disorder: Insights from studies of high-risk infants. *Pediatrics, 123,* 1383-1391.

Zwaigenbaum, L., Bryson, S., Rogers, T., Roberts, W., Brian, J., & Szatmari, P. (2005). Behavioral manifestations of autism in the first year of life. *International Journal of Developmental Neuroscience, 23,* 143-152.

Appendix A
DSM-IV-TR/ICD-10 Checklist

DSM-IV-TR/ICD-10 Checklist

Please indicate "yes" if the following applies to your child/client. Indicate "no" if the item does not apply to your child/client.

1. Impairment in social interaction, such as:
____ a. Impairment in the use of multiple nonverbal behavior, such as eye-to-eye gaze (e.g., eye contact), body posture, or gestures[a, b]
____ b. Failure to develop peer relationships appropriate to developmental level (e.g., little to no interest in forming friendships or lack understanding of how to interact socially with others)[a, b]
____ c. Lack of spontaneous seeking to share enjoyment, interest or achievements with others (e.g., not showing, bringing, or pointing out objects of he/she finds interesting)[a]
____ d. Lack of social or emotional reciprocity (e.g., not actively participating in social play or games, preferring solitary activities)[a, b]
____ e. Rarely seeking or using others for comfort in times of stress or offering comfort or affection to others in stress[c]

2. Impairments in communication, such as:
____ a. Delay in development or lack of spoken language (i.e., not accompanied by an attempt to communicate through alternative ways to communicate such as gestures or mime)[a,b]
____ b. In those with adequate speech, impairment to initiate or sustain conversations with others[a,b]
____ c. Stereotyped and repetitive use of language or idiosyncratic language (e.g., using words in a peculiar or odd way)[a,b]
____ d. Lack of varied, spontaneous make-believe play (e.g., pretend play) or social imitative play (e.g., imitating adults) appropriate to developmental level[a]
____ e. Lack of emotional response to others' verbal or non-verbal communication[c]
____ f. Lack of variation in the rhythm or emphasis of speech (e.g., speech is monotone; without change)[b]
____ g. Impaired use of gestures to aid spoken communication[c]

3. Restricted, repetitive and stereotyped patterns of behavior, interest or activities such as:
____ a. Preoccupation with one or more stereotyped and restricted patterns of interest of abnormal intensity or focus (e.g., few interests)[a,b]
____ b. Inflexible adherence to specific, nonfunctional routines or rituals[a,b]
____ c. Stereotyped and repetitive motor mannerisms (e.g., hand or finger flapping or twisting, or other complex whole-body movements such as rocking, dipping or swaying)[a,b]
____ d. Persistent preoccupation with parts of objects (e.g., buttons, parts of the body)[a,b]
____ e. Specific attachments to unusual objects (e.g., string)[c]
____ f. Distress over changes in small, non-functional details of the environment[b]

4. _____ Delays or abnormal functioning in at least one of the previous areas (#1-3) was present prior to age of 3[a,b]

[a] DSM-IV-TR diagnostic criteria; [b] ICD-10 diagnostic criteria; [c] Items are included in descriptions of ICD-10 for clinical use, but not included as specific diagnostic criteria

Appendix B
Original Items of the ASD-OC

1. Looks when name is called.

2. Curiosity with surroundings

3. Use of gestures to communicate (e.g., pointing, hand motions to aid speech, nodding "yes", shaking head "no", shrug shoulders for "don't know", thumbs up for "okay", wave for "hey")

4. Awareness of the unwritten or unspoken rules of social play (e.g., turn taking, sharing)

5. Initiates make-believe or pretend play

6. Follows along/participates in make-believe or pretend play

7. Initiates joint attention

8. Follows along with joint attention

9. Can imitate simple sounds, words, phrases

10. Can imitate simple sentences

11. Can imitate simple physical gestures or movements

12. Can imitate complex physical gestures or movements

13. Asks for help (verbal and/or nonverbal)

14. Can make facial expression of emotion when asked (happy, sad, scared, surprised, etc.)

15. Can imitate facial expression of emotion by clinician upon request (happy, sad, scared, surprised)

16. Able to understand the subtle cues/gestures/body language of others (e.g., sarcasm, crossing arms to show anger, rolling eyes, winking)

17. Interest in another person's side of the conversation (e.g., talks to people with intention of hearing what others have to say)

18. Understanding of age appropriate jokes, figures of speech or sayings (e.g., idioms, irony, implied meaning, sarcasm)

19. Reaction to praise

20. Reaction to correction

21. Shares enjoyment, interests, or achievement with others (e.g, parents, friends, caregivers)

22. Reaction to sounds

23. Facial expression corresponds to environmental events (e.g., smiling, grimacing, questioning/uncertainty)

24. Eye contact when communicating expressively (i.e., speaking, using gestures, sign language, or other means)

25. Eye contact when being spoken to (i.e., receptive)

26. Respect for others' personal space (e.g., stands too close to others)

27. Excessive interest in inanimate objects

28. Peculiar body postures

29. Reaction when transitioning between activities

30. Abnormal, repetitive hand or arm movements

31. Abnormal, repetitive motor movements involving entire body (e.g., whirling/spinning, darting/lunging, rocking, jumping)

32. Abnormal preoccupation with the parts of an object or objects

33. Staring at hands

34. Abnormal fascination with the movement of spinning objects (e.g., closing doors, electric fan blades)

35. Rituals, insistence on sameness, lining things ups, arranging things

36. Repetitive sniffing, touching, feeling, licking, mouthing, tapping of objects or surfaces

37. Walks or runs on toes/balls of feet

38. Verbal items (if nonverbal, rate as 0 and skip a-n)

    a. Rhythm when speaking (e.g., sing-songy)

    b. Intonation when speaking (i.e., uses higher pitch for questions vs. statements; changes pitch on last word of sentence; uses different pitch when speaking to a child vs adult)

    c. Stress/emphasis when speaking (i.e., pronounces words differently based on context; stresses different words in the sentence based on the question)

    d. Use of phrasing (grouping of words, pauses; calm or excited tone) when speaking to change meaning

    e. Rate of speech (e.g., too slow, to fast)

    f. Echolalia (immediate or delayed)

    g. Pronoun reversal (i.e., refers to self in 3rd person rather than "I")

    h. Idiosyncratic use of words or phrases (e.g., listener would have to be familiar with child to understand; makes up own words, uses words in particular ways)

    i. Speaks overly precise or scholarly (e.g., high vocabulary, speaks very properly)

    j. Pedantic speech (i.e., lengthy, sounds like a rehearsed monologue or reading from a book, provides more information and details and elaborates more than relevant/necessary)

    k. Repeatedly provides facts and data about a specific interest

    l. Saying words and phrases repetitively

    m. Starts/initiates conversation/makes small talk

      n. Asks questions

      o. Responds with relevant/on topic comments during conversation

      p. Can verbally label facial expression clinician makes when asked

39. Shows empathy

Appendix C
IRB Approval

Project Report and Continuation Application

(Complete and return to IRB, 131 David Boyd Hall, Direct questions go to IRB Chairman Robert Mathews 578-8692.)

LSU

IRB#: 2609   **Current Approval Expires On:** Jan 6, 2011

Review Type: Expedited   **Risk Factor:** Minimal

PI: Johnny L. Matson, Ph.D.   **Dept:** Psychology   **Phone:** 225-578-3745

Student/Co-Investigator:

Project Title: Developing the Autism Spectrum Disorder (ASD)

Number of Subjects Authorized: 2000

Please read the entire application. Missing information will delay approval!
IRB Security of Data Agreement: http://www.lsu.edu/irb/IRB%20Security%20of%20Data.pdf

I. PROJECT FUNDED BY: _____   LSU Proposal #: _____

II. PROJECT STATUS: Check the appropriate blank(s) and complete the following:

- ◉ 1. **Active,** subject enrollment continuing; # subjects enrolled: 550
- ○ 2. **Active,** subject enrollment complete; # subjects enrolled: _____
- ○ 3. **Active,** subject enrollment complete; work with subject continues.
- ○ 4. **Active,** work with subjects complete; data analysis in progress.
- ○ 5. Project **start postponed** ___/___/___
- ○ 6. **Project complete;** end date ___/___/___
- ○ 7. **Project cancelled:** no human subjects used.

III. PROTOCOL: (Check one).

- ◉ Protocol continues as previously approved
- ○ Changes are requested*
  - --List (on separate sheet) any changes to approved protocol.

IV. UNEXPECTED PROBLEMS: (did anything occur that increased risks to participants):
- ---State number of events since study inception: ∅   since last report: ∅
- ---If such events occured, describe them and how they affect risks in your study, in an attached report
- ---Have there been any **previously unreported** events? Y/Ⓝ

V. CONSENT FORM AND RISK/BENEFIT RATIO:
Do new knowledge or adverse events change the risk/benefit ratio? Y/Ⓝ
Is a corresponding change in the consent form needed? Y/Ⓝ

VI. ATTACH A BRIEF, FACTUAL SUMMARY of project progress/results to show continued participation of subjects is justified; or to provide a final report on project findings.

VII. ATTACH CURRENT CONSENT FORM (only if subject enrollment is continuing); and check the appropriate blank;
- ○ 1. Form is unchanged since last approved
- ◉ 2. Approval of revision requested herewith: (identify changes)

Signature of Principal Investigator: _____   Date: 11/11/10

IRB Action:	✓ Continuation/approved;	Approval Expires: 1/15/11
	___ Disapproved	
	___ File Closed	
Signed _____		Date 11/16/10

Print Form

Vita

Daniene Neal was born in Thibodaux, Louisiana, in 1981. She received her Bachelor of Arts degree in psychology from Nicholls State University in 2003. Thereafter, she received her Master of Science degree in clinical psychology from Mississippi State University in 2005. Following completion of her degrees, she was employed at North Lake Supports and Services Center, Louisiana, as a psychology associate for adults with intellectual disability. She subsequently enrolled in Louisiana State University's Clinical Psychology Doctoral Program in 2007. Her current clinical and research interests are the assessment and treatment of children and adults with Autism Spectrum Disorders and other developmental disabilities.