Developing and Refining Instruments and Methods for Diagnostic and Language Assessment of Young Children with Autism Spectrum Disorders (ASD)

by

So Hyun Kim

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
2012

Doctoral Committee:

Professor Catherine Lord, Co-Chair
Professor Christopher Stephen Monk, Co-Chair
Professor Susan A. Gelman
Professor Dale A. Ulrich

Dedication

To God: "My grace is sufficient for you, for my power is made perfect in weakness." Therefore I will boast all the more gladly about my weaknesses, so that Christ's power may rest on me. 2 Corinthians 12:9. Without the strength, wisdom, and grace God had poured upon me, this dissertation would not have been possible. This dissertation is a testimony of God working through my weaknesses to reveal His glory.

To my husband, Daniel Cheong, for his love, patience, and humor; Kim and Cheong families, especially my parents, for their unconditional love and support; my advisor, Catherine Lord, for her guidance, determination, and passion.

Acknowledgements

This research was supported by grants from the National Institute of Mental Health (NIMH RO1 MH066469 and MH57167), the National Institute of Child Health and Human Development (HD 35482-01), the Simons Foundation, and the Western Psychological Services to Catherine Lord, the Blue Cross Blue Shield Foundation of Michigan research award, the Edward S., Bordin Graduate Research Fund, and a Rackham Doctoral Research Grant awarded to So Hyun Kim.

I am indebted to my advisor, Dr. Catherine Lord, for her mentorship and inspiration. Her extraordinary passion for research, compassion for children with ASD and their families, and dedication for her students have made her exceptional mentor. I am grateful for her continuous effort to empower and challenge me to higher levels. With her mentorship, the first two studies have been published in 2011 in the Journal of Autism and Developmental Disorders (Chapter II) and the Journal of Child Psychology and Psychiatry (Chapter III).

I thank the children and families who participated in the various research projects. I gratefully acknowledge the faculty and staff at the University of Michigan Autism and Communication Disorders Center (UMACC), particularly Kate Gasparrini, Kristin Houck, Alayna Schreier, Dörte Junker, and Shanping Qiu who assisted in collecting and preparing these data and Katherine Gotham, Somer Bishop, Jennifer Richler, Rhiannon

Luyster, Susan Risi, Fiona Miller, Pamela Dixon Thomas, and Suzi Naguib for their guidance. I also thank Kathy Hatfield, Kathryn Larson, Ellen Bucholz, and Mary Yonkovit for their support. I am especially grateful to Elizabeth Buvinger, Themba Carr, and Vanessa Hus for their wisdom, encouragement, and friendship. I finally want to acknowledge Drs. Susan Gelman, Christopher Monk, and Dale Ulrich for their time and effort to improve this dissertation.

Table of Contents

List of Tables

Table

List of Figures

Figure

ABSTRACT

Earlier provision of services and treatments is associated with better outcomes in Autism Spectrum Disorders (ASD). Researchers and clinicians recognize the increasing need for diagnostic instruments that are appropriate for toddlers and young preschoolers to capture the early signs of autism. However, comprehensive assessment of ASD for toddlers and young preschoolers has been compromised by lower diagnostic validity of preexisting instruments for these children. Therefore, the first two studies in this three-study dissertation focus on improving and expanding the valid use of pre-existing diagnostic measures for toddlers and young preschoolers with ASD from 12 to 47 months of age. The first study achieves this by developing new diagnostic algorithms for a widely used diagnostic instrument. The second study is focused on evaluating different diagnostic methods to use information from the instrument included in the first study and another commonly used diagnostic instrument in a way that maximizes the diagnostic validity of the instruments.

Language skills in young children with ASD have been found to be one of the most important variables predicting better outcomes in later childhood and adulthood. However, there have not generally been standardized instruments that measure spontaneous expressive language of children with ASD in a relatively naturalistic setting. Therefore, the third study of this dissertation focuses on developing a new measure for children

with ASD and other communication disorders from 2 to 12 years of age for the valid description of spontaneous language use in a standardized, but naturalistic, setting.

Overall, this dissertation is focused on developing and refining instruments and methods for the diagnostic and language assessment of young children with ASD. The newly developed and identified diagnostic algorithms and methods for toddlers and preschoolers will enhance the early identification and provision of treatment for these young children. The new language measure will allow clinicians and researchers to describe the current level of language and quantify language impairments in relation to autism symptoms for children with ASD. These newly developed and improved diagnostic and language measures will provide useful information for treatment and education programs promoting more positive outcomes for young children with ASD.

Chapter I

Introduction

In the early 1940, Leo Kanner (1943) provided detailed descriptions of 11 children with autism who shared qualities of social aloofness, insistence on sameness and language delays or oddities. At about the same time, Asperger (1944) described four “little professors” who shared qualities of social awkwardness and circumscribed interests, but who had strengths in vocabulary and syntactic aspects of language. In the 1960’s, it was proposed that ASD was a neurobiological disorder (Rimland, 1964; Rutter and Lockyer, 1967; Rutter & Schopler, 1971). Shortly after, Wing and Gould (1978) conceptualized the disorder as the co-occurrence of a triad of impairments in social reciprocity, language comprehension, and play. These deficits in their most extreme characterize autism, but also occur in individuals with other developmental disorders. The behaviors and deficits identified earlier form the base of conceptualizations of autism spectrum disorders (ASD) even today. From these findings came the broader definitions currently used in the Diagnostic and Statistical Manual of the American Psychiatric Association (APA, 1994) of autism and the term “pervasive developmental disorders (PDD),” also referred to as ASD.

ASD is characterized by the presence of symptoms in three domains including social reciprocity, communication, and restricted and repetitive behaviors and interests (Carter, Davis, Klin, & Volkmar, 2005; Williams White, Koenig, & Scahill, 2007). The central defining characteristic of ASD is impairment in social reciprocity. Examples of deficits in this area are lack of eye contact, a narrow range of facial expressions directed to others, and difficulties initiating social overtures such as asking questions and requesting. Individuals with ASD also show impairment in communication including delay or lack of communication strategies. These difficulties are present in both nonverbal (e.g., minimal use of gestures) and verbal (e.g., echolalia, late onset of phrase speech, stereotyped speech) aspects of communication. The third domain consists of symptoms associated with restricted, repetitive behaviors and interests (RRBs). RRBs include a very broad category of behaviors such as repetitive motor manners (e.g., hand flapping), preoccupation with parts of objects (e.g. peering at the wheels of toy cars while spinning them), and adherence to specific, nonfunctional routines (e.g. insisting on taking a certain route to school).

Based on recent findings on epidemiological research, approximately 1 in 100 to 150 children have an ASD in the UK and US (Baird et al., 2006; Center for Disease Control, 2007). The earliest studies of ASD in the 1960's indicated a prevalence rate for relatively narrowly defined autism of 4-5 out of 10,000 (Fombonne, 2007). These figures began to change, with higher rates of autism, suggesting that ascertainment affected estimate rates (Fombonne, 2009). By the late 1980, many studies began to report higher prevalence rates of 13-16 out of 10,000 for autism with even higher rates of the more broadly defined Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) of up to 20-21 out of 10,000. Although the reasons for the increase in the prevalence rates are not fully known, some of the increase can be explained by broadened diagnostic criteria and the development of diagnostic measures with improved validity and reliability (Bishop, Whitehouse, Watt, & Line, 2008).

Because ASDs typically begin when children are infants or toddlers and continue into adulthood, identification of clearly defined behaviors that are necessary and sufficient to diagnoses during infancy and toddlerhood is an important task for more positive outcomes (Lord, Pickles, DiLavore, & Shulman 1996). Infancy and toddlerhood is a period of time of great change in child development. Children begin walking and become able to manipulate objects with much greater dexterity. They start to understand language and their vocabulary exponentially increases. They also begin to demonstrate imaginative play, complex social cognition, and autonomy, allowing them to develop a more sophisticated understanding of social relationships and events. After children go through this period of rapid development, social and communication deficits become more discriminative of children with ASD from those with other developmental disorders.

However, even though some of the ASD symptoms become more evident during infancy and toddlerhood, many behaviors clearly indicative of autism in older children are common in both ASD and other developmental disorders (e.g., language delays, intellectual disabilities) during this period of time. For instance, older children with ASD have marked impairments in initiating and maintaining reciprocal conversations with others. However, toddlers and young preschoolers are not fully competent at having flexible back and forth conversations with others regardless of whether they have ASD or not. This creates a challenge for differentiating children with ASD from those with other developmental in these very young children. Variability in typical development also poses another challenge in the early identification of ASD. For example, variability has been found in the onset of language acquisition and the strategies and mechanisms of language learning process in very young children, which are all affected by both individual differences and environmental factors (Bates, Bretherton, & Snyder, 1988; Fenson et al., 1994; Goldfield, 1987). Because infants and toddlers show differences in achieving developmental milestones, parents of children with ASD may miss or overlook their children's delayed development or lack of change. Thus, discrimination between ASD and other disorders can be complex for infants and toddlers who are still at very basic levels of development (Charman et al., 1998; Volkmar et al., 1994).

For those children whose parents do observe their child failing to achieve developmental milestones early on, referrals for diagnostic assessments are common. In order to make a diagnostic judgment, clinicians rely on diagnostic instruments including clinician observation, parent interview and questionnaires. However, many of these instruments have shown lower sensitivity (the ability to correctly identify ASD cases) and/or specificity (the ability to correctly exclude non-ASD cases) for the differentiation between very young children with ASD and nonspectrum disorders (NS). This is in part because the behaviors and skills organized in the current diagnostic instruments are still primarily based on older preschool children although the criteria for autism were modified in the third edition of the Diagnostic and Statistical Manual-III-Revised (APA, 1987) to be more appropriate for younger children. For example, the period between the child's fourth and fifth birthdays is considered by the Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003), a parent interview, as the age at which symptoms of ASD are clearest.

As researchers and clinicians recognized the increasing need for diagnostic measures appropriate for very young children to capture the early signs of autism, several screening instruments appropriate for toddlers and young preschoolers were recently developed such as the Communication and Symbolic Behavior Scale (CSBS; Wetherby & Prizant, 2003), Early Social Communication Scales (ESCS; Mundy, Hogan, & Doehring, 1996), and Screening Tool for Autism in Two-Year-Olds (STAT; Stone et al., 2000). Diagnostic measures appropriate for toddlers have been also developed. The Toddler module of the Autism Diagnostic Observation Scale (ADOS-T; Lord et al., in press) and the Toddler version of the ADI-R (Rutter et al., 2003) are examples of diagnostic instruments that have been developed for toddlers and young preschoolers.

The ADI-R is a standardized, semistructured, investigator-based interview for parents or caregivers of individuals who have been referred for a possible diagnosis of ASD. The ADI-R provides formal, diagnostic algorithms, summaries of items combined to generate cutoffs for the classifications of "autism" or "autism spectrum." However, toddlers and preschoolers whose caregivers are administered the standard or toddler version of the ADI-R have so far been classified based on the list of items of a *current behavior algorithm* form (as distinguished from an empirically supported diagnostic algorithm). Because the current behavioral algorithm form was adapted from the algorithms developed for older children rather than empirically validated based on younger children, this has resulted in lower sensitivities and/or specificities compared to that of older children and adults (Gotham, Risi, Pickles, & Lord, 2007; Lord, Storoschuk, Rutter, & Pickles, 1993; Wiggins & Robins, 2008). Thus, the focus of the first study was to develop a set of empirically supported diagnostic algorithms for the toddler and regular versions of the ADI-R for toddlers and young preschoolers.

Diagnostic validity of ASD increases when information from multiple sources is combined together. The National Research Council has suggested that a child's developmental history, parent descriptions and current cognitive, social, language and adaptive functioning across a variety of contexts, as well as the judgment of a skilled clinician, are all necessary for appropriate diagnosis and recommendations (National Research Council, 2001). Past studies have also suggested that combining information from multiple sources across raters and measures enhances diagnostic accuracy for the diagnosis of ASD as well as other developmental disorders. For this reason, the ADI-R, a parent interview, and the ADOS, a clinician observation, are intended to be used in combination. Indeed, Risi et al. (2007) found that when the ADI-R and ADOS were used in combination, well balanced, higher sensitivity and specificity were obtained compared to when the instruments were used alone. Le Couteur et al. (2008) also found that combining information from both the ADI-R and ADOS for preschoolers provided a greater level of diagnostic clarity than when each instrument was used in isolation.

Even though the past studies have shown the enhanced diagnostic validity when information from both the ADI-R and ADOS are used together, there has been no systematic attempt to examine the combined use of these instruments for toddlers and young preschoolers using newly revised and developed algorithms for the ADI-R and ADOS. Thus, the aim of the second study was to systematically evaluate ways to combine information from parent interviews and clinician observations using the new ADI-R algorithms for toddlers and young preschoolers (Kim & Lord, in press), revised ADOS algorithms (Gotahm et al., 2007), and ADOS-Toddler algorithms (Luyster et al., 2009) for toddlers and young preschoolers with ASD.

An assessment of language is another crucial part of identifying and describing behaviors of children with ASD. A wide range of verbal abilities from being nonverbal to verbally fluent accompany ASD. Language level affects how ASD symptoms are manifested and the severity of impairment. Specific patterns of language impairment such as echolalia, pronoun reversal, and odd intonation have been found to be associated with ASD (Tager-Flusberg, Paul & Lord, 2005). A subset of children with ASD shows features of specific language impairment (SLI; shorter utterances, more variable use of word endings, articulation problems; Leyfer, Tager-Flusberg, Dowd, Tomblin, & Folstein, 2008; Rice, Wexler, & Cleave, 1995). Gotham et al. (2005) also found that many of the social communicative behaviors measured by the ADOS, such as the frequency of gestures, were strongly associated with a child's language level. Thus, language level should be considered as an important factor for the assessment of ASD symptoms, even though language impairment is neither necessary nor sufficient for a diagnosis of ASD. In addition, because expressive language skills are one of the most important variables predicting later outcomes, most interventions target spoken language acquisition as a main component of treatment outcome studies. Thus, the development of appropriate language measures that can be used to evaluate the efficacy of interventions is crucial. Recognizing these, the third study focuses on developing a new language measure for children from 2 to 12 years of age, the Observation of Spontaneous Expressive Language (OSEL).

Overall, this dissertation is focused on developing and refining instruments and methods for the diagnostic and language assessment of young children with ASD. The newly developed and identified diagnostic algorithms and methods to combine information from clinician observations and parent interviews for toddlers and preschoolers will aid in early identification and provision of treatment for these young children. The new language measure will allow clinicians and researchers to describe the current level of language and quantify language impairments in relation to autism symptoms for children with ASD. These newly developed and improved diagnostic and language measures will provide useful information for treatment and education programs promoting better outcomes for young children with ASD.

References

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

Asperger, H. (1944). Die "autistischen Psychopathen" im Kindesalter. *Archiv fur Psychiatrie*
und Nervenkrakheiten, 117, 76-136.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Cambridge University Press.

Carter, A. S., Davis, N. O., Klin, A., & Volkmar, F. R. (2005). Social development in autism. In F. R. Volkmar, R. Paul, A. Klin, & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders: Vol. 1. Diagnosis, development, neurobiology, and behavior.* Hoboken, NJ: John Wiley & Sons.

Charman, T., Baron-Cohet, S., Swettenham, J., Cox, A., Baird, G., & Drew, A. 1998, An
experimental investigation of social-cognitive abilities in infants with autism: clinical implications. *Infant Mental Health Journal, 19*, 260–275.

Fenson L, Dale PS, Reznick JS, Bates E, Thal DJ, Pethick SJ. Variability in early communicative development. *Monographs of the Society for Research in Child Development, 59*(5), i-185

Fombonne, E. (2007) ‘Epidemiological surveys of pervasive developmental disorders’, in F. Volkmar (ed.), *Autism and Pervasive Developmental Disorders*. 2nd edn. New York, Cambridge University Press. pp 33-68.

Fombonne, E. (2008) ‘Thimerosal disappears but autism remains’, *Archive of General Psychiatry, 65*(1): 15-16.

Fombonne E. (2009) ‘Epidemiology of pervasive developmental disorders’, *Pediatric Research, 65*(6): 591-8.

Goldfield, A.B. (1987). The contributions of child and caregiver to referential and expressive language. *Applied Psycholinguistics, 8,* 267-280.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The autism diagnostic observation schedule (ADOS): Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders, 37*(4), 613–627.

Leyfer, O.T., Tager-Flusberg, H., Dowd, M., Tomblin, J.B., & Folstein, S.E. (2008). Overlap between autism and specific language impairment:

comparison of Autism Diagnostic Interview and Autism Diagnostic Observation Schedule scores. *Autism Research, 1*(5):284-96.

Lord, C., Luyster. R, Gotham, K., & Guthrie, W.J. (in press). *Autism Diagnostic Observation Schedule – Toddler Module.* Los Angeles, CA: Western Psychological Services.

Lord, C., Pickles, A., DiLavore, P. C., & Shulman, C. (1996). *Longitudinal studies of young children referred for possible autism*. Paper presented at the biannual meeting of the International Society for Research in Child and Adolescent Psychopathology, Los Angeles.

Lord, C., Storoschuk, S., Rutter, M., & Pickles, A. (1993). Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal, 14*(3), 234-252.

Mundy, P., Hogan, A., & Doehring, P. (1996). *A preliminary manual for the abridged Early Social-Communication Scales.* from www.psy.miami.edu/ faculty/pmundy.

National Research Council (2001). *Educating children with autism.* Washington, DC: National Academy Press.

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

Rice, M.L., Wexler, K., & Cleave, P. L. (1995). Specific language impairment as a period of extended optional infinitive. *Journal of Speech and Hearing Research, 38*(4), 850-63.

Rimland, B. (1964) *Infantile Autism: The Syndrome and its Implications for a Neural Theory of Behavior.* East Norwalk, CT, US: Appleton-Century-Crofts.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(9), 1094-1103.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised*. Los Angeles: Western Psychological Services.

Rutter, M. and Lockyer, L. (1967) A five to fifteen year follow-up study of infantile psychosis. I. Description of sample. *British Journal of Psychiatry, 113*(504): 1169-82.

Rutter, M. and Schopler, E. (1978) *Autism: A Reappraisal of Concepts and Treatment.* New York, Plenum.

Rice, M.L., Oetting, J., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research, 37*, 106-122.

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

Tager-Flusberg H, Paul R, Lord CE. Language and communication in autism. In: Volkmar F, Paul R, Klin A, Cohen DJ, editors. *Handbook of autism and pervasive developmental disorder.* 3rd ed. Vol. 1. New York: Wiley; 2005. pp. 335–364.

Volkmar, F. R., Klin, A., Siegel, B., Szatmari, P., Lord, C., Campbell, M., et al. (1994). Field trial for autistic disorder in DSM-IV. *American Journal of Psychiatry, 151,* 1361-1367.

Wetherby, A.M., & Prizant, B.M. (2002). *Communication and Symbolic Behavior Scales Developmental Profile.* Baltimore: Paul H. Brookes.

Wiggins, L. D. & Robins, D. L. (2008). Excluding the ADI-R behavioral domain improves diagnostic agreement in toddlers. *Journal of Autism and Developmental Disorders, 38*(5), 972-976.

Williams White, S., Koenig, K., & Scahill, L. (2007). Social skills development in children with autism spectrum disorders: A review of the intervention research. *Journal of Autism and Developmental Disorders, 37*, 1858-1868.

Wing, L. and Gould, J. (1979) Severe impairments of social interaction and associated abnormalities in children: epidemiology and classification. *Journal of Autism and Developmental Disorders, 9,* 11–29.

Chapter II

New Autism Diagnostic Interview-Revised (ADI-R) Algorithms for Toddlers and Young Preschoolers from 12 to 47 Months of Age

The Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) is a standardized, semistructured, investigator-based interview for parents or caregivers of individuals referred for a possible Autism Spectrum Disorder (ASD). The ADI-R includes 93 items in three domains of functioning – language/communication; reciprocal social interactions; and restricted, repetitive, and stereotyped behaviors and interests, as well as other aspects of behaviors. Up to 42 of the interview items are systematically combined to produce a formal, diagnostic algorithm for autism (Rutter, Le Couteur, & Lord, 2003) based on the ICD-10 (World Health Organization, 1992) and DSM-IV (American Psychiatric Association; APA, 1994) definitions of autism as specified by the authors. Other criteria such as using lower cutoffs with the same set of items have been used to create an algorithm for broader classification of autism spectrum disorders (ASD) as used in several collaborative studies (Dawson, Webb, Carver, Panagiotides, & McParland, 2004; Risi et al., 2006). Previous analyses suggested that the diagnostic algorithm was useful for children with a non-verbal mental age above 2 years (Lord et al., 1994). Because most toddlers and preschool children with ASD are not yet at this level of skill, the ADI-R algorithm has not been appropriate to characterize very young children with severe delays (Ventola et al., 2006).

A 'Toddler' version of the ADI-R was developed several years ago to provide descriptive data to be used for research purposes with children under 4 years of age. It includes 32 new questions and codings about the onset of autism symptoms and general development with a total 125 items. Other items in both versions of the ADI-R are identical except that the Toddler ADI-R does not have codes for behaviors between 4 and 5 years of age (referred to as *most abnormal 4 to 5*). No diagnostic algorithm was generated for the toddler version of the ADI-R.

Because of the belief that earlier provision of services and treatments is associated with better outcomes, in the past few years, research has flourished concerning detection of ASD symptoms in the first 2 years of life (National Research Council, 2001). In recent studies, the average age of first parental concern was between 15 and 18 months (Chawarska et al., 2007; DeGiacomo & Fombonne, 1998). Advocacy and funding agencies have also joined together to promote the study of infant siblings of children with autism and other very young children at risk for ASD as seen in the establishment of the Baby Siblings Research Consortium (Yirmiya and Ozonoff, 2007). Thus, researchers and clinicians recognize the increasing need for diagnostic measures that are appropriate for toddlers to capture the early signs of autism at such young ages. For example, the Toddler ADI-R was used to study the parental recognition of developmental problems in toddlers with ASD (Chawarska et al., 2007). Lord, Shulman, and DiLavore, (2004) examined regression and word loss in toddlers with ASD using the Toddler ADI.

Another study focused on restricted and repetitive behaviors in young children with ASD based on the Toddler ADI-R in addition to other measures (Richler, Bishop, Kleinke, & Lord, 2007). The purpose of the present study is to propose a first set of diagnostic algorithms for toddlers and young preschool children developed on a sample of children whose ages ranged from 12 to 47 months with nonverbal mental ages down to 10 months. Although the initial intent was to refine the existing toddler ADI-R into a specific instrument for toddlers, in creating algorithms specifically for young children, priority was given to items that overlapped between the toddler and standard versions of the ADI-R because of the wider availability of the standard ADI-R.

The published algorithms for the ADI-R include a *current behavior algorithm* form (as distinguished from an empirically supported diagnostic algorithm) for children whose ages range from 2 years, 0 months to 3 years, 11 months. Age 4 is a natural dividing point because the standard ADI-R contains questions about children's behavior between age of 4 and 5 (48-59 months) that are not applicable to younger children. The list of items on this form has been used to describe toddlers whose caregivers were administered either the Toddler or standard version of the ADI-R (Wiggins & Robbins, 2008). However, sensitivity and specificity of this list of items with very young children have not yet been carefully examined. In fact, the study that provided the psychometric properties of the existing ADI-R algorithms was based on a sample of children from 36 to 59 months of age, with mental ages ranging from 21 to 74 months (Lord, Rutter, & Le Couteur, 1994). Using the existing algorithms, the group of children with autism over 36 months of chronological age was well differentiated from children with nonspectrum disorders showing high sensitivity and specificity (both over .90). Further analyses of data from preschoolers revealed that the ADI-R algorithms significantly differentiated between children over 2 years with ASD from other developmental disorders. However, discrimination between nonverbal children with ASD and nonverbal children without ASD under 2 years of age, especially for those with mental ages under 18 months was poor, resulting in low specificity (Lord, Storoschuk, Rutter, & Pickles, 1993). Analyzing a larger sample, Risi et al. (2006) also showed high sensitivity (above 80%) of the ADI-R for the classification of children with ASD under 3 years of age, but lower specificity for these children in the comparison of ASD versus nonspectrum disorders (around 70 %).

In contrast, Ventola et al. (2006) reported that the algorithm for the ADI-R resulted in lower sensitivity when compared with Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2000), Childhood Autism Rating Scale (CARS; Schopler, Reichler, & Renner, 1980), and clinical judgment using DSM-IV criteria for children whose chronological age ranged from 16 to 30 months. In the study by Ventola et al. (2006), parents did not report sufficient Repetitive and Restrictive Behaviors (RRBs) to meet the criteria for the RRB domain on the ADI-R. Wiggins and Robins (2008) also found that ADI-R algorithms resulted in poor sensitivity for children between 16 to 37 months of age when the standard cutoff for the RRB domain was included in the diagnostic criteria. The authors suggested that, although RRBs are common in children with ASD, the RRB domain is not as relevant to the ADI-R classification as other diagnostic criteria when evaluating very young children since these behaviors are less often reported by their parents. Thus, despite the relatively good sensitivity and specificity of the ADI-R diagnostic algorithms for older preschool and school age children, there is a need to develop diagnostic algorithms with improved sensitivity and specificity that can differentiate toddlers and young preschoolers with ASD from children with nonspectrum disorders.

Recently, large datasets from several collaborative efforts were used to generate improved algorithms for another autism diagnostic instrument, the Autism Diagnostic Observation Schedule (Lord et al., 2000). In the course of these analyses, it was found that the use of different algorithms for subsets of children defined by age and language level reduced the effects of age, language level, and IQ and increased the sensitivity and specificity of the measure (Gotham, Risi, Pickles, & Lord, 2007). Consequently, we used a similar strategy in developing the new ADI-R algorithms for toddlers and young preschoolers. In addition, because recent studies have shown that classifications within ASD, that is, diagnoses of autism versus Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS) are relatively unstable in contrast to overall diagnoses of ASD versus other nonspectrum disorders that are consistent over time (Lord et al., 2006; Turner & Stone, 2006; Kleinman et al., 2008) and in keeping with the proposed DSM-V ASD criteria (www.dsm5.org), the new algorithms include only two classifications: ASD or nonspectrum.

In order to formally acknowledge the less clear stability of diagnoses in younger children, as we had for the ADOS-Toddler Module (ADOS-T; Lord et al., in press; Luyster et al., 2009), in the present paper, we propose *ranges of concerns* to be used as the primary outcome of the algorithms following the structure of the ADOS-T algorithms. The ranges can be used by clinicians for clinical monitoring and follow-up (rather than yes/no cutoffs for ASD), reflecting little-to-no, mild-to-moderate, or moderate-to-severe concern. However, reflecting the necessity in some research and clinical settings for more strictly stratified groupings, two cutoff scores, one for research (more restrictive; higher specificity with lower sensitivity) and one for clinical purposes (more inclusive; higher sensitivity with lower specificity), were selected for each algorithm. These alternatives allow researchers and clinicians to be transparent about the choices they make, recognizing that diagnostic decisions about ASD in very young children are less stable and precise than for older children and adolescents.

As with other diagnostic measures, the algorithm scores are not intended to be used alone in generating a diagnosis. The assumption is that both clinicians and researchers will make diagnostic decisions taking into account children's developmental history, clinical observations, cognitive skills, and social, language and adaptive functioning in various contexts (Lord & Corsello, 2005). The ADI-R can be also used to describe characteristics of ASD in individual children. Theoretically generated and empirically derived factors are used to organize autism symptoms into different domains. Reliability at the item level allows researchers the opportunity to develop different constructs to study behavioral differences, as well as clinicians the opportunity to describe a child's strengths and weaknesses.

In order to examine the domain structure of autism symptoms included in the ADI-R for children under 4 years of age, we performed factor analyses using data obtained from toddlers and preschoolers. The existing diagnostic algorithms for the ADI-R capture abnormalities in 3 symptom domains, *Reciprocal Social Interaction, Communication*, and *Restricted and Repetitive Behaviors*, which is consistent with the current framework of the DSM-IV. Several recent studies have found other organizations of autism symptom domains in the ADI-R. The content areas that comprise these domains have varied across different samples and analytic techniques. One study found two factors, *social and communication* and *restricted and repetitive behaviors* using confirmatory and exploratory factor analyses based on the data from the ADI-R for children from 4 to 18 years of age (Snow, Lecavalier, & Houts, 2008). Another study found that a three-factor solution (*peer interaction and imaginative play*, *social and communication*, and *stereotyped behavior/restricted interests*) fit better than one- or two- factor solutions based on the data from the ADI-R for children and adolescents from 4 to 20 years of age (Van Lang et al., 2006). The present paper uses data from toddlers and young preschoolers to replicate the results of the past studies. Factor analyses are used to identify new algorithm domain scores that can represent patterns of autism symptoms as well as to generate guidelines for combining domain scores in ways that maximize sensitivity and specificity.

Methods

Participants

Analyses were conducted on 695 different *participants*. Some participants had repeated assessments yielding a total of 829 *cases* (each case was defined by complete data of an ADI-R, contemporaneous ADOS, nonverbal IQ, and best estimate clinical diagnosis). The majority of participants were recruited from two projects, Early Diagnosis of Autism (EDX) and First Words and Toddlers (FW/T) at the University of Michigan Autism and Communication Disorders Center (UMACC) and University of Chicago Developmental Disorders Clinic. The rest of participants were school- or physician-referred clinic patients at UMACC. Children in the EDX project were assessed at ages 2 and 3. Children in the FW/T were assessed at entry to the study (approximately at 2 years of age) and at 3.

The sample was limited to participants aged 12 to 47 months with a nonverbal mental age of at least 10 months. Out of 695 participants, 535 were males. The mean age of the participants was around 33 months (Mean (*M*) = 33.3, Standard Deviation (*SD*) = 9.4). The dataset included 491 participants with ASD, 136 participants with non-spectrum disorders (NS), and 67 typically developing participants. Non-spectrum participants had a range of diagnoses, including language disorders (37 %), nonspecific developmental delays (cases with either the nonverbal or verbal IQ scores more than one standard deviation below average; 21%), non-specific intellectual disability (cases with the full scale IQ score less than 70 with significant delays in adaptive skills; 16 %), and behavioral disorders (8 %). Ethnicity was not associated with diagnosis; 78% of participants were Caucasian, 16% African American, 2.5% Asian American, 2.5% biracial, and about 1% of participants were Native Americans or others.

We created three developmental cells to obtain more homogeneous groups to reduce the effect of language level and age. These groups were defined by age and verbal status during the assessment. The ‘‘phrase speech (PH)’’ group included children whose scores on the item in the ADI-R, ‘‘Overall Level of Language” were ‘0’ = spontaneous language that involves phrases of three words or more including verbs. The “single words (SW)” group included children who scored ‘1’ = use of speech that involves at least five different spontaneous, meaningful words used daily. The ‘‘nonverbal (NV)’’ group included children whose scores on this item were either ‘2’ = fewer than five words total used on a daily basis or ‘3’ = no words used in a daily basis. Consistent with the past finding by Luyster et al. (2009) based on the ADOS-T scores, we found that the ADI-R score distributions of younger children (from 12-20 months, 30 days) and older nonverbal children (from 21-47 months, 30 days) were similar to each other. In contrast, children 21 months or older who had single words or phrase speech differed in the score distributions. Thus, the developmental groups were assigned as follows: (1) all children 12 to 20 months, 30 days of age as well as nonverbal children 21 to 47 months, 30 days of age (hereafter referred to as ‘‘12–20/NV21–47’’); (2) children 21 to 47 months, 30 days of age with single words (“SW21-47”); (3) all children 21 to 47 months, 30 days of age with phrase speech (‘‘PH21–47’’).

Some analyses (factor analyses, logistic regression analyses) were run for data from all 829 assessments (refer to as “*all cases*”) in order to take advantage of the larger sample size afforded by including repeated measurements. However, in order to eliminate effects of repeated participation on domain totals and algorithm totals, the majority of analyses (ROC curves, correlations, and internal consistency) were run with “*unique cases,*” a dataset in which each participant was represented only once per group. Mean ages, NVIQ scores, and gender distributions varied by diagnosis and developmental cells for both *all cases* and *unique cases*. In the *unique cases* groups, the results from the one-way ANOVA indicated that nonverbal IQ differed significantly across three diagnostic groups for "12-20/NV21-47" and "PH21-47" groups with typically developing children showing the highest mean NVIQ scores ($p < .001$). Typically developing children were younger than children with ASD and NS ($p < .001$). Therefore, we controlled for age and IQ scores in all analyses. As expected, the nonverbal group of children between 21 and 47 months of age only included children with ASD and NS because there were no nonverbal typically developing children at this age. Similar patterns were examined for the *all cases* sample. See Table 2.1 for more details on mean ages, NVIQ scores, and gender distributions by diagnosis and developmental cells for *all cases*.

Measures

Either the toddler or standard version of the ADI-R was administered by a clinical psychologist or a trainee who had completed research training and met standard requirements for research reliability (See Risi et al., 2006). A developmental hierarchy of psychometric measures, most frequently the Mullen Scales of Early Learning (MSEL; Mullen, 1995; 52% of participants received the MSEL) and the Differential Abilities Scale (DAS; Elliott, 1990) were used to determine IQ scores. For the MSEL, ratio IQ scores were used as estimates of ability in the present study as in many past studies (e.g. Richler et al., 2007) based on the finding that the ratio IQ scores derived from the MSEL using the age equivalents were found to have a good convergent validity with the DAS (Bishop, Guthrie, Coffing, & Lord, 2011). The MESL ratio IQs were calculated by averaging the age equivalents of all subtests to obtain mental age, and then dividing mental age by chronological age and multiplying by 100. Cognitive testing generally took place immediately before the ADOS administration within the same week as the ADI-R. The ADOS-T (Lord et al., in press), Module 1 or 2 of the ADOS (Lord et al. 1999), or Pre-Linguistic ADOS (PL-ADOS; DiLavore, Rutter & Lord, 1995) was administered depending on the age and developmental level of the child. Out of 829 cases, 294 cases were administered the PL-ADOS because they were assessed before the ADOS modules 1 and 2 were developed.

A Consensus Best Estimate Diagnosis

For children in the EDX study, an experienced clinical researcher used the ADOS and ADI-R scores and observations made during the testing to generate independent best estimate diagnoses of autism, PDD-NOS, and nonspectrum disorders (APA, 1994; See Lord et al., 2006). For children in the FW/T project, scores on the ADI-R, ADOS, and clinical observations were used by two clinicians to make a best estimate diagnosis operationalizing DSM-IV criteria for autism and PDD-NOS (APA, 1994; See Luyster et al., 2009). For clinic cases, a diagnosis was made by a psychologist and/or psychiatrist after review of all data mentioned above.

Design and Analyses

First, we replicated the previously developed algorithm in order to be able to compare the predictive validity obtained from the previous algorithm to the new algorithms. For the development of the new algorithms, the first step was to examine distributions for all of the items in toddler and standard version of ADI-R and identify those that best differentiated ASD and NS diagnoses (referred to as "best items"). Most items in the ADI-R are coded with *no definite behavior of the type specified* (score of 0), *behavior of the type specified probably present but defining criteria not fully met* (score of 1), and *definite abnormal behavior of the type described in the definition and coding* (score of 2), with a score of 3 used occasionally to indicate *extreme severity* (Rutter, Le Couteur & Lord, 2003). Scores of 2 and 3 are collapsed to "2" in the algorithm. The intended selection criteria for the best items were that no more than 20% of ASD cases scored a zero on an item, and no more than 20% of non-spectrum cases scored a 2 or 3. For some items that were theoretically important, these criteria were modified to include items with slightly less clear distributions for certain developmental cells (i.e. *hand and finger mannerisms, compulsions and rituals* for "PH21-47" group). ANOVAs were also used to confirm the results of the diagnostic differences. Items that did not differentiate between diagnoses according to the ANOVAs were excluded from the final sets. Correlations were generated between items and participant characteristics such as age and IQ separately for each developmental cell, and items that were highly correlated with participant characteristics were not included in the final set of items. In addition, when items were highly correlated with each other (above an r of .5), the item that better differentiated ASD versus NS was selected.

After we selected the items that best differentiated between diagnoses, we performed exploratory factor analyses separately for each of the three developmental cells. Scores of 2 and 3 were collapsed and scores of 7, 8 and 9 were treated as missing data and excluded. Because the ADI-R data are ordinal and scores do not represent equal intervals, analyses were run as ordinal probit item response models using Mplus Version 5.21 software. Geomin oblique rotations were used to examine the pattern of factor loadings, and some items were added and deleted based on the loadings to maximize the model fit. Confirmatory factor analyses were next used to verify goodness-of-fit for the factor structure. Then, we used logistic regression to examine the weighting of the three domains created from factor analyses in view of the relative predictive value of scores from the different factors. Item correlations for each individual item with domain-totals-minus-the-item and the correlations between the domains and participant characteristics were examined. ROC curves were calculated to examine the sensitivity and specificity of the selected cutoff scores.

Results

Replication of Sensitivity and Specificity of the Previous Algorithm

We first used ROC curves (Siegel, Vukiceciv, Elliot, & Kraemer, 1989) to examine sensitivity and specificity of the previous *current behavior algorithm* in order to compare the predictive validity of the new diagnostic algorithms to application of the previous algorithm (Rutter, LeCouteur & Lord, 2003). For the younger or nonverbal group ("12-20/NV21-47"), the previous algorithm yielded excellent sensitivity (97%) but very poor specificity (43%) for ASD versus NS. When typically developing children were combined with the NS group, specificity was 47 %.

For the SW21-47 group, the original algorithm resulted in 91% sensitivity and 82% specificity when children with ASD were compared to NS. For the PH21-47 group, the existing *current behavioral algorithm* showed 70% sensitivity and 68% specificity for the comparison of ASD versus NS. Specificity improved slightly to 70% when the TD group was included. Thus, the biggest challenges in discriminating children with ASD were the youngest children, nonverbal children with the most significant delays ("12-20/NV21-47"), and preschool children with more advanced verbal abilities ("PH21-47").

Development of the New Algorithms

Best Items.

Items were selected using the criteria mentioned above for each of the three developmental cells defined by age ranges and language level. Items from both the standard and toddler ADI-R were considered. Initial analyses determined that items that appeared in both measures were consistently more informative than items in only one of the measures. Thus, further analyses were restricted to items used in both the standard and toddler versions. All items in the ADI-R are coded in terms of whether the behavior is "currently" occurring (behavior occurred within the past 3 months; referred as the "current" item), and whether it "ever" occurred (referred as the "ever" item, which also includes current behavior). Selected items included a combination of *current* and *ever* items.

Factor Analyses.

Results from the factor analyses showed that 3-factor solutions fitted well for all three developmental cells (Table 2.2). For the "12-20/NV21-47" and "SW21-47" groups, items loaded onto three factors, Social Affect (SA), Restricted and Repetitive Behaviors (RRB), and Imitation, Gestures, and Play (IGP). For the "PH21-47" group, items associated with SA and IGP loaded onto a first factor, also referred to as "Social Communication (SC)," and items associated with RRBs comprised a second factor. A third factor emerged for the "PH21-47" group comprised of items related to "Reciprocal and Peer Interaction (RPI)." In most cases, a loading on the exploratory factor analyses (EFA) of .4 was set as a threshold for the inclusion criteria to include in the algorithms except for several items that were theoretically important. These items had high specificity though they were not common across all diagnostic groups (e.g. *ever: unusual preoccupation* for the "SW21-47" algorithm, *current: use of other's body to communicate* for the "PH21-47" algorithm; See Table 2.2 for more details). Although the loadings for these items were lower than the threshold in the EFA, confirmatory factor analyses showed notably higher loadings. All of the RRB items selected were *ever* items because *ever* RRB items consistently showed higher sensitivity and specificity than *current* RRB items. The one exception was *current*: *stereotyped language* for children with phrase speech, which differentiated diagnoses better than *ever: stereotyped language,* consistent with the results from the past study using the original ADI-R (Rutter et al., 2003). All non-RRB items in the proposed algorithms are *current* items.

Confirmatory factor analyses (CFA) were performed to examine the model fit for each group, and the results showed a 3-factor model to fit substantially better than 1-, and 2- factor models for all three groups. In the "12-20/NV21-47" group, the goodness-of-fit rating yielded a Comparative Fit Index (CFI) of 0.952 (CFI between 0.9 and 1 indicating good fit; Skrondal & Rabe-Hesketh, 2004) and Root Mean Square Error Approximation (RMSEA) of 0.069 (RMSEA of 0.08 or less is considered a satisfactory fit; Browne & Cudeck, 1993).

For the "SW21-47" group, two items, *current: interest in other children* and *current: response to approach of other children*, were not initially included in the EFA. However, because these items emerged in both other algorithms, they were added in the CFA to the first factor, SA, to maximize the number of algorithm items directly comparable across the three age by language groupings. The addition of these items in the first factor also increased the predictive validity of the algorithm resulting in higher sensitivity and specificity. The model fit from the confirmatory factor analyses showed satisfactory fit even after these changes were made (CFI of 0.943 and RMSEA of 0.062).

For the "PH21-47" group, the goodness-of-fit rating was satisfactory with CFI of 0.952 and RMSEA of 0.057. In addition, for all three groups, the item, *current: inappropriate facial expression*, always loaded onto the RRB domain; however, we placed the item into the first domain of the algorithm because it made more sense theoretically.

Logistic Regression Check on the Prediction of Diagnosis for Each Domain.

In order to describe each child's individual profile, item scores are totaled for each domain. Item scores that were most predictive of diagnosis based on logistic regressions are then combined to yield a single cutoff for categorical classification of ASD. Because factor scores were not uniformly better at prediction of diagnosis than simple totals for each domain, simple item totals for each factor were used in logistic regression analyses, predicting clinical diagnoses of ASD as the outcome. Logistic regression for ASD versus nonspectrum disorders (NS) indicated that among the domains, the SA domain significantly predicted diagnosis when IQ scores, chronological age, and the two other factors were controlled for the younger and older nonverbal group and the older children with single words ($\beta = .17$, $p < .001$ for "12-20/NV21-47"; $\beta = .402$, $p < .001$ for "SW21-47"). The RRB domain was also a significant predictor of diagnosis while controlling for the other factors for both groups ($\beta = .337$, $p < .001$ for "12-20/NV21-47"; $\beta = .236$, $p < .05$ for "SW21-47").

In contrast, Imitation, Gestures, and Play (IGP) did not predict diagnosis for ASD versus NS when the other two factors and participant characteristics were controlled. For this reason, the final algorithm cutoff score for the "12-20/NV21-47" and "SW21-47" algorithms included only the two domains, SA and RRBs, though the third factor will remain on algorithm forms in order to allow quantification of these deficits.

Item totals using SA and RRB factors were combined, and the combined scores also predicted diagnosis when the third factor, IGP, and participant characteristics were controlled (β = .232, p < .001 for "12-20/NV21-47"; β = .327, p < .001 for "SW21-47"). For the children with phrase speech, each domain significantly predicted diagnosis while controlling for the other two factors as well as IQ scores and age (β = .148, p < .05 for SC; β = .312, p < .05 for RRBs; β = 19, p < .05 for RPI). Thus, all three factors (SC, RRBs, and RPI) were included when generating diagnostic thresholds for the "PH21-47" algorithm.

For all three algorithms, all domain scores included in the algorithms for *unique cases* were significantly higher for the ASD sample than the nonspectrum or typically developing groups (See Table 2.3).

Correlations between Domain Totals and Participant Characteristics.

Correlations between domain totals and participant characteristics were examined using *unique cases* only, in order to eliminate effects of repeated participation on domain totals. For the "12-20/NV21-47" group, most of the correlations between the domain totals and participant characteristics did not exceed an r of .5. Nevertheless, for the "12-20/NV21-47" group, the correlation between verbal IQ and the IGP domain was higher (r = -.52, p < .001), which is one of the reasons why it is not included in the diagnostic algorithm. For the "SW21-47" group, most of the correlations between the domain totals and participant characteristics did not exceed an r of .4. For the "PH21-47" group, no correlations exceeded an r of .4.

Internal Consistency of Algorithm.

Internal consistency of each algorithm was examined through correlations between item and domain-total-minus-item correlations (item-total correlation) and Cronbach's Alpha using *unique cases*. The item, *current: inappropriate facial expression*, always loaded onto the RRB domain. As expected, because we placed the item into the SA domain, the item-total correlation for the particular item was often the lowest among all the items in the same domain. For the younger or nonverbal children receiving the "12-20/NV21-47" algorithm, item-total correlations ranged from .65 (*current: inappropriate facial expression*) to .81 (*current: social smiling*) for the SA domain; from .71 (*ever: hand and finger mannerisms*) to .77 (*ever: repetitive use of objects*) for the RRB domain; and from .64 (*current: imaginative play*) to .82 (*current: instrumental gestures*) for the IGP domain. Cronbach's alpha was .9 for the SA domain; .73 for the RRB domain, and .87 for the IGP domain, indicating strong internal consistency (Cronbach, 1951).

For the older children with single words receiving the "SW21-47" algorithm, item-total correlations for *unique cases* ranged from .61 (*current: inappropriate facial expression*) to .73 (*current: direct gaze*) for the SA domain; from .49 (*ever: unusual preoccupation*) to .69 (*ever: repetitive use of objects*) for the RRB domain, and from .6 (*current: offering to share*) to .78 (*current: instrumental gestures*) for the IGP domain. Cronbach's alpha showed strong internal consistency for the items in each domain (.85 for the SA domain; .62 for the RRB; .74 for the IGP).

For the children with phrase speech receiving "PH21-47" algorithm, item-total correlations from .32 (*current: use of other's body to communicate*) to .7 (*current: quality of social overtures*) for the SC domain; from .62 (*ever: hand and finger mannerisms*) to .76 (*current: stereotyped language*) for the RRBs domain; from .81 (*current: appropriateness of social response*) to .86 (*current: interest in other children*) for the RPI domain. Cronbach's alpha showed strong internal consistency for all items in each domain (.83 for the SC domain; .79 for the RRB; .72 for the RPI).

Ranges of Concern.

Recognizing that diagnoses of ASD in very young children may be less stable than diagnoses at older ages, *ranges of concern* were identified for all three algorithms to be used for clinical purposes. Three *ranges of concern* were set for each algorithm such that at least 80% of children with ASD and no more than about 5% of children with TD would fall in the two ranges of clinical concern (mild-to-moderate and moderate-to-severe ranges). See Figure 2.1 for results. For all three groups, 67% to 81% of children with NS were accurately assigned to the little-to-no range depending on the developmental group.

Sensitivity and Specificity of the New Algorithms.

We next used ROC curves to generate two sets of cutoffs for the new algorithms; one for clinical purposes with maximum sensitivity and adequate specificity (above 70%) for the comparison of ASD vs. NS and one for research purposes with maximum specificity and adequate sensitivity (above 80%) for the comparison of AUT vs. NS. These cutoffs are tied to the endpoints of the mild-to-moderate range from the *ranges of concern* described above.

The clinical cutoffs yielded sensitivities ranging from 80 to 94% and specificities ranging from 70 to 81% for ASD vs. NS depending on developmental cells. For research cutoffs, sensitivities ranged from 80 to 84% and specificities range from 82 to 90 % for AUT vs. NS. See Table 2.4 for more details.

Comparison of the New Algorithms to the Previous Algorithm

Figure 2.2 shows the significant gains in predictive validity, using the new algorithms compared to the previous algorithm. As intended, the groups that were harder to differentiate using the previous algorithm showed the most predictive improvement when the new algorithms were used: For the younger and older nonverbal children ("12-20/NV21-47"), specificity improved significantly when either clinical or research cutoffs were used compared to the original algorithm even though sensitivity dropped; for the children with phrase speech ("PH21-47"), specificity and sensitivity improved significantly when either clinical or research cutoffs were used compared to the *current behavior algorithm*. For the older children with single words ("SW21-47"), the new and original algorithms were comparable.

Discussion

The algorithms presented in this study are the first algorithms developed on data obtained from toddlers and young preschoolers whose ages ranged from 12 to 47 months. These new algorithms offer theoretically updated and more valid ways of using caregiver information in the diagnosis of young children, while expanding the lowest age of application to 12 months with a lowest nonverbal developmental level of 10 months. Compared to the existing algorithms which contain over 30 items, the new algorithms showed improved predictive validity with fewer items (13-20 items). In particular, the new algorithms showed substantial gains in specificity (37-42%) for the "12-20/NV21-47" group and modest gain in specificity (2-14%) with consistent improvements in sensitivity (10-14%) for the "PH12-47" group.

One of the advantages of the new algorithms for toddlers is that they provide clinicians and researchers with several different options for the diagnostic classification of young children. For clinical purposes, *ranges of concern* are proposed that represent the severity of autism symptoms. Depending on where a child falls in among the three *ranges of concern,* a clinician or a researcher can decide about whether or not the child should be followed up with further assessments and enter into treatment irrespective of diagnostic cutoffs. Scores that fall into the little-to-no range indicate that the child is reported to have no more behaviors associated with ASD than children in the same age range who do not have ASD. On the other hand, a child who scores in the mild-to-moderate range has a number of behaviors consistent but not unique to ASD. For clinical purposes, these children, just as those in the moderate-to-severe range should receive further evaluation and follow-up, including other cognitive and language assessments, and recommendations for treatment.

On the other hand, researchers conducting expensive and time consuming procedures such as neuroimaging may wish to stratify ASD cases in order to more strictly exclude likely NS cases by using the research cutoffs that include only the moderate-to-severe range. In contrast, researchers such as geneticists who are casting a broader net for children with autism-related difficulties and clinicians needing to avoid wrongly denying a child access to services can choose to use the clinical cutoffs. In the past, these cutoffs might have been linked to differences between autism and PDD-NOS, but since it is clear that, in this sample, these differences were quantitative, not qualitative, designating them as *ranges of concern* seems more appropriate.

As found in previous research, results from the present study showed that social and communication items primarily loaded into one factor, the *Social Affect* domain for the younger children and older nonverbal children as well as for children with single words and the *Social Communication* domain for children with phrase speech. These results are consistent with past studies using the ADI-R with older children that have also shown that items associated with social and communication loaded onto a single factor (Frazier et al., 2008; Snow, Lecavalier, & Houts, 2008; Van Lang et al., 2006). In addition, the present study showed that for all of the three developmental groups, a second factor was associated with RRBs. Items were similar across the groups, but children with phrase speech had additional items such as *stereotyped language* due to their advanced verbal abilities. Cronbach's alphas were lower than expected for the RRB domain even though they were all above .7, possibly because the domains encompass a diverse set of behaviors. It is also interesting to note that the item, *inappropriate facial expression,* consistently loaded on this domain, raising the possibility that the domain may not only represent RRBs but also unusual behaviors of other types.

Items associated with *Imitation, Play and Gestures* (IGP) loaded onto a third factor for the first two groups of children with minimal language (nonverbal children and children with single words), not for those with phrase speech. This finding is consistent with a recent study done by Frazier et al. (2008) using the items in the ADI-R in which the authors found a third factor related to *Play*. It is interesting that even though the IGP factor consists of "best items" that differentiated children with ASD from those with NS and typically developing children, it did not differentiate between diagnoses when age, IQ scores, and the other domain scores were covaried. This was why we did not incorporate it into the cutoffs for ASD in the "12-20/NV21-47" and "SW21-47" algorithms along with the *SA* and *RRB* domains. For the children with phrase speech, the third factor was associated with *Reciprocal and Peer Interaction,* also consistent with past studies (Van Lang et al., 2006). The third factor uniquely contributed to diagnostic differentiation, and it was found to be independent of age and NVIQ, which was why it was included in the algorithm total.

With the new algorithms, children do not have to have RRBs as long as they score high enough on the other domain(s) to exceed the cutoffs for ASD. This may partially ease the concern raised in past studies that parents might not report RRBs in very young children (Wiggins & Robbins, 2008). Nevertheless, RRB domain totals were consistently higher for children with ASD than children with NS and TD in all of the three developmental cells. Furthermore, in past studies, RRBs had added to stability of diagnoses over time and diagnostic predictability across measures (Lord et al., 2006; Risi et al., 2006). Thus, all of the domains including the RRB domain clearly contributed to the diagnostic validity of the new algorithms.

The goal of creating algorithms less dependent on age was met relatively easily by dividing the sample into cells by age. Minimizing the effect of nonverbal IQ and language level was more complex but low correlations (below .4) between each domain score and the participant characteristics were maintained by creating different algorithms for different language levels. For the "12-20/NV21-47" group, verbal IQ scores showed a moderate correlation with the IGP domain even after the sample was divided into different language levels. This is one of the reasons why the IGP domain was not included in the diagnostic algorithm even though the domain seems sufficiently important to make it readily available on algorithm forms.

Limitations

Even though we were able to create similar algorithms across the three groups, different thresholds across cells were necessary in order to obtain the best sensitivity and specificity within each developmental cell. This limits the interpretation of data when clinicians and researchers want to measure changes over time because children will move from algorithm to algorithm as they grow older. However, it is not surprising that the algorithms contain slightly different items; some of abnormalities in social interaction and communication as well as RRBs become less or more salient with development. Clinicians can compare items that overlap across algorithms to see the changes in the severity in the specific behaviors measured by each item.

Sensitivity and specificity of the measure may vary in different research samples due to factors such as participant characteristics, socio economic status of the family, and skills of the examiner. In particular, there were few NS children in the 12-20 age group in the present study. However, these children were combined with nonverbal children up to 47 months into the "12-20/NV21-47" group because of similarities in score distributions, and this provided us a sufficient sample size for this group.

Replications across sites with well-defined populations with and without ASD will be critical. For replications, the total scores can be calculated currently by adding scores from the items listed under the first two domains for the "12-20/NV21-47" and "SW21-47" groups and those listed under three domains for the "PH21-47" group (See Table 2.2). Replications will be needed for each of the two different thresholds for ASD (See Table 2.1) as well as *ranges of concern* (See Figure 2.1).

Conclusion

In sum, new ADI-R algorithms presented in this study extend the valid use of the ADI-R to toddlers and young preschoolers ranging from 12-47 months of age and down to nonverbal mental age of 10 months. Algorithms can be used for either the standard or toddler version of the ADI-R. We hope that researchers and clinicians alike find them a useful tool in supporting families and children with ASD to advance our understanding of these conditions through quantifying autism symptom domains at individual and domain levels and along with clinical observations and other information, contributing to the reliable diagnoses of toddlers and young preschoolers with ASD.

Table 2.1 Description of sample of *all cases*

	12-20			NV21-47		SW21-47		PH21-47		
	ASD	NS	TD	ASD	NS	ASD	NS	ASD	NS	TD
N (male, female)	43 (37, 6)	24 (16, 8)	47 (30, 17)	318 (264, 54)	36 (20, 16)	140 (103, 37)	32 (19, 13)	107 (91, 16)	62 (49, 13)	20 (13, 7)
Chronological age										
Mean	17.5	16.7	15.6	34.2	29.8	36	33.6	40.9	38.7	28.4
Standard Deviation	2.2	2.4	2.6	7.2	6.4	6.4	5.9	5.4	5.7	5.2
Minimum-Maximum	13-20	12-20	12-20	21-47	22-43	21-47	22-46	24-47	22-47	21-37
Nonverbal IQ [a]	ASD	NS	TD	ASD	NS	ASD	NS	ASD	NS	TD
Mean	85.3	86.4	111	60.7	64.9	71.4	69.1	87.2	92.1	121
Standard Deviation	20.2	23.9	16.6	20.7	30.9	18.8	18.1	22.2	17.9	17.8
Minimum-Maximum	41-119	38-123	68-150	15-144	17-132	32-122	47-114	31-156	50-129	83-155

[a] Full scale scores were used for 39 children with TD in the 12-20 group, 1 child with NS in the SW21-47 group, 1 child with ASD, 2 children with NS, and 15 children with TD in the PH21-47 group because no nonverbal scale was available. *12-20* Children from 12 to 20 months of age, *NV21-47* Nonverbal children from 21 to 47 months of age, *SW21-47* Children with single words from 21-47 months of age; *PH21-47* Children with phrase speech from 21-47 months of age. *ASD* Autism Spectrum Disorder, *NS* Nonspectrum disorder, *TD* Typical Development.

Table 2.2 Algorithm mapping for groups defined by chronological age and expressive language level

12-20/NV21-47	**Factor Loadings**		**SW21-47**	**Factor Loadings**		**PH21-47**	**Factor Loadings**	
	EFA	**CFA**		**EFA**	**CFA**		**EFA**	**CFA**
Social Affect			**Social Affect**			**Social Communication**		
C. Attention to Voice*	0.69	0.79	C. Attention to Voice*	0.75	0.74	C. Attention to Voice*	0.71	0.81
C. Direct Gaze*	0.76	0.77	C. Direct Gaze*	0.85	0.76	C. Direct Gaze*	0.79	0.74
C. Social Smiling†	0.95	0.84	C. Social Smiling†	0.87	0.75	C. Nodding to mean yes	0.58	0.6
C. Seeking to Share Enjoyment*	0.47	0.85	C. Seeking to Share Enjoyment*	0.78	0.71	C. Seeking to Share Enjoyment*	0.82	0.71
C. Range of Facial Expression*	0.56	0.69	C. Range of Facial Expression*	0.63	0.72	C. Range of Facial Expression*	0.51	0.59
C. Inappropriate Facial Expression†	0.58	0.64	C. Inappropriate Facial Expression†	0.41	0.64	C. Offers Comfort	0.43	0.61
C. Appropriateness of Social Response*	0.87	0.84	C. Appropriateness of Social Response*	0.68	0.69	C. Pointing to Express Interest*	0.59	0.71
C. Interest in Children*	0.91	0.81	C. Interest in Children[a]*	-	0.76	C. Showing and Directing attention*	0.77	0.76
C. Response to Approaches of Children*	0.93	0.77	C. Response to Approaches of Children[a]*	-	0.81	C. Quality of Social Overtures†	0.80	0.76
			C. Quality of Social Overtures†	0.71	0.72	C. Social Chat	0.57	0.73
						C. Use of Other's Body to Communicate	0.32	0.3
Repetitive & Restricted Behaviors			**Repetitive & Restricted Behaviors**			**Repetitive & Restricted Behaviors**		
E. Repetitive Use of Objects*	0.55	0.8	E. Repetitive Use of Objects*	0.53	0.76	C. Stereotyped Language*	0.5	0.84
E. Hand and Finger Mannerisms *	0.46	0.67	E. Hand and Finger Mannerisms *	0.67	0.53	E. Hand and Finger Mannerisms *	0.6	0.47
E. Other Complex Mannerisms*	0.55	0.74	E. Other Complex Mannerisms*	0.64	0.63	E. Other Complex Mannerisms*	0.41	0.61
E. Unusual Sensory Interests*	0.65	0.72	E. Unusual Sensory Interests*	0.34	0.65	E. Unusual Sensory Interests*	0.53	0.71
			E. Unusual Preoccupations†	0.25	0.38	E. Unusual Preoccupations†	0.43	0.62
			E. Compulsions/Rituals†	0.44	0.44	E. Compulsions/Rituals†	0.49	0.64
Imitation, Gestures & Play			*Imitation, Gestures & Play*			**Reciprocal and Peer Interaction**		
*C. Pointing to Express Interest**	*0.72*	*0.86*	*C. Pointing to Express Interest**	*0.59*	*0.71*	C. Appropriateness of Social Response*	0.5	0.89
C. Conventional/Instrumental Gestures†	*0.68*	*0.88*	*C. Conventional/Instrumental Gestures†*	*0.69*	*0.79*	C. Interest in Children*	0.86	0.84
C. Spontaneous Imitation of Actions†	*0.77*	*0.83*	*C. Spontaneous Imitation of Actions†*	*0.84*	*0.78*	C. Response to Approaches of Children*	0.73	0.8
C. Offering to Share†	*0.71*	*0.83*	*C. Offering to Share†*	*0.5*	*0.62*			
C. Imaginative Play†	*0.82*	*0.69*	*C. Imaginative Play†*	*0.58*	*0.58*			
*C. Showing and Directing Attention**	*0.63*	*0.89*						
	EFA	**CFA**		**EFA**	**CFA**		**EFA**	**CFA**
	CFI			**CFI**			**CFI**	
	0.991	0.952		0.988	0.943		0.988	0.96
	RMSEA			**RMSEA**			**RMSEA**	
	0.064	0.069		0.047	0.062		0.054	0.053

[a] Items added only for the Confirmatory Factor Analyses; * Items that overlap across all three algorithms; † **I**tems that overlap across two algorithms. Factors that are *not* included in the algorithm cutoffs are *italicized*. *C* Current; *E* Ever; *12-20/NV21-47* Children from 12-20 months of age and nonverbal children from 21-47 months of age; *SW21-47* Children with single words from 21-47 months of age; *PH21-47* Children with phrase speech from 21-47 months of age; *EFA* Exploratory Factor Analyses; *CFA* Confirmatory Factor Analyses; *CFI* Comparative Fit Index; *RMSEA* Root Mean Square Error Approximation. *ASD* Autism Spectrum Disorder, *NS* Nonspectrum disorder, *TD* Typical Development.

Table 2.3 Mean algorithm domain scores by diagnostic group

Domain	12-20/NV21-47			SW21-47		PH21-47		
	ASD	NS	TD	ASD	NS	ASD	NS	TD
	n = 263	*n* = 47	*n* = 47	*n* = 127	*n* = 30	*n* = 101	*n* = 59	*n* = 20
Social Affect	11.8 (4.18)	5.7 (4.25)	1.9 (1.73)	11 (4.56)	3.4 (2.88)	9.7 (4.42)	5.1 (3.77)	1.4 (1.35)
RRB	4.8 (2.32)	2 (2.01)	0.9 (1.24)	5.5 (2.64)	2.3 (1.66)	5.8 (2.99)	2.9 (2.9)	0.7 (1.09)
RPI[a]	-	-		-	-	3 (1.76)	1.6 (1.5)	1 (1.26)
Total algorithm score	16.5 (5.51)	7.7 (5.25)	2.8 (2.22)	16.6 (6.3)	5.7 (3.72)	18.5 (7)	9.5 (6.71)	3.1 (2.96)

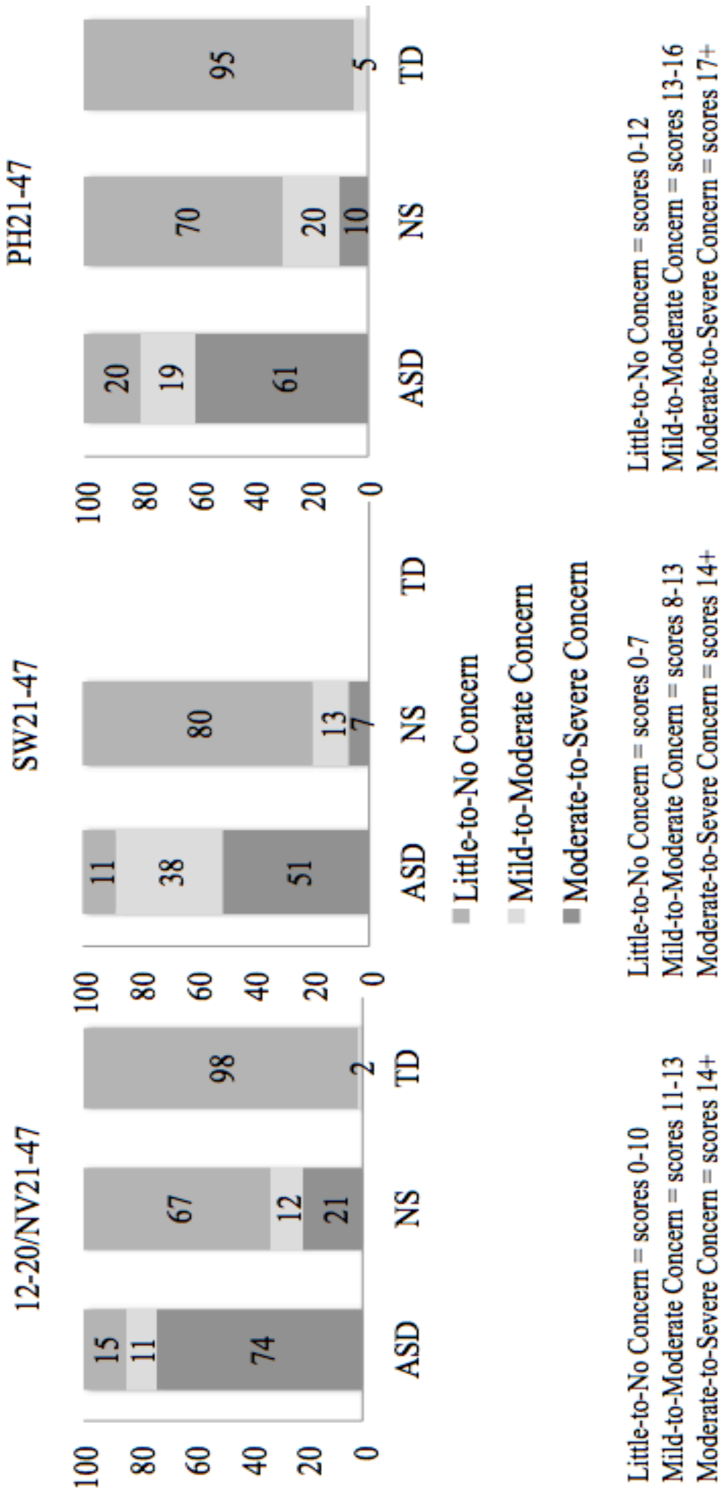
Standard deviations in parentheses; [a] Reciprocal and Peer Interaction domain was included only in the "PH21-47" algorithm; *12-20/NV21-47* Children from 12-20 months of age and nonverbal children from 21-47 months of age; *SW21-47* Children with single words from 21-47 months of age; *PH21-47* Children with phrase speech from 21-47 months of age*;* *ASD* autism spectrum disorder; *NS* non-spectrum disorder; *TD* typical development; *RRB* Restricted and Repetitive Behaviors.

Table 2.4 Sensitivity and specificity of research and clinical cutoffs

		Sensitivity		Specificity	
		AUT	ASD	NS	NS & TD
12-21/NV21-47	Research Cutoff = 13	84	77	**85**	93
	Clinical Cutoff = 11	91	**85**	70	86
SW21-47	Research Cutoff = 13	80	71	**90**	-
	Clinical Cutoff = 8	99	**94**	81	-
PH21-47	Research Cutoff = 16	84	70	**82**	86
	Clinical Cutoff = 13	93	**80**	70	76

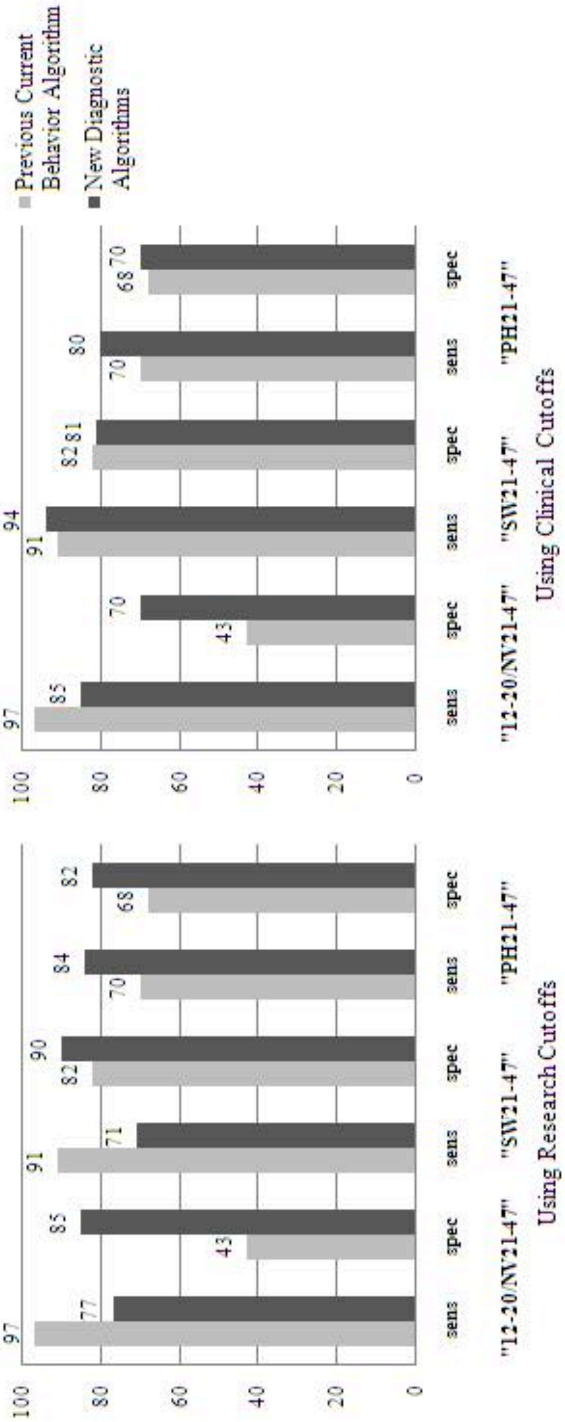
AUT Autism, *ASD* Autism Spectrum Disorder, *NS* Nonspectrum disorder, *TD* Typical Development; *12-20/NV21-47* Children from 12-20 months of age and nonverbal children from 21-47 months of age; *SW21-47* Children with single words from 21-47 months of age; *PH21-47* Children with phrase speech from 21-47 months of age. **Bolded** numbers indicate maximized specificities and sensitivities depending on criteria used in selecting cutoff scores.

Figure 2.1 Percent of participants falling into ranges of concern by diagnostic group



12-20/NV21-47 Children from 12-20 months of age and nonverbal children from 21-47 months of age; *SW21-47* Children with single words from 21-47 months of age; *PH21-47* Children with phrase speech from 21-47 months of age; *ASD* autism spectrum disorder; *NS* non-spectrum disorder; *TD* typical development.

Figure 2.2 Sensitivities and specificities of new diagnostic algorithms (using research and clinical cutoffs) and a previous current behavior algorithm



Sens Sensitivity; *Spec* Specificity; *12-20/NV21-47* Children from 12-20 months of age and nonverbal children from 21-47 months of age; *SW21-47* Children with single words from 21-47 months of age; *PH21-47* Children with phrase speech from 21-47 months of age.

References

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4^{th} ed.). Washington, DC: Author.

Bishop, S., Guthrie, W., Coffing, M., & Lord, C. (2011). Convergent validity of the mullen scales of early learning and the differential ability scales in children with autism spectrum disorders. *American Journal on Intellectual and Developmental Disabilities*, *116*(5), 331-343.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing Structural Equation Models* (pp. 136–162). Newbury Park, CA: Sage.

Chawarska, K., Paul, R., Klin, A., Hannigen, S., Dichtel, L., & Volkmar, F. (2007). Parental recognition of developmental problems in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38* (1), 67-72.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika,* 16, 297–334.

Dawson, G., Webb, S., Carver, L., Panagiotides, H., & McPartland, J. (2004). Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Developmental Science, 7*(3), 340-359.

DeGiacomo, A., & Fombonne, E. (1998). Parental recognition of developmental abnormalities in autism. *European Journal of Child and Adolescent Psychiatry, 7*, 131–136.

DiLavore, P., Lord, C., & Rutter, M. (1995). The pre-linguistic autism diagnostic observation schedule (PL-ADOS). *Journal of Autism and Developmental Disorders, 25*, 355–379.

Elliott, C. D. (1990). *Differential Abilities Scale (DAS).* San Antonio, TX: Psychological Corporation.

Frazier, T., Youngstrom, E., Kubu, C., Sinclair, L., Rezai, A. (2008). Exploratory and confirmatory factor analysis of the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders, 38*(3), 474-480.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The autism diagnostic observation schedule (ADOS): Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders, 37*(4), 613–627.

Kleinman, J. M., Ventola, P. E., Pandey, J., Verbalis, A. D., Barton, M., Hodgson, S., et al. (2008). Diagnostic stability in very young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*(4), 606–615.

Lord, C., & Corsello, C. (2005). Diagnostic instruments in autistic spectrum disorders. In F.R. Volkmar, A. Klin, & R. Paul (Eds.), *Handbook of Autism and Pervasive Developmental Disorders*, 3rd Edition, (pp. 730-771). Hoboken, NJ: John Wiley & Sons, Inc.

Lord, C., Luyster. R, Gotham, K., & Guthrie, W.J. (in press). *Autism Diagnostic Observation Schedule – Toddler Module.* Los Angeles, CA: Western Psychological Services.

Lord, C., Risi, S., DiLavore, P., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from two to nine. *Archives of General Psychiatry, 63*(6), 694–701.

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.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *Autism Diagnostic Observation Schedule: Manual.* Los Angeles: Western Psychological Services.

Lord, C., Storoschuk, S., Rutter, M., & Pickles, A. (1993). Using the ADI-R to diagnose autism in preschool children. *Infant Mental Health Journal, 14*(3), 234-252.

Lord, C., Shulman, C, & DiLavore, P. (2004). Regression and word loss in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry, 45* (5), 936–955.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., et al. (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*, 1305–20.

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

Muthen, L. K., Muthen, B. O. (2007). *M-plus User's Guide, Version 5*. Los Angeles, CA: Muthen and Muthen.

National Research Council (2001). *Educating children with autism*. Washington, DC: National Academy Press.

Richler, J., Bishop, S. L., Kleinke, J. R. & Lord, C. (2007). Restricted and repetitive behaviors in young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*, 73-85.

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(9), 1094.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised*. Los Angeles: Western Psychological Services.

Schopler, E., Reichler, R. J., & Renner, B. R. (1980). *The Childhood Autism Rating Scale*. Los Angeles: Western Psychological Services.

Siegel, B., Vukicevic, J., Elliott, G., & Kraemer, H. (1989). The use of signal detection theory to assess DSM-III-R criteria for autistic disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 28, 542–548.

Snow, A., Lecavalier, L., Houts, C. (2008). The structure of the Autism Diagnostic Interview-Revised: diagnostic and phenotypic implications. *Journal of Child Psychology and Psychiatry, 50*(6), 734-742.

Turner, L., & Stone, W. (2007). Variability in outcome for children with an ASD diagnosis at age 2. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 48*(8), 793–802.

Van Lang, N.D.J., Boomsma, A., Sytema, S., de Bildt, A.A., Kraijer, D.W., Ketelaars, C., & Minderaa, R.B. (2006). Structural equation analysis of a hypothesized symptom model in the autism spectrum. *Journal of Child Psychology and Psychiatry, 47*, 37–44.

Ventola, P. E., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J., Robins, D., & Fein, D. (2006). Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders, 36*(7), 839-47.

World Health Organization. (1990). *International classification of diseases* (10th revision). Geneva: Author.

Wiggins, L. D. & Robins, D. L. (2008). Excluding the ADI-R behavioral domain improves diagnostic agreement in toddlers. *Journal of Autism and Developmental Disorders*, *38*(5), 972-976.

Yirmiya, N. & Ozonoff, S. (2007). The very early autism phenotype. *Journal of Autism and Developmental Disorders, 37* (1), 1-11.

Chapter III

Combining Information from Multiple Sources of Information for the Diagnosis of Autism Spectrum Disorders in Toddlers and Preschoolers from 12 to 47 months of Age

The Autism Diagnostic Interview-Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003) and the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 2001) have been widely used together, particularly for research, and sometimes in clinical settings for individuals who have been referred due to possible autism spectrum disorders (ASD). The ADI-R is a standardized, semi-structured, investigator-based interview for caregivers. The ADOS is a standardized, semi-structured, clinician-administered observation of communication, social interaction, and play. Both instruments provide diagnostic algorithms for autism. The ADOS also includes an algorithm for a broader classification of ASD; an equivalent algorithm for the ADI-R has been used in several collaborative studies (Dawson, Webb, Carver, Panagiotides, & McParland, 2004; Risi et al., 2006) based on the ICD-10 (WHO, 1992) and DSM-IV (APA, 1994).

Combining information from multiple sources across raters and instruments enhances accuracy for the diagnosis of ASD when a best estimate clinical diagnosis is treated as the gold standard. For example, the Social Responsiveness Scale resulted in high diagnostic specificity for children and adolescents with ASD when information from both parent and teacher reports were combined (Constantino et al., 2007). Bishop and Baird (2001) reported improved validity of the Children's Communication Checklist when information from both parents and professionals were used for 151 children with pervasive developmental disorders (PDD) or other developmental disorders between 5 to 17 years of age. Corsello et al. (2007) reported enhanced diagnostic validity by combining information across instruments, either the Social Communication Questionnaire (SCQ; Rutter, Bailey, & Lord, 2003) or the ADI-R with the ADOS for the diagnosis of children with ASD between age 2 and 16 years.

Risi et al. (2006) found a better balance of sensitivity and specificity when the ADI-R and ADOS were used in combination compared to when each instrument was used alone. For example, the combined use of these instruments resulted in sensitivity and specificity of 82% and 86% for children with autism compared to children with non-spectrum disorders over age 3 years. For younger children, sensitivity and specificity for the same diagnostic comparison using both instruments were 81% and 87%, respectively. In contrast, when each instrument was used alone, specificities ranged from 59% to 72%, with sensitivities remaining above 80%. In addition, using revised ADOS algorithms, Le Couteur et al. (2008) examined the combined use of the ADOS and ADI-R for preschoolers with ASD. Consistent with the past study (Risi et al., 2006), combining information from both instruments provided improved diagnostic accuracy compared to either instrument in isolation.

In recent papers, newly developed ADI-R algorithms for toddlers and young preschoolers from 12 to 47 months of age as well as revised ADOS and new ADOS-Toddler algorithms showed improved validity compared to pre-existing algorithms used in past studies (Gotham et al., 2007; Kim & Lord, 2011; Luyster et al., 2009). Thus, the present study focuses on the validity of the combined use of the ADI-R and ADOS for toddlers and preschoolers from age 12 to 47 months using the new and revised algorithms.

In very young children, diagnostic differentiation between non-autism ASD (e.g. PDD-NOS) and autism is less stable than for older children and adolescents (Lord et al., 1999; Szatmari et al., 2002; Wiggins, Robins, Adamson, Bakeman, & Henrich, in press). Consequently, the new ADI-R algorithms for toddlers and young preschoolers and the ADOS-T algorithms provide only a single classification of ASD. In addition, in order to formally acknowledge the less clear stability of diagnoses in younger children, these algorithms provide *ranges of concern* (little-to-no, mild-to-moderate, or moderate-to-severe concern), to be used in clinical monitoring and follow-up. However, because more strictly stratified groupings are necessary for some purposes, the new ADI-R algorithms also provide two cutoffs, one for research (more restrictive; higher specificity with lower sensitivity) and one for clinical purposes (more inclusive; higher sensitivity with lower specificity).

Past studies examining validity of the ADI-R and ADOS have found that parent reports and clinician observations do not always agree. Agreement between these instruments has varied across samples and analytic techniques. In a sample of 797 ASD and 163 non-spectrum cases over 36 months of age, Risi et al. (2006) found that the Pearson r correlation between ADI-R and ADOS algorithm totals was 0.57. Correlations differed by domains in the study by Le Couteur et al. (2008), ranging from 0.51 to 0.71 for a sample of 77 preschoolers with ASD and 24 with other developmental disorders. Agreement between the instruments using Kappa ranged from 0.48 to 0.62. In another study (de Bildt et al., 2004), correlations ranged from 0.52 to 0.54 between the ADI-R and ADOS algorithm totals for 123 children aged 5 to 20 years with ASD *and* intellectual disability and 62 with intellectual disability *only*. In contrast, Ventola et al. (2006) compared the performance of the ADI-R and the ADOS to each other and clinical diagnosis in a sample of 36 ASD and 9 non-spectrum cases aged 16 to 31 months. Significant levels of agreement were found between the ADOS and clinical judgment (κ=0.59, p<.001) but agreement between the ADI-R and clinical judgment (κ=0.15, ns) and between the ADI-R and the ADOS (κ=0.07, ns) was poor.

Because the combined use of the ADI-R and ADOS has shown better diagnostic validity than either individual instrument, it is recommended that clinicians and researchers use information from both instruments when making diagnoses. However, due to constraints in time, cost, or expertise, often only one of the instruments is actually used. Relatively little is known about ways to maximize validity in this case. One approach would be to determine scores on the instruments associated with a very high (or low) probability of receiving the classification of ASD on the "alternative instrument" (referred to as "*positive* (or *negative*) *screening estimate*" hereafter). For instance, if a child's score reaches a *positive screening estimate* on the ADI-R, a clinician could presumably omit the ADOS assuming that the probability of the child receiving the ASD classification on the ADOS would be very high. The same strategy could be used with a *negative screening estimate.*

Another approach is to conduct similar analyses using best estimate clinical (BEC) diagnoses based on all available information as the gold standard and then to determine if there are scores on each instrument that result in 100% specificity for ASD. That is, we can examine what score on each instrument successfully excludes all cases determined to not have ASD (henceforth referred to as "*high specificity case scores*") and then describe the sensitivities of these scores. For example, if a child meets or exceeds a *high specificity case score* on the ADOS, a clinician evaluating the child could assume that the chance of the child receiving a BEC diagnosis of ASD would be very high and consequently omit the ADI-R.

In sum, the purpose of this study is to examine the combined use of the ADI-R and ADOS for children under age 4 using the new and revised algorithms. Often, a misdiagnosis that results in a child failing to receive necessary services is the greatest concern. On the other hand, over-diagnosis has negative consequences for individual children, public health strategies and research. Consequently, we present data supporting alternative methods for using both research and clinical cutoffs from the new ADI-R algorithms. Agreement between the two instruments is also evaluated by examining the overlap between the ADI-R and ADOS-T *ranges of concern* and correlations between algorithm totals.

Methods

Participants

All 604 children with complete data from a contemporaneous ADOS, ADI-R, nonverbal IQ, and BEC diagnosis were included from two projects, Early Diagnosis of Autism (EDX) and First Words and Toddlers (FW/T) and for clinic patients at the University of Michigan Autism and Communication Disorders Center (UMACC).

Children in the FW/T projects entered the study between 12 to 18 months and were administered the ADI-R and ADOS-T. The remaining children were administered the ADI-R and either the Pre-Linguistic ADOS (PL-ADOS; DiLavore, Rutter & Lord, 1995), or ADOS Module 1 to 3 depending on their age and language level. Out of 604 children, 195 children, who were nonverbal or had single words only, received the PL-ADOS, which was re-coded to the ADOS Module 1.

All participants, aged 12 to 47 months, were walking at the time of assessment. Mean age was 31.8 months (*SD*=9.6), and 435 children had ASD (345 males), 113 children non-spectrum disorders (NS; 81 males), and 47 children typical development (TD; all younger than 21 months; 31 males). NS participants had a range of diagnoses, including language disorders (53%), intellectual disability of unknown etiology (18%), Down syndrome (6.4%), externalizing disorders (5.5%), internalizing disorders (2.7%), and general, mild developmental delays (14.4%). Ethnicity was not associated with diagnosis; 74% of participants were Caucasian, 15% African American, 3% Asian American, 3% biracial, and 5% Native American or other races. The sample in the present study was a subset of children (about 30%) from the sample used to develop the new ADI-R algorithms for toddlers and young preschoolers (Kim & Lord, 2011). In addition, approximately at least 30% and 15% of the sample also used for the development of revised ADOS algorithms and new ADOS-T algorithms, respectively (Gotham et al., 2007; Luyster et al., 2009).

Participants were divided into three developmental cells by the child's age and language level following the structure of the developmental groupings of the new ADI-R algorithms: (1) all children between 12 and 20 months, 31 days of age and nonverbal children between 21 and 47 months, 31 days of age (“12–20/NV21–47”); (2) children between 21 and 47 months, 31 days of age with single words (“SW21-47”); and (3) children between 21 and 47 months, 31 days of age with phrase speech (‘‘PH21–47’’).

As shown in Table 3.1, children with TD and NS were significantly younger and had significantly higher NVIQ and Vineland Adaptive Behavior Composite scores (Sparrow, Balla, & Cicchetti, 1984) than children with ASD for the “12-20/NV21-47” group (p<.001). For both “SW21-47” and “PH21-47” groups, Vineland composite scores were significantly higher for children with NS than ASD (p<.001). A significant age difference emerged for the “SW21-47” group (children with ASD were older than children with NS, p<.05).

Measures

In the new ADI-R algorithms for toddlers and young preschoolers, item scores in *Social Affect* (SA) and *Restricted and Repetitive Behaviors* (RRBs) for the “12-20/NV21-47” and “SW21-47” groups and *Social Communication* (SC), *RRBs*, and *Reciprocal and Peer Interaction* (*RPI*) for the “PH21-47” group are combined to generate cutoffs for the classification of ASD. Thirteen to 20 items comprise the new ADI-R algorithms depending on children’s ages and language levels. For the revised ADOS and new ADOS-T algorithms, the total number of items in the algorithms is 14, with the composition of items in each algorithm differing by children’s ages and language levels.

Procedure

Each caregiver was administered the ADI-R and the Vineland. The ADOS and cognitive testing were then completed by the same or by a different clinical psychologist or a trainee within a few days' time. A standard hierarchy of cognitive measures, most frequently the Mullen Scales of Early Learning (n=438; Mullen, 1995) or the Differential Ability Scales (n=61; Elliott, 1990) was used to determine IQ scores. Examiners in the study had completed research training and met standard requirements for research reliability for the ADI-R and ADOS. Inter-rater reliability was monitored through periodic observations and scoring by two examiners and scoring of videotapes. Caregivers signed an Institutional Review Board approved informed consent to participate in research before participation.

Consensus Best Estimate Clinical Diagnosis

For children in the EDX study, an experienced clinical researcher used the videotaped ADOS and ADI-R scores and observations made during the testing to generate an independent BEC diagnosis of autism, PDD-NOS, or non-spectrum disorders (APA, 1994). For children in the FW/T project, scores on the ADI-R, ADOS, and clinical observations were used by two clinicians to make a BEC diagnosis operationalizing DSM-IV criteria (APA, 1994; See Luyster et al., 2009). For clinic cases, a diagnosis was made by a psychologist and/or psychiatrist after review of all information.

Analyses

Sensitivities and specificities for single and combined use of the ADI-R and ADOS algorithms were compared with BEC diagnoses. Sensitivities and specificities (Siegel, Vukicevic, Elliott, & Kraemer, 1989) were considered in each of these conditions: 1) Meeting ADI-R criteria; 2) Meeting ADOS criteria; 3) Meeting *either* ADI-R *or* ADOS criteria when both were administered; 4) Meeting criteria on *both* the ADI-R *and* ADOS. Characteristics of children correctly or incorrectly classified were examined. Correlations were used to assess the agreement between the ADI-R and ADOS algorithm totals as well as between domain totals for three different developmental cells ("12-20/NV21-47," "SW21-47," "PH21-47). Correlation coefficients were compared using Fisher's Z transformations (Steiger, 1980). Seventy children who received both new ADI-R *and* ADOS-T algorithms were selected to examine the overlap between the *ranges of concern* from both instruments. Odds ratios were calculated to assess the likelihood of receiving a diagnosis of ASD when a child was classified by the ADI-R and/or ADOS in these ranges.

Positive/negative screening estimates were identified for each instrument by selecting scores associated with very high/low percentages of cases that received a classification of ASD on the other instrument. Sensitivities and specificities for these scores were then evaluated. In addition, *high specificity case scores* were selected for each instrument by examining total scores that resulted in high specificities (100%, 90%, and 80%) of the BEC diagnoses for the comparison of ASD vs. NS cases. Sensitivities for these scores were also examined.

Results

Sensitivities and specificities for the comparison of ASD vs. NS

Not surprisingly, as shown in Table 3.2, the most satisfactory results were obtained when the most stringent condition, requiring a child to meet criteria on *both* the ADI-R (using clinical cutoffs) *and* the ADOS was used. In these cases, sensitivity and specificity for ASD vs. NS were consistently above 80%. For example, using both instruments yielded comparable sensitivities and significant improvements in specificities (4-22%) beyond when only ADI-R algorithms were used. Compared to when ADOS algorithms were used alone, using both instruments resulted in significant gains in specificities (10-31%) and slightly lower sensitivities, though they were still above 80% when ADI-R clinical cutoffs were used. As noted in previous papers (Gotham et al., 2007; Risi et al., 2006), when children with nonverbal mental ages below 15 months were included, specificities were slightly lower. Because evaluating children with low non-verbal mental age is a reality in clinical practice, these specificities are reported in parentheses (See Table 3.2). However, we also present separate results from data without children whose non-verbal mental ages fell below 15 months for researchers who wish to restrict their samples for better diagnostic accuracy.

As expected, the least restrictive condition, requiring a child to meet *either* the ADI-R *or* ADOS criteria, resulted in excellent sensitivities for ASD cases (97-99%), but poor specificities (45-85%). As in past studies, for all developmental cells, sensitivities improved when children whose BEC diagnoses were PDD-NOS were excluded. Although comparisons between the ASD and TD cases are not very informative clinically, because much research with younger children contrasts ASD and mixed TD and NS samples (as in studies with baby siblings), it is useful to know that not surprisingly, specificities also improved when TD cases were included. Likelihood ratios for the comparison of ASD vs. NS were most satisfactory when both instruments were used in combination using the conventional criteria (Likelihood ratio above 5 is considered satisfactory; Jaeskchke, Guyatt, & Lijmer, 2002).

Characteristics of misclassified children

We then compared the characteristics of true positives (TPs) and false negatives (FNs) for each instrument as well as false positives (FPs) and true negatives (TNs). The most common trend was that FPs (NS cases misclassified as ASD) were significantly older and had significantly lower NVIQ and Vineland scores than TNs (correctly classified NS cases). On the other hand, FNs (ASD cases misclassified as NS) were younger and showed higher NVIQ and Vineland scores than TPs (correctly classified ASD cases). See Table 3.3.

Overlap between the ADI-R and ADOS-T ranges of concern

Most children (71%) whose scores were in the little-to-no range of concern in the ADI-R fell in the same range in the ADOS-T. Similarly, 64% of children whose scores fell in the moderate-to-severe range in the ADI-R fell in the same range in the ADOS-T. If a child was classified as at risk (mild-to-moderate and moderate-to-severe ranges) by only one instrument (23%), the odds ratio for the child to be placed in a risk group by the other instrument was 12.69 (χ^2=19.2, p<.001, See Figure 3.1). When children were placed in a risk group by *both* instruments (50%), the odds ratio of having a BEC ASD diagnosis was 56.19 (χ^2=19.2, p<.001).

Agreement across the instruments

A correlation between the ADI-R and ADOS algorithm totals for the "12-20/NV21-47" group (r=0.75) was also significantly greater than those of the "SW21-47" and "PH21-47" groups (r=0.47, Z=4.7; r=0.59, Z=2.7, both p<.01). The correlation between the ADI-R and ADOS SA domains for the "12-20/NV21-47" group (r=0.69) was significantly greater than that of the "SW21-47" group (r=0.49, Z=3.1, p<.01). The correlation between the ADI-R and ADOS RRB domains for the "12-20/NV21-47" group (r=0.62) was significant greater than those of the "SW21-47" and "PH21-47" groups (r=0.44, Z=2.5, p<.05; r=0.29, Z=3.9, p<.01, respectively).

Positive and negative screening estimates

Total scores on the ADI-R and ADOS algorithms which resulted in very *high* probabilities (100%) of receiving an ASD classification on the other instrument (*positive screening estimate*) for all ASD cases, ranged from 18 to 25 and 18 to 22 respectively. Total algorithm scores resulting in very *low* probabilities (less than 5%) of receiving an ASD classification on the other instrument (*negative screening estimate*) ranged from 4 to 5 and 8 to 11 respectively. See Table 3.4 in Electronic Appendix for sensitivities and specificities for these scores.

High specificity case scores

We then identified scores on the ADI-R and ADOS that resulted in high specificities for BEC diagnoses of ASD (Table 3.5). The lowest scores on both instruments that resulted in 100% specificities were first identified for each developmental cell. For the ADOS, when scores were selected by 100% specificity, sensitivities ranged from 17 to 80% depending on developmental cells. Clinical cutoffs on the ADI-R are reported. As expected, when the ADI-R research cutoffs were used, a similar pattern emerged but with lower sensitivities and higher specificities. For high specificity case scores (100%) on the ADI-R, sensitivities ranged from 14 to 41%. Scores that resulted in specificities around 90% and 80% were also identified. An example of sequential assessment strategies using the PSE, NSE and high specificity case scores are described in Figure 3.2 and the Discussion.

Discussion

Consistent with findings from older children (Risi et al., 2006), use of information from both the new ADI-R algorithms for toddlers and young preschoolers and the revised ADOS and new ADOS-T algorithms together better reflected clinical best estimate diagnoses of ASD than when either single instrument was used. The ADI-R includes a developmental history and a detailed description of individual's functioning in a variety of social contexts as well as caregivers' perceptions of the level of impairment and/or frequency of different behaviors. The ADOS provides a summary of an experienced clinician's standardized observations of individual's behaviors within contexts that elicit social initiations and responses as well as communication interchanges. As suggested by low to moderate correlations between the ADI-R and ADOS in this study and in previous research (de Bildt et al., 2004; Ventola et al., 2006), the instruments provide overlapping but not identical information. Though the lack of high agreement between the instruments is frustrating in terms of each instrument's diagnostic validity, it increases their additive value. In fact, the combination of new and revised algorithms revealed even higher validity for toddlers and preschoolers than expected from studies using the original algorithms (Risi et al., 2006).

These newly developed and revised algorithms were created in a way that the influence of age and IQ scores on the algorithm scores was minimized. Nevertheless, we found differences in age, IQ, and adaptive functioning between children who were correctly identified and those misclassified by the instruments. For example, ASD cases misclassified as NS tended to be younger toddlers who had higher nonverbal intellectual and adaptive functioning. On the other hand, NS cases misclassified as ASD were older preschoolers with lower intellectual and adaptive functioning. These results are consistent with past studies showing that differentiating children with ASD from other developmental disorders is more difficult for very young children, children with severe delays (with lower IQ scores and/or who are nonverbal), and the most able toddlers and young preschoolers (with very high IQ scores and/or phrase speech; Gotham et al., 2007; Lord, Storoschuk, Rutter, & Pickles, 1993).

More able children, in this case, primarily older preschoolers, showed lower correlations between the ADI-R and ADOS than the younger and/or nonverbal group. In addition, mean ADI-R domain and algorithm total scores were lower for the "PH21-47" group than the "SW21-47" and "12-20/NV21-47" groups whereas mean ADOS scores were similar across all three groups. This may indicate that parents of preschoolers with more advanced levels of language, children who almost always also have stronger nonverbal skills, perceive their children's symptoms as less severe than clinicians evaluating the same children based on direct observations. This supports the usefulness of integrating perspectives from both caregivers and experienced clinicians especially when evaluating more complex cases.

Different sequential strategies could be used to determine when use of a single instrument might be sufficient. Each strategy has a distinct process in terms of obtaining a diagnostic classification. For example, as in 3.2, if a clinician first administered the ADOS and the child's score on the ADOS was above the PSE (or the high specificity case score), unless other information suggested otherwise, the clinician could reasonably assume that the child would be likely to receive an ASD diagnosis without administering the ADI-R. Based on the clinic referrals from the dataset used in the present study, such an approach was appropriate for about 72% of the clinic referrals (with 52% very likely ASD and 20% likely not ASD). However, about 28% of the referrals obtained less decisive scores, showing that such an approach may not be appropriate for all children (See Table 3.4 and Table 3.5 in Electronic Appendix). It is also important to note that UMACC is an autism clinic; in general developmental disorders clinics, autism cases would comprise a smaller proportion of likely diagnoses, so that the percent of cases with scores below or equal to NSE and possibly those in the less decisive range would increase. Studies of baby siblings of children with autism also suggest that the proportion of less decisive cases may be higher when children are not specifically referred for an autism assessment (Landa & Garrett-Mayer, 2006; Zwaigenbaum et al., 2005).

Although distributions of children by *ranges of concern* did not overlap perfectly between the ADI-R and ADOS-T, the majority of children classified as those needing follow-up evaluations and treatments by one instrument were also classified as at risk by the other instrument. In addition, the high likelihood ratio of receiving a BEC diagnosis of ASD for children classified into the risk groups by both instruments supports the validity of ASD risk categories even in very young children.

Limitations

Compared to the samples in past studies, the sample used in the present study was smaller because we selected only children with a contemporaneous ADI-R and ADOS for each case. Thus, restricted size and possible recruitment biases of more complex cases in the control groups (children with NS and TD) may have resulted in lower specificities in the present study compared to the original studies (Gotham et al., 2007; Kim & Lord, 2011; Luyster et al., 2009). In addition, because samples included a subset of children from previous studies mentioned above, replications from different sites will be critical.

The ADI-R and ADOS were administered by the same clinician for 75% of children; for 66% of these children, the ADI-R was administered before the ADOS. Thus, in about half of the cases, clinicians were not blind to developmental history and the caregiver's descriptions, which might have affected their ADOS administration and coding. However, the correlation between the algorithm total scores for the two instruments was slightly higher when different clinicians versus the same clinician administered the instruments (r of 0.66 vs. 0.59).

Conclusion

The ADI-R and ADOS provide both unique and overlapping information important for clinicians and researchers making diagnostic decisions. When both instruments were used in combination, well-balanced sensitivities and specificities were obtained. Even with young children, validity for combined use of the instruments in the present study was comparable or higher than in past research (Risi et al., 2006). Taking into account information from both a skilled clinician and a caregiver contributes to diagnostic differentiations especially for more complex cases. Alternative combinations with other instruments besides or in addition to the ADI-R and/or ADOS, such as the SCQ (Rutter et al., 2003), SRS (Constantino & Gruber, 2005), CCC-2 (Bishop, 2003), and the Screening Tool for Autism in Two-Year-Olds (STAT; Stone, Coonrod, & Ousley, 2000) may be equally effective. In addition, sequential assessment strategies may be appropriate for some children allowing cost- and time- effective research and clinical practice.

Table 3.1. Description of sample

Diagnosis	12-20			NV21-47		SW21-47		PH21-47	
	ASD	NS	TD	ASD	NS	ASD	NS	ASD	NS
Gender (male, female)	30, 6	13, 5	31, 16	175, 35	19, 8	86, 37	17, 11	54, 12	32, 8
Age in months, Mean	17.6	16.8	15.6	32.7	28.8	36.1	33	41.1	39.3
(Standard Deviation)	(2.2)	(2.6)	(2.6)	(7)	(5.8)	(6.6)	(6.1)	(5.1)	(5.4)
Mean NVIQ*	91.3	97.4	112.6	65.8	74.7	69.9	69.2	91.4	97.8
(Standard Deviation)	(15.6)	(14.4)	(15.3)	(19.9)	(27.6)	(17.5)	(18)	(20.8)	(16.5)
Mean VABC	77.1	83.3	96.5	59.7	65.9	65.3	72.6	76.3	83.9
(Standard Deviation)	(12.9)	(10.4)	(10.1)	(8.1)	(11.3)	(10.3)	(11.2)	(13)	(10.7)

NVIQ nonverbal IQ, *VABC* Vineland Adaptive Behavior Composite standard score; *12-20* all children 12-20 months, *NV21-47* nonverbal children 21-47 months, *SW21-47* children 21-47 months with single words, *PH21-47* children 21-47 months with phrase speech.

*For some children, NVIQ scores were not available, thus replaced by full scale IQ scores: 37 TD cases in "12-20" group; 1 ASD and 2 NS cases in "PH21-47" group.

Table 3.2 Validity of all conditions tested

		Sensitivity		Specificity	ASD vs. NS
		Autism	ASD	NS	Likelihood Ratio Positive
12-21/NV21-47	ADI-R (CLI) and ADOS	92	84	86 (80)	6
	ADI-R (RES) and ADOS	85	76	96 (89)	19
	ADI-R (CLI) alone	92	85	73 (69)	3
	ADI-R (RES) alone	85	78	89 (84)	7
	ADOS alone	99	95	76 (64)	4
	ADI-R (CLI) or ADOS	100	98	57 (49)	2
	ADI-R (RES) or ADOS	100	98	70 (60)	3
SW21-47	ADI-R (CLI) and ADOS	98	91	92	11
	ADI-R (RES) and ADOS	79	71	96	18
	ADI-R (CLI) alone	99	94	85	6
	ADI-R (RES) alone	80	73	92	9
	ADOS alone	99	97	65	3
	ADI-R (CLI) or ADOS	100	99	62	3
	ADI-R (RES) or ADOS	100	97	85	6
PH21-47	ADI-R (CLI) and ADOS	90	82	80	4
	ADI-R (RES) and ADOS	83	71	88	6
	ADI-R (CLI) alone	90	83	58	2
	ADI-R (RES) alone	83	73	75	3
	ADOS alone	100	97	68	3
	ADI-R (CLI) or ADOS	100	99	45	2
	ADI-R (RES) or ADOS	100	98	55	2

12-20/NV21-47 all children 12-20 months and nonverbal children 21-47 months , *SW21-47* children 21-47 months with single words, *PH21-47* children 21-47 months with phrase speech, *ADI-R* Autism Diagnostic Interview-Revised, *ADOS* Autism Diagnostic Observation Schedule, *CLI* Clinical Cutoff, *RES* Research Cutoff.
Numbers in parentheses are when children with nonverbal mental age lower than 15 were included.

Table 3.3 Characteristics of misclassified children

	12-20/NV21-47				SW21-47				PH21-47			
	ADI-R PPV = 95 NPV = 47 ADOS PPV = 94 NPV = 73				ADI-R PPV = 96 NPV = 76 ADOS PPV = 93 NPV = 85				ADI-R PPV = 76 NPV = 68 ADOS PPV = 83 NPV = 93			
	N	Age	NVIQ	VABC	N	Age	NVIQ	VABC	N	Age	NVIQ	VABC
ADI-R TPs	209	30.8(8.4)	68.4(20.6)	60.9(9.8)	116	36.3(6.6)	70.3(17.9)	64.8 (9.8)	55	41.4(5.2)	89.6(19.8)	73.7(11.9)
ADI-R FNs	36	28.3(8.6)	76.9(24.1)	70 (13.3)	7	33.1(5.5)	63.7(8.9)	75.3(14.3)	11	40 (4.4)	103.8(24.5)	90.8(9)
ADI-R FPs	14	26.7(8.1)	74.5(29.5)	67.1(15.4)	4	38 (2.9)	72.5(10.6)	64 (5)	17	40.5(4)	91.8(16.9)	82.7(12.5)
ADI-R TNs	31	22.8(7.1)	88 (23)	74.6(12.4)	22	32.6(6.2)	70 (19.3)	75.4(10.8)	23	38.3(6)	100.4(16.6)	84.8(9.4)
ADOS TPs	234	30.8(8.4)	68.7(20.9)	61.8(10.5)	120	36.2(6.4)	69.5(17.2)	65.1(9.9)	64	41.3(5.1)	91.4(19.6)	76 (13.1)
ADOS FNs	11	24.7(8.0)	86.7(24.1)	72.3(15.4)	3	-*	-*	-*	2	-*	-*	-*
ADOS FPs	16	25.6(8.4)	69.8(24.8)	65.6(14.1)	9	31.3(6.2)	72.4(14)	77.9(12.7)	13	40.6(3.8)	91.5(13.5)	83.3(9.6)
ADOS TNs	29	23 (7)	91.6(23)	76.1(12.2)	17	34.5(6)	69.2(20.3)	71.4(9.4)	27	38.6(5.8)	99.2(18.2)	84.2(11.3)

PPV Positive Predictive Value, *NPV* Negative Predictive Value, *12-20/NV21-47* all children from 12-20 months and nonverbal children from 21-47 months, *SW21-47* children from 21-47 months with single words, *PH21-47* children from 21-47 months with phrase speech, *ADI-R* Autism Diagnostic Interview-Revised, *ADOS* Autism Diagnostic Observation Schedule, *TPs* True Positives, *FNs* False Negatives, *FPs* False Positives, *TNs* True Negatives. Clinical cutoffs were used for the ADI-R.
*Sample size is too limited for the comparison. Significant differences emerged between FPs and TNs using the ADOS and between TPs and FNs using the ADI-R for the “12-20/NV21-47” group for NVIQ and VABC scores, between TPs and FNs using the ADI-R for the “SW21-47” group for VABC scores, between TPs and FNs by ADI-R for the “PH21-47” group for NVIQ and VABC scores (all results $p<.05$).

Table 3.4 Sensitivities and specificities of Positive and Negative Screening Estimates (PSE/NSE)

		12-20/NV21-47			SW21-47			PH21-47		
ADI-R	*N (ASD vs. NS)*	*289 (244, 45)*			*115 (85, 30)*			*105 (65,40)*		
		score	sens	spec	score	sens	spec	score	sens	spec
	PSE (100%)	18	50%	95%	25	10%	100%	19	57%	78%
	NSE (<5%)	4	96%	34%	5	95%	42%	5	97%	20%
ADOS		ADOS-T			Module 1			Module 2		
	N (ASD vs. NS)	*46 (31, 18)*			*113 (90, 13)*			*61 (52, 39)*		
		score	sens	spec	score	sens	spec	score	sens	spec
	PSE (100%)	22	6%	100%	16	64%	100%	18	24%	97%
	NSE (<5%)	10	100%	72%	11	93%	77%	8	89%	69%

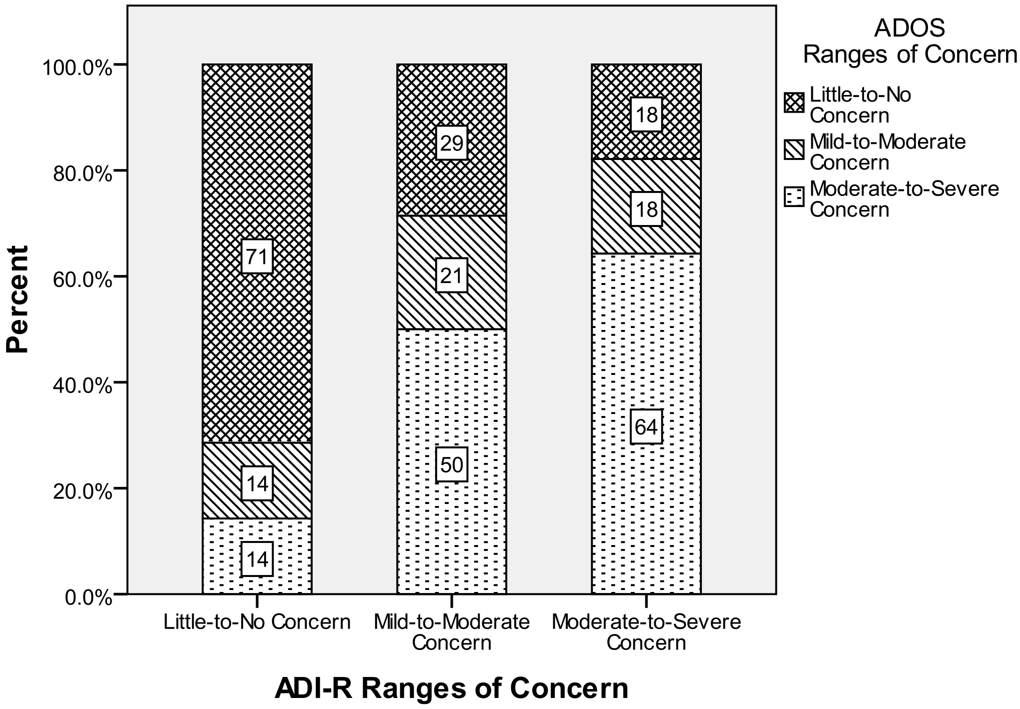
12-20/NV21-47 all children from 12-20 months and nonverbal children from 21-47 months, *SW21-47* children from 21-47 months with single words, *PH21-47* children from 21-47 months with phrase speech, *ADI-R* Autism Diagnostic Interview-Revised, *ADOS* Autism Diagnostic Observation Schedule.
The chance for the children whose scores are equal to or higher than PSE on one measure receiving the ASD classification on the other measure is very high (100%); the chance for the children whose scores are equal to or lower than NSE on one measure receiving the ASD classification on the other measure is very low (<5%).

Table 3.5 High specificity (100%, 90%, and 80%) case scores and sensitivities

			12-20/NV21-47		SW21-47		PH21-47	
ADI-R	*N (ASD vs. NS)*		*289 (244, 45)*		*115 (85, 30)*		*105 (65,40)*	
		specificity	*score*	*sensitivity*	*score*	*sensitivity*	*score*	*sensitivity*
		100%	22	22%	18	41%	28	14%
		90%	17	55%	13	70%	23	34%
		80%	12	77%	8	83%	21	43%
ADOS			**ADOS-T**		**Module 1**		**Module 2**	
	N (ASD vs. NS)		*46 (31, 18)*		*113 (90, 13)*		*61 (52, 39)*	
		specificity	*score*	*sensitivity*	*score*	*sensitivity*	*score*	*sensitivity*
		100%	18	35%	14	80%	20	17%
		90%	-	-	13	80%	11	68%
		80%	17	45%	12	90%	9	79%

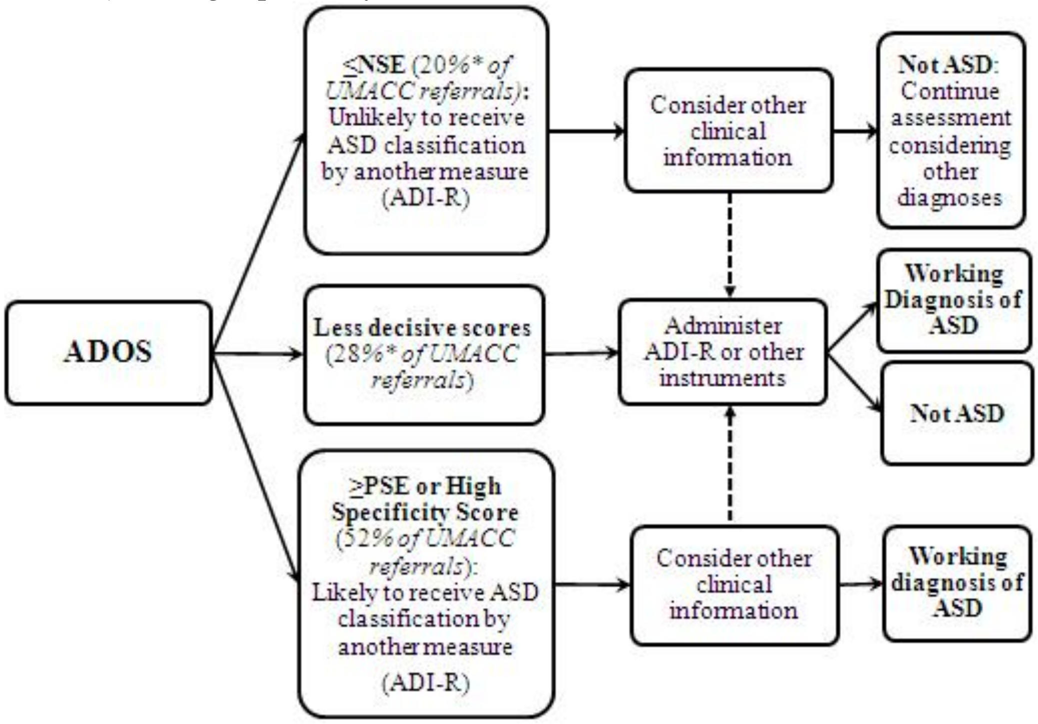
12-20/NV21-47 all children 12-20 months and nonverbal children 21-47 months, *SW21-47* children 21-47 months with single words, *PH21-47* children 21-47 months with phrase speech, *ADI-R* Autism Diagnostic Interview-Revised, *ADOS* Autism Diagnostic Observation Schedule, *ADOS-T* ADOS-Toddler.
High specificity case scores are available from the ADOS-T for 12-20/NV21-47 group, from Module 1 for SW21-47 group, from Module 2 for PH21-47 group.

Figure 3.1 Overlap between the ADI-R and ADOS ranges of concern



ADI-R Autism Diagnostic Interview-Revised, *ADOS* Autism Diagnostic Observation Schedule.

Figure 3.2 Sequential assessment strategies using positive/negative screening estimates (PSE/NSE) and high specificity case scores



ADI-R Autism Diagnostic Interview-Revised, *ADOS* Autism Diagnostic Observation Schedule. *In general developmental disorders clinics, autism cases would comprise a smaller proportion of likely diagnoses, thus the percent of cases with scores below or equal to NSE and/or possibly the less decisive range would increase.

References

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.). Washington, DC: Author.

Bishop, D. (2003). *The Children's Communication Checklist-2*. London: Psychological Corporation.

Bishop, D., & Baird, G. (2001). Parent and teacher report of pragmatic aspects of communication: use of the Children's Communication Checklist in a clinical setting. *Developmental Medicine and Child Neurology, 43*, 809-818.

Constantino, J., & Gruber, C. (2005). *Social Responsiveness Scale*. Los Angeles, CA: Western Psychological Services.

Constantino, J., LaVesser, P., Zhang, Y., Abbacchi, A., Gray, T., & Todd, R. (2007). Rapid quantitative assessment of autistic social impairment by classroom teachers. *Journal of American Academy of Child and Adolescent Psychiatry, 46*(12), 1668-1676.

Corsello, C. Hus, V., Pickles, A., Risi, S., Cook, E., Leventahl B., et al. (2007). Between a ROC and a hard place: decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, 48*(9), 932-940.

Dawson, G., Webb, S., Carver, L., Panagiotides, H., & McPartland, J. (2004).Young children with autism show atypical brain responses to fearful versus neutral facial expressions of emotion. *Developmental Science, 7*(3), 340-359.

de Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Erik Mulder, Volkmar, F., et al. (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(2), 129-137.

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

Elliott, C. D. (1990). *Differential Abilities Scale (DAS).* San Antonio, TX: Psychological Corporation.

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The Autism Diagnostic Observation Schedule (ADOS): Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders, 37*(4), 613–627.

Jaeskchke R, Guyatt G, Lijmer J. (2002) Diagnostic tests. In: Guyatt G, Rennie D, (Eds). *User's Guide to the Medical Literature* (pp. 121-140). Chicago: AMA Press.

Kim S, & Lord C. (2011) New Autism Diagnostic Interview-Revised (ADI-R) algorithms for toddlers and young preschoolers from 12 to 47 months of age. *Journal of Autism and Developmental Disorders, 42*(1), 82-93.

Kim S, & Lord C. (2010). Restricted and repetitive behaviors in toddlers and preschoolers with autism spectrum disorders based on the Autism Diagnostic Observation Schedule (ADOS). *Autism Research,* 3(4), 162-173.

Landa, R., & Garrett-Mayer, E. (2006). Development in infants with autism spectrum disorders: a prospective study. *Journal of Child Psychology and Psychiatry, 47* (6), 629-638.

Le Couteur A, Haden G, Hammal D, McConachie H. (2007). Diagnosing autism spectrum disorders in preschoolers using two standardised assessment instruments: The ADI-R and the ADOS. *Journal of Autism and Developmental Disorders, 38*(2), 362-372.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (1999). *Autism Diagnostic Observation Schedule: Manual.* Los Angeles: Western Psychological Services.

Lord, C., Storoschuk, S., Rutter, M., & Pickles, A. (1993). Using the ADI-R to diagnose autism in preschoolers. *Infant Mental Health Journal, 14*(3), 234-252.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., et al. (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*, 1305–20.

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

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., et al. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(9), 1094.

Rutter, M., Bailey, A., & Lord. C. (2003). *The Social Communication Questionnaire*. Los Angeles: Western Psychological Services.

Rutter, M., Le Couteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised*. Los Angeles: Western Psychological Services.

Siegel B, Vukicevic J, Elliott G, Kraemer H. (1989). The use of signal detection theory to assess DSM-III-R criteria for autistic disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*, 542–548.

Sparrow S, Balla D, Cicchetti D. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines: American Guidance Service.

Steiger J. (1980). Test for comparing elements of a correlation matrix. *Psychological Bulletin, 87*(2), 245-251.

Stone, W., Coonrod, E., & Ousley, O. (2000). Screening Tool for Autism in Two-Year-Olds (STAT): Development and preliminary data. *Journal of Autism and Developmental Disorders, 30,* 607-612.

Szatmari, P., Merette, C., Bryson, S., Thivierge, J., Roy M., Cayer, M., et al. (2002). Quantifying dimensions in autism: A factor-analytic study. *Journal of American Academy of Child and Adolescent Psychiatry 41:*467-474.

Ventola, P. E., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J., et al. (2006). Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders, 36,* 839–847.

Wiggins, L., Robins, D., Adamson, L., Bakeman R., & Henrich C. (in press). Support for a dimensional view of autism spectrum disorders in toddlers. *Journal of Autism and Developmental Disorders.*

World Health Organization. (1990). *International Classification of Diseases* (10^{th} revision). Geneva: Author.

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

Chapter IV

Observation of Spontaneous Expressive Language:

A New Measure for Spontaneous and Expressive Language of Children with Autism Spectrum Disorders and Other Communication Disorders

Since Kanner (1943) defined the characteristics of Autism Spectrum Disorders (ASD) in his seminal article, communication impairments have been recognized as one of the core features of ASD along with social deficits and restricted and repetitive behaviors. For example, approximately 20% of the ASD population does not acquire any functional expressive language (Lord, Risi, & Pickles, 2004). Communication impairments in ASD include a variety of characteristics such as failure to acquire speech without compensating through alternative communication methods, use of stereotyped speech or delayed echolalia (e.g. repeating lines from a Disney movie), and difficulty initiating and maintaining meaningful conversation (e.g. not responding to others' leads or questions; Lord & Corsello, 2004).

A valid assessment of communicative functioning in children with ASD, in particular their spoken language skills, has significant implications for interventions and treatments. The emergence of spoken language in children with ASD is one of the most important variables predicting better outcomes in later childhood and adulthood (Gillberg & Steffenburg, 1987; Howlin et al., 2004; Venter, Lord, & Schopler, 1992). Thus, improvements in communication skills have become one of the main goals in early treatments of ASD (Kasari, 2010; Smith, Groen, & Wynn, 2000). However, despite numerous intervention studies that have focused on language acquisition, lack of uniform measurement approaches to assessing language skills has been problematic for the comparison of treatment outcomes across different intervention research studies (Tager-Flusberg et al., 2009).

Previous studies have shown that, for many children, difficulties in social use of language far exceed what we might expect given their delays in other areas (Bishop 2002; Condouris, Meyer, & Tager-Flusberg, 2003; Tomblin et al., 2004). In addition, a number of researchers have emphasized that assessments of spontaneous language are more effective for evaluating language deficits in children with specific language impairment than structured measures (Goffman & Leonard, 2000; Rescorla, Roberts, & Dahlsgaard, 1997). Almost all models of language disorders include some attention to how children use language spontaneously in social contexts, but in most structured language tests, spontaneous language (defined as language not directly elicited as part of the test) is not taken into account. Some omnibus language tests such as the Reynell Developmental Language Scales (Reynell & Gruber, 1990) require clinicians to obtain short language samples, but they are obtained in a relatively short amount of time in fairly constrained contexts (i.e. while looking at a picture or during a very brief activity with dolls). Transcriptions or formal linguistic analyses (e.g. Child Language Data Exchange System [CHILDES], Systematic Analysis of Language Transcripts [SALT]) can be used, but they are time consuming and require technical skills to score. Recognizing this, a new language assessment tool, the Observation of Spontaneous Expressive Language (OSEL), has been developed to measure children's social use of language based on spontaneous expressive language.

The OSEL is a 30-45 minute observational assessment which focuses on children's spontaneous expressive language use in standardized, but natural contexts. The OSEL is intended to be used for children with ASD and other communication disorders from 2 years up to 10-12 years of age whose language levels are equivalent to typically developing children from 2 to 5 years who are beginning to use syntax (e.g., emerging 2- to 3-word phrases) to expected levels of discourse (e.g., complex sentences with 2- to 3- clauses combined). The OSEL is designed to be used by a speech language pathologist, clinician, or researcher familiar with basic aspects of language structure.

Built on the model of the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000), the examiner administering the OSEL presents semi-structured, highly motivating materials and activities organized in 7 tasks that were specifically developed to elicit spontaneous use of expressive language in natural contexts (see Table 4.1). The first activity, the *Mr. Potato Head* task, gives a child the opportunity to interact within a relatively easy and usually familiar play context, namely to construct different Potato Head puppets. The relatively low demand of the task provides a gentle warm-up activity for a child and creates opportunities for him/her to use different nouns, adjectives, and verbs as well as to answer and ask questions. In the second activity, the *Picture Stories* task, a child is presented with four pictures that depict a story or plot and then is asked to describe the story to the examiner, thereby offering the opportunity for the child to demonstrate narrative skills. The third task, *Conversations*, may occur throughout the assessment as opportunities present themselves. This task requires the examiner to create contents in which the child initiates conversations about points of interest or experiences and responds to conversational leads provided by the examiner. The fourth task, the *Camping Trip*/*Picnic*, is highly motivating and provides an opportunity to observe a child's spontaneous language production within a loosely structured pretend-play activity (e.g., getting ready for a camping trip or picnic, building a house/tent/castle, cooking and eating dinner). It also involves some exploratory/sensory activities (e.g., exploring with a flashlight, fishing on a pretend lake made out of bubble wrap) that a child has the opportunity to lead. Again, consistent with the other tasks, various morphological structures (e.g., verb phrases, verb tenses) as well as sentence structures (e.g., coordination, subordination), and pragmatic skills (e.g., asking for clarifications, making comments to express interests) can be observed during this task. The next task, the *Throwing Game*, is more focused than the other tasks; it is intended to elicit the use of particular syntactic forms, spatial prepositions. This task is meant to give the child opportunities to verbally communicate locations of objects that are out of reach by using different prepositions. The next task, *Retell a Story: Where Are My French Fries?,* provides an opportunity for a child to re-tell a familiar story that incorporates theory of mind. This gives the examiner a chance to observe the child's semantic and narrative skills (e.g., synthesizing information, understanding of cause and effect relationships). In the last task, the *Picture Vignette*, a child is shown a picture that illustrates an open-ended story with some surprising elements (e.g., a frog jumping in a salad bowl during a family's picnic). The vignette provides opportunities for the child to describe objects, people and events at multiple levels. The child is expected to use different word classes (e.g., nouns, verbs, adjectives), verb tenses (e.g., future, past, present) and forms (e.g., modal auxiliary) while describing the vignette.

During the administration of these different activities, the examiner codes an extensive list of syntactic skills and some pragmatic and semantic skills using a real time coding system (see Appendix A). Both grammatical and ungrammatical uses of many different aspects of syntax are coded. Several semantic and pragmatic skills are also coded during the administration. After the administration, these codes are transferred to a Summary Coding table (see Appendix B). In addition, more detailed pragmatic and semantic skills are also coded after the assessment using the Pragmatic Semantic Profile (see Appendix C). The OSEL systematically captures syntax, pragmatics, and semantics based on children's spontaneous expressive language, through codes specifically focused on these aspects of language created for the real time coding system and the Pragmatic Semantic Profile. In addition to these newly created items, the OSEL includes some items from the ADOS that assess pragmatic skills and unusual features of language (e.g., conversations, asking for information, stereotyped language, immediate echolalia). These items have been expanded and elaborated for more detailed and comprehensive descriptions of pragmatic skills and unusual features of language.

The OSEL covers various aspects of expressive language skills identified through research on language development in typically developing children. Many of these aspects of language have also been commonly reported to be areas of difficulty for children with ASD and/or language impairments. More specifically, past studies noted that children with ASD have significant impairments in their morphosyntactic skills. For example, children with ASD were found to use definite articles (e.g. the, an, a) less often than children with intellectual disability (ID; Bartolucci et al., 1980). Children with ASD were also reported to produce less complex and more repetitive and formulaic grammatical structures rather than novel ones compared to children with Down syndrome, other developmentally delays, or typical development (Scarborough, Rescorla, Tager-Flusberg, Fowler, & Sudhalter, 1991; Tager-Flusberg & Calkins, 1990). In particular, pronoun confusions (e.g. I-you confusion, I-he/she confusion) were more prevalent in children with ASD than typically developing peers and children with other developmental disorders (Walenski, Tager-Flusberg, & Ullman, 2006). Bartolucci et al. (1980) also found significant differences between children with ASD and ID in their use of grammatical morphemes. In this study, children with ASD omitted Brown's fourteen grammatical morphemes (e.g. plurals [-s], possessives ['s], present-progressive [-ing], past tense [-ed]; see Brown, 1973 for a complete list) more often than children with ID.

The semi-structured, play-based environment of the OSEL was deliberately designed to evoke spontaneous use of a variety of these syntactic skills. For example, the use of different verb forms including regular and irregular past tenses (e.g., "I *caught* the fish!"), regular and irregular plurals (e.g., "There are *geese* in the lake."), and adjectives (e.g., "S'mores are my *favorite* snack.") are elicited through several activities. Other syntactic areas in which children with ASD may show impairments are also coded in the OSEL (e.g., Infinitive phrases, gerunds, negations, modal auxiliary verbs).

Children with ASD have been also found to show significant impairments in pragmatic and semantic skills. Pragmatic skills involve children's ability to use social aspects of language and conversation (Walenski et al., 2006). Semantic skills involve children's ability to interpret meanings in language (Harris et al., 2006). Impairments in pragmatic and semantic abilities are affected by social cognitive deficits such as joint attention and theory of mind, which have been found to be areas of considerable difficulty for children with ASD (Tager-Flusberg, Skwerer, & Joseph, 2006). In the course of tasks that elicit social initiations and responses, various aspects of pragmatics and semantics are coded in the OSEL. For example, throughout the OSEL administration, children are provided with opportunities to ask questions, provide information about their experiences, comment on the materials and the examiner's actions, and clarify what the examiner says. Other examples of pragmatic and semantic skills include reporting main ideas, synthesizing cause-and-effect information, and maintaining back-and-forth conversations. In addition, unusual features of language that children with ASD and other communication disorders have been reported to show are also coded in the OSEL (e.g., echolalia, impolite and inappropriate language, stereotyped/idiosyncratic use of words or phrases; see Appendix C).

In order to assess these different aspects of children's expressive language, while administering the OSEL, the examiner structures the tasks by adjusting the support for the child to show language skills spontaneously rather than initially eliciting those skills directly. For example, during the *Camping Trip/Picnic* task, the examiner follows a predetermined hierarchy of prompts to see if a child will request. The hierarchy begins with the examiner waiting to see if the child initiates interaction when he wants the examiner to give him a fishing pole. If the child does not spontaneously request, the examiner then looks deliberately at the child to see if he will say anything to request. Finally, if the child does not initiate a request, the examiner asks, "What do you need?". Thus, the OSEL coding reflects both how the child responded to the "press" for social behavior by using his pragmatic and semantic skills as well as how much the examiner had to structure the situation to elicit these responses.

This approach of using a predetermined hierarchy of prompts and structures in the administration of the OSEL is similar to that used in the ADOS. However, whereas the ADOS provides opportunities to elicit behaviors associated with core autism symptoms, the OSEL is designed to create opportunities to observe different aspects of expressive language skills which may or may not be associated with symptoms of ASD. As a result, the use of the OSEL is not limited to children suspected of having an ASD but also is intended for children who have specific language impairments (SLI), intellectual disabilities, or other developmental disabilities, and those who are suspected of having a language delay or disorder (beyond a phonological or articulation disorder) and are using language at less than a 5 year-old level (i.e., in chronological age from 2 to 10-12).

Before the OSEL is used to assess language skills of these clinical populations, a standardization of the instrument is needed to produce a meaningful metric of how a child's language level compares with that of the general population. The purpose of this study is to describe psychometric properties including the validity and reliability of the OSEL using a preliminary sample of typically developing children from 2 to 5 years of age. This normative group is not a population norm, but it will provide a preliminary comparison group to which to compare the language skills of children with ASD and other communication whose language levels are comparable to that of typical 2 to 5 year olds. The OSEL will be particularly useful for young preschoolers with ASD and other developmental disorders who have relatively mild language delays, as well as older, school-aged children with moderate communication impairments. Continuity in scoring from beginning syntax to expected levels of discourse as well as from young preschoolers to elementary school-aged children allows comparisons of children across different chronological ages, language levels, and genders. Based on the normative data, age equivalent scores for syntax, pragmatics, and semantics will be created to provide meaningful language profiles for the comparisons. The use of the quantified profiles for syntax, pragmatics and semantics can also inform programming intervention goals for children with ASD.

Methods

Participants

A sample was collected consisting of 176 typically developing children (96 males) between ages 2 to 5 years with a mean age of 42 months (*SD*=10.7) at the University of Michigan Autism and Communication Disorders Center (UMACC). All caregivers signed an Institutional Review Board approved informed consent to participate in research before participation. Children were divided into 7 groups by age (Table 4.2). Originally, the sample was divided into 6 age groups based on 6-month-intervals. However, the youngest age group showed a larger variance in score distributions, and consequently was divided further into 2 groups based on 3-month-intervals. Participant characteristics including age, IQ scores, and scores of other language instruments are described in Table 4.2. Out of 176 children, 83% of children were Caucasian, 9% Biracial, 5% Asian, 2% African American, and 1% other races or unknown. With regards to maternal education, 54% of mothers had degrees at a graduate or professional level, 28% from a four-year college, 17% at an associate or vocational level, and the rest were unknown. All children used English as their native language.

Measures and Procedures

Both language and cognitive assessments were completed by the same or by a different clinical psychologist, graduate student, or a trainee on the same day or within a few days' time. All of the data were collected by doctoral level students in clinical psychology and UMACC staff members who had achieved research reliability on the OSEL under the supervision of a licensed clinical psychologist. For the cognitive testing, the Abbreviated Battery from the Stanford Binet Intelligence Scale (SBIS; Thorndike, Hagen & Sattler, 1986) was used to estimate verbal and nonverbal intellectual levels. To assess children's language level, all children were administered the Preschool Language Scale, Fourth Edition (PLS; Zimmerman, Steiner, & Pond, 2003) and/or the Comprehensive Assessment of Spoken Language (CASL; Carrow-Wodlfolk, 1999) in addition to the OSEL. Each caregiver completed the Vineland Adaptive Behavior Scales (VABS; Sparrow, Balla, & Cicchetti, 1984).

Statistical Analyses

We first examined the means and standard deviations for the number of grammatical uses (coding only different "types" of uses by excluding the grammatical uses of the same word or utterance. For example, counting the use of a question, "what's this," was coded only once for a WH- question even if the child used the same question again). The distribution of each syntax item was examined separately for different age and gender groups. The usefulness of each item was assessed based on distributions of means, medians, and standard deviations. A total of 24 out of 28 items that showed gradual increases in means and medians from younger to older ages were included in a final set of items to create the "OSEL syntax totals." The OSEL syntax totals were intended to reflect the developmental progression of syntactic skills across development. Thus, two items were excluded from the final set of items (declarative and imperative sentences) because they were equally prevalent in all age groups. Two other items added toward the end of data collection (wh- infinite phrases, gerunds) were also excluded because there was not yet sufficient data to be analyzed.

Among the 24 items included in the final set, grammatical uses for 12 items were counted from 0 to 4; grammatical uses for two items were counted up to 3. Grammatical uses of 10 other items were counted up to ceilings of 6 to 18. These 10 items were recoded into the 0- to 4-point scale with higher scores indicating more grammatical uses. For 10 items that were recoded, no grammatical use (0) was coded as a score of 0, and remaining counts were recoded to a score of 1, 2, 3, and 4 based on the original score distributions. When recoded, the original scores were collapsed such that the number of children distributed across different original scores was as close to that of the recoded scores. Scores from all 24 items on either the 3-point or 4-point scale were combined to generate the OSEL syntax totals.

Similarly, we examined means, medians, and standard deviations for item scores in the pragmatic semantic profile (PSP) separately by gender and age groups. The scores for all of these items ranged from 0 to 3 with higher scores indicating more impairment. A total of 15 items was included in the final set that comprised "OSEL PSP totals."

Three items added toward the end of data collection (level of support required for conversation, intonation/volume/rhythm/rate, intelligibility) were not included in the OSEL PSP totals because of limited data available.

The factor structure of all syntax items used to generate the OSEL syntax totals was examined by performing an exploratory factor analysis (EFA) using Mplus software (Muthen & Muthen, 1998). A confirmatory factor analysis (CFA) was then performed. An EFA was also performed to examine the factor structure of all PSP items, with a CFA conducted to confirm the factor structure of the PSP. Based on the factors emerging from these analyses, the PSP subdomain totals were calculated by adding the scores of items within each factor.

The effects of gender, age, and verbal IQ on the OSEL syntax and PSP totals were examined using General Linear Models. Reliability and validity of the OSEL were then assessed based on the OSEL syntax and PSP totals. Intraclass correlations (ICC) were calculated for inter-rater and test-retest reliabilities using the OSEL syntax and PSP totals. Because of the high proportion of children scoring 0s on ungrammatical uses of many syntax items, these scores were not included in these analyses. Internal consistency was examined by calculating Cronbach's alpha for all participants. For this, age group 1 and 2 were collapsed to obtain sample sizes large enough for the analyses, which resulted in a total of 6 age groups. Concurrent and convergent validity was assessed through calculating correlations between the OSEL syntax and PSP totals and scores from other measures such as the Communication domain totals in a measure of adaptive behavior (e.g. VABS) and domain standard scores from other language measures (e.g. PLS, CASL) based on scores from all participants. The relations between the OSEL scores and chronological age and verbal and nonverbal IQ scores were also examined as indicators of concurrent validity and specificity of the OSEL.

Age equivalents were derived from the OSEL syntax and PSP totals separately by gender using a standard method (Ward, Stoker, & Murray-Ward, 1996). First, median scores from the syntax and PSP totals were computed for each age group and gender. These median scores were then plotted across age groups. A smooth line was fitted to the plotted points. Age equivalents corresponding to particular syntax and PSP totals were then read from the smooth lines separately by gender based on the equations of the lines. Thus, the age equivalents represent ages corresponding to median syntax and PSP scores predicted by the fitted line. This is consistent with other commonly used language measures for which age equivalents were developed based on the same method (CASL [Carrow-Wodlfolk, 1999]; PLS [Zimmerman, Steiner, & Pond, 2003]).

Ungrammatical items were examined by combining the occurrences of all ungrammatical uses (referred to as "OSEL syntax error totals") on 25 syntactic items (24 items for which both grammatical and ungrammatical uses were coded plus an additional item, subject-verb agreement error). For each item, there could be 6 possible ungrammatical uses. Mean OSEL syntax error totals were examined separately by each age and gender group. Correlations between the prevalence of grammatical errors and age and the OSEL syntax and PSP totals were also examined.

Creating Syntax and PSP Totals based on Factor Analyses

Results from the Exploratory Factor Analysis (EFA) using the OSEL syntax items showed that a 1-factor solution fitted well (Table 4.3). A Confirmatory factor analysis (CFA) was performed to examine the model fit for each group, and the result consistently showed that a 1-factor model fitted substantially better than 2-, and 3- factor models. The goodness-of-fit rating yielded a Comparative Fit Index (CFI) of 0.99 and 0.977 for the EFA and the CFA respectively (CFI between 0.9 and 1 indicating good fit; Skrondal & Rabe-Hesketh, 2004) and a Root Mean Square Error Approximation (RMSEA) of 0.047 and 0.059 for the EFA and the CFA respectively (RMSEA of 0.08 or less is considered a satisfactory fit; Browne & Cudeck, 1993).

Results from the EFA using the OSEL PSP items showed that a 3-factor solution fitted well (Table 4.4). Items loaded onto three factors, *Initiation of Reciprocal Communication, Narrative Skills*, and *Unusual Features* (See Table 4.4 for the item loadings). One of the items, Stereotyped Language, was excluded from the EFA due to the large portion of children scoring 0s (more than 90% of children in the sample). However, the item was included for the CFA because it is anticipated that many more children with AD will have scores other than 0 on this item. The goodness-of-fit rating yielded a CFI of 0.995 and 0.996 and a RMSEA of 0.05 and 0.04 for EFA and CFA respectively. Based on the 3 factors emerging from the analyses, PSP subdomain totals were calculated by combining item scores under each domain. “PSP 3 domain totals” were also created by adding item scores under all three domains. The mean syntax and PSP totals by gender and age groups are presented in Table 4.5.

Reliabilities

For inter-rater reliabilities, intraclass correlation (ICC) between raters was 0.96 for the syntax totals and 0.83 for the PSP totals (both $p<0.001$). For test-retest reliabilities, ICC for test-retest reliabilities was 0.95 for the syntax totals and 0.92 for the PSP totals (both $p<0.001$).

Internal Consistency

For all syntax items, Cronbach's alpha was 0.918 for Age Groups 1 and 2 combined, 0.904 for Age Group 3, 0.9 for Age Group 4, 0.837 for Age Group 5, 0.919 for Age Group 6, and 0.842 for Age Group 7 (all $p<0.001$). Cronbach's alpha across all age groups was 0.938 ($p<0.001$). For all PSP items, Cronbach's alpha was 0.642 for Age Groups 1 and 2 combined, 0.796 for Age Group 3, 0.761 for Age Group 4, 0.724 for Age Group 5, 0.660 for Age Group 6, and 0.677 for Age Group 7 (all $p<0.001$). Cronbach's alpha across all age groups was 0.8 for the PSP items ($p<0.001$).

Concurrent and Convergent Validity

As expected, the Pearson r correlation between the OSEL syntax totals and chronological age was 0.6 ($p<0.01$). Across all age groups, the correlation between the OSEL syntax totals with the PLS Expressive Communication domain scores was 0.4 ($p<0.01$). The correlation between the OSEL syntax totals and the PLS Auditory

Comprehension domain scores was also 0.4 (p<0.01) for all participants. The OSEL syntax totals were also correlated with the CASL Syntax Construction domain standard scores (r=0.6, p<0.01) and the CASL Pragmatic Judgment domain standard scores (r=0.5, p<0.01) using a subset of 112 children. The correlation between the OSEL and the VABS Commination domain was minimal (r=0.1, n/s) for all participants. Correlations between the OSEL scores and the estimated verbal and nonverbal IQ scores were r of 0.3 (p<0.01) for both verbal and nonverbal IQ scores.

The OSEL PSP 3 domain totals were also moderately correlated with age (r=-0.6, p<0.01). Across all age groups, the correlation between the OSEL PSP 3 domain totals (combined scores of items under all three domains; higher scores indicating absence/abnormality of skills specified) and the PLS Expressive Language was -0.4 (p<0.01). The correlation between the OSEL PSP 3 domain totals and the PLS Auditory Comprehension domain scores was -0.4 (p<0.01). The OSEL PSP 3 domain totals were also correlated with the CASL Syntax Construction and Pragmatic Judgment standard scores (both r=-0.5, p<0.01, n=112). The correlation with the VABS Commination domain was minimal (r=-0.1, n/s). Correlations between the OSEL PSP 3 domain totals with verbal and nonverbal IQ scores were both -0.3, (p<0.01).

Effects of Gender, Age, and Verbal IQ as Predictors of OSEL Syntax and PSP Totals

The General Linear Model showed that gender was a significant predictor of the OSEL syntax totals (F=7.57, p<0.05) and for the PSP Initiation of Reciprocal Communication domain totals, and the PSP 3 domain totals (F=6.62 and F=5.37 respectively, all p<0.05) while controlling for age and verbal IQ. Age significantly predicted the syntax totals (*F*=188.64, p<0.01) and all PSP totals (*F*=82.41 for Initiation of the Reciprocal Communication domain totals, *F*=107.77 for the Narrative Skills domain totals, *F*=25.01 for the Unusual Features domain totals, *F*=161.57 for the PSP 3 domain totals, all *p*<0.001). Verbal IQ was a significant predictor of the syntax totals (*F*=54.16, *p*<0.001) and all PSP totals (*F*=27.72 for the Initiation of Reciprocal Communication domain totals, *F*=22.23 for the Narrative Skills domain totals, *F*=22.71 for the Unusual Features domain totals, *F*=51.6 for the PSP 3 domain totals, all *p*<0.001).

Deriving Age Equivalents for Syntax and Pragmatic Semantic Profile Totals

The fit for the smooth lines based on the median syntax totals across age groups was R^2 of 0.86 for males and 0.89 for females. The fit for the PSP totals ranged from R^2 of 0.82 to 0.98 for different factors. Figure 4.1 shows an example of the smooth line fitted for the medians PSP Factor 1 (Initiation of Reciprocal Communication) totals for females. Age equivalents calculated from the smooth lines for the OSEL syntax and PSP totals for males and females are presented in Table 4.6 and Table 4.7 respectively (Ward, Stoker, & Murray-Ward, 1996). Because the behaviors specified under the PSP items that loaded onto the Unusual Features factor (e.g., stereotyped/idiosyncratic use of words or phrases, immediate echolalia) were rare in typically developing children included in the normative data, scores for these items were relatively low. The mean totals for this factor ranged from 0.1 to 2.3 with standard deviations ranging from 0.3 to 2.3 (See Table 4.5). Thus, age equivalents were not created for this factor due to the limited variability across age groups. However, item scores from this domain were included for the OSEL PSP 3 factor totals.

Ungrammatical Uses of Syntax Items

The mean OSEL syntax error totals are presented in Table 4.5. The lowest mean totals were obtained from children in Age Group 1 (from 24 to 27 months; 0.73 for males and 1.6 for females). The trend was that the errors generally increased with age and peaked at around 42-47 months for males and at around 48-53 months for females and decreased afterwards. Mean errors were slightly correlated with age (r=0.2, p<0.01) and OSEL syntax and PSP totals (both r=-0.3, p<0.01).

Discussion

The OSEL is a measure of children's spontaneous expressive language obtained in standardized, but natural contexts. Results indicate strong internal consistency for the OSEL syntax and PSP items. Concurrent and convergent validity were observed through moderate to strong associations between the OSEL syntax and PSP totals and other language measures (e.g. both Expressive and Receptive domains from the PLS and Pragmatic Judgment and Syntax Constructions subtests from the CASL). The OSEL is different from other structured language measures in several ways. For example, the OSEL is designed to tap into the morphosyntactic complexity and pragmatic and semantic skills based on children's spontaneous use of expressive language whereas the other language instruments provide global measures of receptive and expressive language skills obtained in highly structured settings.

The OSEL is a semi-structured observation which occurs in a brief time period (about 30-45 minutes). The results of the present study showed that, even in a relatively short amount of time, the OSEL successfully captured different aspects of expressive language skills (i.e., syntax, pragmatics, and semantics) in the normative sample of 2-to-5-year-olds. OSEL scores reflected developmental progressions in these different areas of expressive language skills. Age showed a strong positive correlation with the OSEL syntax totals (higher scores indicating more grammatical uses) and a negative association with scores on pragmatic semantic profiles (higher scores indicating more impairments). Results from the general linear regression analysis also showed that older children and/or children with higher verbal scores demonstrated more advanced grammatical, pragmatic, and semantic skills. The gradual progression of language levels observed in the normative data allowed derivation of age equivalent scores. These age equivalent scores will be particularly useful with a target population for the OSEL, children with ASD and other communication disorders from 2 to 10-12 years of age whose language levels are comparable to that of typical 2 to 5 year olds. The age equivalents provide a reference point to which a child's spontaneous use of language can be compared.

Results from the general linear regression analysis showed that gender made a significant independent contribution to OSEL syntax and PSP totals. Consistent with past research suggesting that language acquisition is more rapid for females than for males during toddler and early preschool years (Galsworthy, Dionne, Dale, & Plomin, 2000; Bauer et al., 2002), females showed significantly more grammatical uses and advanced pragmatic and semantic skills than males across all ages on the OSEL. Not surprisingly, the gap between males and females decreased over time. As a result, age equivalents were created separately by gender.

When the correlations between verbal IQ and the OSEL syntax and PSP totals were examined, they each remained at *r* of 0.3. Even though the correlations were minimal, they were still significant. In fact, the general linear regression analysis also showed that verbal IQ scores made significant independent contributions to the OSEL scores. This was expected given the role of language skills in the measurement of cognitive skills in young children. In fact, some of the items in the Abbreviated Battery from the Stanford Binet Intelligence Scale are highly associated with language skills (Thorndike, Hagen & Sattler, 1986).

Based on age equivalent scores, researchers and clinicians using the OSEL can obtain quantified profiles of spontaneous expressive language for children with ASD and other communication disorders for their syntactic, pragmatic, and semantic skills. The use of the quantified language profiles obtained from the OSEL can provide important information to researchers and clinicians about the changes in communicative functioning over development. More importantly, the OSEL may be used to identify specific areas of language skills that require intervention, as well as to capture the changes in expressive language skills that may occur over the course of treatment. Therefore, with the readily available quantified profiles of expressive language skills, the OSEL can contribute to the uniform use of language assessments, allowing comparisons across different treatment outcome as well as genetic and neuroimaging studies. Another advantage of the OSEL is that it focuses on children's spontaneous use of expressive language in standardized, but natural, contexts (e.g. while playing with a variety of toys, telling stories from a picture vignette, and interacting with an examiner during imaginative play). This is different from most standardized testing, which elicits responses that are knowledge based or highly tied to concepts (e.g., This chain is long, this chain is…) rather than spontaneous expressive skills. By using various play-based tasks in the OSEL, researchers and clinicians can obtain more meaningful profiles of spontaneous language skills, which reflect the language skills that children demonstrate in everyday activities (e.g., at home while interacting with parents and siblings, at school while interacting with teachers and peers).

The quantified profiles obtained from the OSEL for different domains of expressive language skills can also provide opportunities to identify potential subgroups of ASD and other communication disorders. A substantial number of children with these diagnoses may show significant impairments in pragmatic and semantic skills but have fairly intact syntactical skills. On the other hand, some children might show significant difficulties with syntax, but relatively stronger pragmatic and semantic skills. OSEL scores can also facilitate further investigations of the associations between these subgroups and possible genetic and neurobiological correlates.

Limitations

One of the limitations of this study was the limited sample size for the youngest female group. In addition, a large proportion (more than a half) of the sample included children whose parents had higher educational backgrounds (a graduate or professional level education). Eventually, age equivalents derived from the present sample will need to be replicated with a larger, more representative sample, before they are made available to the broader research and clinical communities. With the larger normative sample, the score distributions of items that were added toward the end of the data collection process (gerunds and conjunctions for syntax items; level of support required for conversation, intonation/volume/rhythm/rate, and intelligibility for PSP items) can be examined further to test the feasibility of including these items in the total scores and age equivalents.

Because the OSEL was validated with typically developing children, codes for ungrammatical uses of syntactic items that were originally designed for children with ASD and other communication disorders were consistently low for the normative data across different age groups. On average, both males and females showed fewer than 1 error at around age 2, and 4 to 5 errors around the ages of 3 to 4 years. Ungrammatical uses were only slightly associated with age. We expect that the patterns will be different for children with ASD (e.g., more errors than typically developing children across all ages). Similarly, as expected for typically developing children, raw item scores under the PSP Unusual Features domain, also originally created for clinical populations, were lower than the scores on the other two domains. Thus, the item score distributions under this domain (e.g., stereotyped/idiosyncratic use of words or phrases, immediate echolalia, semantic errors) should be reexamined in clinical populations. It is expected that children with ASD will have significantly higher scores than children in the normative sample on these items. Further research is needed with clinical populations to identify the pattern of language impairments in this area.

For reliability assessments, weighted kappas are commonly calculated (Fleiss, 1986). However, due to the small number of children included for reliability calculations (n of 10) and low variability in score distributions for some items (e.g., all of the 10 children scoring ceilings on items such as progressive verbs and a number of nouns), a weighted kappa coefficient for each item was not calculated in this study. Instead, intraclass correlations for OSEL syntax and PSP total scores were calculated. In addition, percentage agreements at the item level will be examined for inter-rater and test-retest reliabilities. Weighted kappas will be calculated as well to replicate these results with the larger sample.

Conclusion

The OSEL is a measure of children's spontaneous use of language in standardized, but natural contexts. In a relatively brief time period (about 30-45 minutes), the OSEL provides quantified profiles of spontaneous expressive language skills in typically developing children from 2 to 5 years of age using syntax and pragmatic-semantic totals and age equivalents. It is hoped that the OSEL can be used in combination with other language measures to evaluate strengths and weaknesses of expressive language skills in children with ASD and other communication disorders from 2 up to 10-12 years (Tager-Flusberg et al., 2009). In the near future, using a sample of children with ASD and other developmental disorders (e.g., language delays, intellectual disabilities), the validity of the measure will be further evaluated by comparing the distributions of item scores across different diagnostic categories. Children with ASD and other communication disorders would show more impairment in morphosyntactic skills when compared to typically developing children. It is also expected that children with ASD will show more considerable difficulty in pragmatic and semantic skills compared to children with other communication disorders and/or typically developing children. Based on a larger normative sample, standard scores for syntax, pragmatics and semantics in addition to age equivalents will be created. These scores will allow researchers and clinicians to quantify the use of spontaneous expressive language skills for children with ASD and other communication disorders, which can be compared to the scores acquired from the normative sample. Due to their primary impairments in pragmatics and social reciprocity, children with ASD may not use the range of vocabulary and grammatical constructions spontaneously in natural settings even though they can do so during highly structured testing. Therefore, the OSEL targets spontaneous expressive language that children with ASD demonstrate in less structured, more naturalistic settings. In addition, it is hoped that the quantified profiles obtained from the OSEL will provide useful information for treatment and educational programs promoting more positive outcomes for children with ASD and other communication disorders.

Table 4.1 OSEL Tasks

Tasks
1. Mr. Potato Head
2. Telling a Picture Story
3. Conversation*
4. Camping Trip
5. Throwing Game
6. Retell a Story: Where Are My French Fries?
7. Picture Description: (Balloon Vignette, Painting Vignette)

* Conversation occurs throughout the administration.

Table 4.2 Participant characteristics by age groups and gender

	Age Group 1		Age Group 2		Age Group 3		Age Group 4		Age Group 5		Age Group 6		Age Group 7	
Age Range*	24-27		28-30		31-35		36-41		42-47		48-53		54-60	
Gender	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
N	11	5	12	11	12	13	18	12	11	12	17	14	15	13
Mean (SD)														
Average age* (SD)	25.5	25.8	29.1	28.9	34	34.2	39.5	38.8	45.6	44.6	51.5	51.2	57.7	57.5
	(1.1)	(0.8)	(0.8)	(0.8)	(1.64)	(1.7)	(1.9)	(1.7)	(1.8)	(2)	(1.9)	(1.8)	(1.8)	(1.7)
SBIS ABVIQ	11	12	9.5	10.8	10.33	11	10.6	11.25	9.4	9.9	11.2	9.6	9.7	9.1
(average = 10)	(1.9)	(2.5)	(3.1)	(3.3)	(1.72)	(2.8)	(3.5)	(2.9)	(1.8)	(1.8)	(3.3)	(2.7)	(3.1)	(2.3)
SBIS ABNVIQ	9.34	9.4	8	8.5	10.92	11.5	10.9	11	10	9.5	9.5	11.6	10.5	12.4
(average = 10)	(2.9)	(4.7)	(4.3)	(3.9)	(3.29)	(2.6)	(2.9)	(2.8)	(2.5)	(2)	(2.2)	(4.4)	(2.5)	(3.8)
PLS Expressive	109	128	110.8	123.1	119.08	117.2	113.2	127.5	108.6	115.6	109.1	113.5	107.8	113.9
Communication	(19.6)	(19.7)	(20.5)	(20.1)	(12.54)	(12.9)	(19.7)	(16.5)	(8.9)	(15.7)	(11.6)	(13.7)	(10.2)	(19)
PLS Auditory	111.4	112.4	107.3	119.5	116.67	115.3	108.8	120.4	110	114.3	107.2	113.7	114.6	114.1
Comprehension	(10.8)	(20.2)	(19)	(13.4)	(14.47)	(10.3)	(14.7)	(11.6)	(12)	(13.8)	(11.8)	(11.4)	(9.6)	(16)
CASL Syntax							103.5	101	96.8	100	106.6	104.6	102.9	106
Construction							(15.4)	(12.5)	(11.4)	(16.3)	(15.1)	(15.2)	(13.2)	(21.6)
CASL Pragmatic							99.1	100.5	99.9	109.4	109.7	114.1	107.3	116
Judgment							(14.3)	(11.1)	(15.9)	(14.7)	(16.3)	(14.8)	(12.3)	(24.3)
VABS	108.6	116	106.6	106.7	115.3	100.9	106.9	107.7	101.7	110.3	104.1	109.4	102.2	107.1
Communication	(6.9)	(14.8)	(4.8)	(11.27)	(9.5)	(16)	(9.7)	(9.1)	(4.3)	(8.8)	(10)	(11.1)	(12.8)	(11.4)

*Age in months; *SBIS ABVIQ* Stanford Binet Intelligence Scale Abbreviated Battery Verbal IQ, *SBIS ABNVIQ* Stanford Binet Intelligence Scale Abbreviated Battery nonverbal IQ, *PLS* Preschool Language Scale, CASL Comprehensive Assessment of Spoken Language, *VABS* Vineland Adaptive Behavior Scales.

Table 4.3 Factor structure of the OSEL syntax items

	Factor Loadings	
Items	EFA	CFA
Responses to WH- questions	0.51	0.73
Responses to Y/N questions	0.64	0.84
Articles	0.77	0.91
Adjectives	0.77	0.87
Verbs	0.96	0.94
Nouns	0.84	0.89
Regular Past Verbs	0.77	0.85
Irregular Past Verbs	0.88	0.92
Future Tense Verbs	0.85	0.92
Progressive Tense Verbs	0.71	0.84
Modal Auxiliary Verbs	0.84	0.92
Copula Verbs	0.80	0.89
Infinitive Phrases	0.82	0.88
Negation	0.83	0.90
Regular Plurals	0.78	0.88
Irregular Plurals	0.50	0.60
Prepositions	0.66	0.79
Object Pronouns	0.74	0.87
Subject Pronouns	0.84	0.96
Possessive Pronouns	0.75	0.87
Subordination	0.88	0.94
Coordination	0.91	0.95
Longest Sentence (number of words)	0.92	0.95
Questions	0.60	0.76
	CFI	
	0.990	0.977
	RMSEA	
	0.047	0.059

EFA Exploratory Factor Analysis, *CFA* Confirmatory Factor Analysis, *CFI* Comparative Fit Index, *RMSEA* Root Mean Square Error Approximation.

Table 4.4 Factor structure of the OSEL pragmatic semantic profile items

	Factor Loadings	
	EFA	CFA
Factor 1: Initiation of Reciprocal Communication		
Verbal requests to get needs met	0.86	0.79
Asks for information about thoughts, feelings, or experiences	0.47	0.85
Comments or offers information about thoughts, feelings, or experiences	0.83	0.81
Maintains a conversation	0.92	0.97
(Absence of) Preoccupation with specific interests	0.91	0.79
Factor 2: Narrative Skills		
Repairs/Request clarification	0.45	0.86
Reports main ideas	0.97	0.83
Reports sequence of events/story	0.92	0.66
Comments on characters' emotional and/or mental states	0.38	0.72
Synthesizes cause-and-effect information	0.69	0.78
Factor 3: Unusual Features		
Interrupts the examiner or dominates conversations	0.32	0.84
Stereotyped/Idiosyncratic use of words or phrases	N/A	0.91
Unspecific language and/or semantic errors	0.37	0.76
Immediate echolalia	0.77	0.75
Impolite or inappropriate language	0.53	0.85
	CFI	
	0.995	0.996
	RMSEA	
	0.050	0.040

EFA Exploratory Factor Analysis, *CFA* Confirmatory Factor Analysis, *CFI* Comparative Fit Index, *RMSEA* Root Mean Square Error Approximation.

Table 4.5 The OSEL score distributions by age groups and gender

	Age Group 1		Age Group 2		Age Group 3		Age Group 4		Age Group 5		Age Group 6		Age Group 7	
Age Range in Months	24-27		28-30		31-35		36-41		42-47		48-53		54-60	
Gender	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
N	11	5	12	11	12	13	18	12	11	12	17	14	15	13
Mean (SD)														
OSEL syntax	30.1	49	31.5	42.4	60.3	58.9	63.3	64.7	66	69.3	68.2	77	72.3	73.5
	(12.1)	(20.9)	(10.8)	(13.8)	(11.9)	(15)	(16.7)	(17.7)	(11.2)	(12.3)	(17. 7)	(7.6)	(9.6)	(11.3)
PSP Factor 1: Initiation of	6.8	4.8	8	5.6	3.1	4	3.9	3.2	3.3	3.2	3	1.9	2.73	2
Reciprocal Communication	(1.8)	(2.3)	(1.8)	(2.7)	(1.9)	(2)	(2.2)	(2.2)	(1.6)	(2.7)	(3.1)	(1.3)	(2.2)	(1.1)
PSP Factor 2: Narrative Skills	9.9	8.6	9.3	9.6	7.8	8.1	7.7	6.2	6.2	6.7	5.8	5	4.8	4
	(2.3)	(1.1)	(1.6)	(1.9)	(3.1)	(2.9)	(3.1)	(2.5)	(2.4)	(2.9)	(2.6)	(2.2)	(2.6)	(2.7)
PSP Factor 3: Unusual	1.3	0.8	2.3	2	1	0.8	0.8	0.1	0.5	0.4	0.5	0.7	0.2	0.4
Features	(1.4)	(0.5)	(2.3)	(3.1)	(1.5)	(1.2)	(1)	(0.3)	(0.5)	(0.7)	(0.8)	(0.7)	(0.6)	(0.7)
PSP 3 Factor Total	18	14.2	19.5	17.2	11.9	12.9	12.3	9.4	9.9	10.3	9.2	7.6	7.7	6.4
	(3.9)	(2.4)	(3.1)	(5.8)	(5.4)	(5.2)	(4.5)	(4.2)	(3)	(4.7)	(5.3)	(3.2)	(4.2)	(3.1)
OSEL syntax errors	0.7	1.6	2.4	3.3	3.4	3.2	3.2	4	4.8	3.4	2.7	5.6	4.2	3.5
	(0.9)	(1.8)	(1.8)	(2.6)	(1.8)	(2.4)	(2.1)	(2.1)	(2.1)	(3.2)	(2)	(5.4)	(3.4)	(2.1)

PSP Pragmatic Semantic Profile.

Table 4.6 Age equivalents (months) corresponding to the OSEL syntax totals

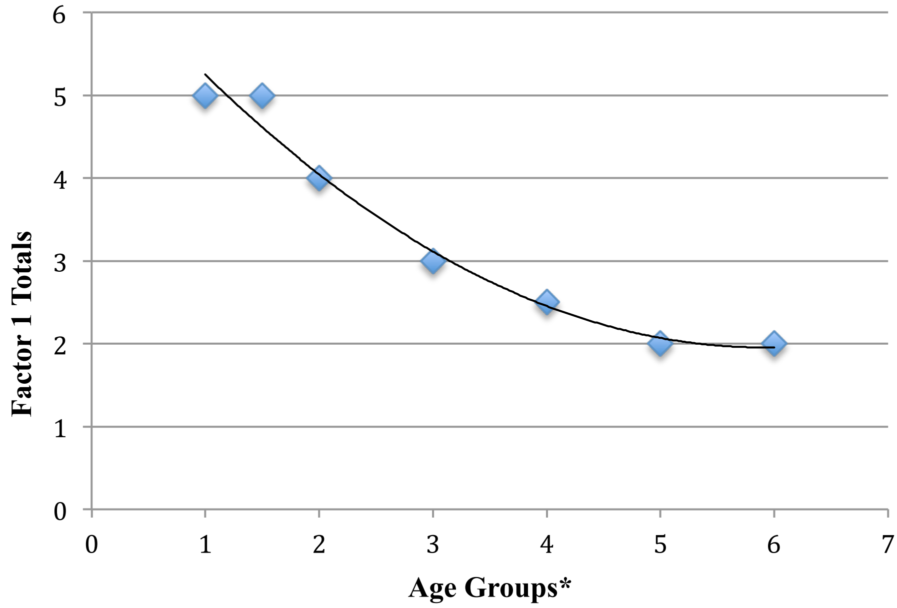
MALE		FEMALE	
Syntax Totals	**Age Equivalent**	**Syntax Totals**	**Age Equivalent**
0-29	< 24	0-34	< 24
30-34	24	35-39	24
35-38	26	40-43	26
39-42	28	44-46	28
43-45	30	47-49	30
46-48	31	50-52	31
49-50	32	53-54	32
51-52	34	55-56	34
53-54	36	57-58	36
55-58	38	59-61	38
59-61	41	62-64	41
62-63	44	65-66	44
64-66	47	67-69	47
67-68	50	70-71	50
69-70	53	72-73	53
71	58	74	58
≥ 72	≥ 61	≥ 74	≥ 61

Table 4.7 Age equivalents (months) corresponding to the OSEL pragmatic semantic profile (PSP) totals

MALE: Age Equivalents in Months					
Factor 1: Initiation of Reciprocal Communication		**Factor 2: Narrative Skills**		**3 Factor Totals: Factor 1 + Factor 2 + Factor 3**	
PSP Totals	**Age Equivalent**	**PSP Totals**	**Age Equivalent**	**PSP Totals**	**Age Equivalent**
≥8	< 24	≥11	< 24	≥19	< 24
7	25	10	25	18	25
6	29	9	32	16-17	29
4-5	33	7-8	39	15	33
3	43	6	48	12-14	39
2	55	5	57	10-11	45
0-1	> 60	0-4	> 60	8-9	52
				7	57
				0-6	> 60
FEMALE: Age Equivalent in Months					
Factor 1: Initiation of Reciprocal Communication		**Factor 2: Narrative Skills**		**3 Factor Totals: Factor 1 + Factor 2 + Factor 3**	
PSP Totals	**Age Equivalent**	**PSP Totals**	**Age Equivalent**	**PSP Totals**	**Age Equivalent**
≥6	< 24	≥11	< 24	≥16	< 24
5	27	10	25	15	25
4	33	9	29	14	29
3	39	8	33	13	33
2	52	7	39	11-12	39
0-1	> 60	6	45	9-10	45
		5	51	8	51
		4	57	7	57
		0-3	> 60	0-6	> 60

Age equivalents for Factor 3: Unusual Features were not calculated due to the limited prevalence of the scores under the factor. However, item scores under Factor 3 was included in the PSP 3 Factor Totals.

Figure 4.1 Fitting a smooth line to derive age equivalents for the PSP Factor 1 (Initiation of Reciprocal Communication) Totals for males



*Total of 7 age groups were created: Age group 1 (24-27 months), Age Group 1.5 (28-30 months), Age Group 2 (31-35 months), Age Group 3 (36-41 months), Age Group 4 (42-47 months), Age Group 5 (48-53 months), Age Group 6 (54-60 months).
Age equivalents were calculated based on the smooth line ($y= 0.1362x^2 - 1.6122x + 6.7244$); The fit of this line for the data was $R^2 = 0.978$.

Appendix A: OSEL Real time Coding Sheet

Child's name:____________

Observation of Spontaneous Expressive Language (OSEL) Coding Sheet
Fill boxes for correct use: "\" = 1 correct use; "X" = 2 correct uses; Fill Circles for error: "\" = 1 error;
Circle the sub-item for the first correct use, and use "\" and "/" for the second and third correct uses

Sentence Forms:

N+N (2) □ N+V (2) □ N+other (2) □

Coordination (4) □ □ O O
(connecting sentences; ex: …and, …or, …but)

Subordination (4) □ □ O O
(e.g. …because, …after, if…then, …(and) then, …while;
i.e. "I know what you want!" "Let's see what's in there!")

Declarative sentence (4) (e.g. I see the bird!) □ □

Imperative sentence (4) (e.g. Let's find some birds!) □ □

Wh-/How Infinitive phrases (2) □ O
(e.g. I don't know *what to* choose. I know *how to* play soccer.)

S+V agreement (6) O O O

Conjunctions (2) O
(connecting words; 1 each) and or

Reporting main ideas (**circle the elements mentioned spontaneously, underline the elements mentioned after specific questions**)
Pool: (boy jumps, water splashes, water on girl)
Frog: (frog jumps into the bowl, frog in the bowl, frog taken out)
Wake up: (mom tries to wake boy up, sister drums, boy wakes up)

Balloon: (child in tree, balloon, parents help)
Painter: (paint spilt, foot prints, paint spilling on woman)

***Conversation turns**: @+@+@+@+ (up to 8)

Clarifications to questions(2) **Clarifications to comments**(2)
Ga-Bang (bag) Gu-Rim (picture) U-San (umbrella) Bong-Too (envelope)

Questions asked by the child
Who/Where/When (3) □ □ What/Which (3) □ □
(including "what/how about…?")

Why/How (3) □ □ Y/N (3) □ □

One word questions (e.g. What/Why/What else?) (3) □ □

Questions marked only by intonation (e.g. I can do that?) (3) □ □

Response to questions

Wh: (3 each) Verbal response No response

Y/N: (3 each) Verbal response No response

Leads: (3 each) Verbal response No response

Prepositions (4) O O
(1 each) in on under behind in front of next to through
over other:

Note: note any odd phrases, echolalia, stereotyped use of language, odd intonation, articulation difficulties, impolite or inappropriate statements/questions, etc

Adjectives (6) □ □ □ O O O
(any colors/numbers are counted only up to 1)

Articles (6) O O O
(3 examples in each group) a /an /the this/that/these/those

Regular Plurals (6) □ □ □ O O O

Irregular Plurals (3) O O
(1 each) **feet teeth fish knives leaves mice geese other:**

Negations (3) □ □ O O
(e.g. n't - counted up to 2, no, never - ask "have you ever" Qs)

Gerunds as nouns (-ing) (2) □ O
(e.g. Swimming is fun!)

Subject pronouns ***I-you:*** O O O ***other***: O O O
(3 each) I you he/she one/this/that it we/they

Object pronouns O O O
(3 each) me you him/her/them/us one/this/that it

Possessive pronouns O O O
(3 each) My/mine your/yours/our/ours their/theirs/his/her/hers

General all purpose verbs (1 each)

make want do did go get look need put work come

Verb tenses:

Regular Past (-ed) (4) □ □ O O
(include regular past participles if '-ed' forms)

Irregular Past (4) □ □ O O

Progressive (v+ing) (3) □ □ O O

Future (2) O
(1 each) going to/gonna will

Verb forms:

Modal Auxiliary (3) O O
(1 each) can/could shall/should may/might will/would

Copula (4) O O
(1each) am is are was/were

Infinitive phrases (4) □ □ O O
(e.g. I want <u>to swim</u>.)

Summary codes:

Nouns ___ <5 6-10 11-15 > 15

Verbs ___ <5 6-10 11-15 > 15

Adjectives ___ <5 6-10 11-15 > 15

Longest Sentence (write out up to first 8 words)

NOT TO BE REPRODUCED IN WHOLE OR IN PART WITHOUT WRITTEN PERMISSION OF WPS Revised 1-10-2012

Appendix B: OSEL Summary Coding Table

Child's Name: **Date:**

Item (ceiling #)	**Grammatical**	**Errors**
Wh Questions answered by child:		
Response (3)		
No response (3)		
Y/N Questions answered by child:		
Response (3)		
No Response (3)		
Leads:		
Response (3)		
No response (3)		
Adjectives ≤5, 6-10, 11-15 >15		
Articles Total (6)		
a/an/the (3)		
this/that/these/those (3)		
Regular Plurals (6)		
Irregular Plurals (3)		
Negation (3)		
Gerunds (2)		
Subject Pronoun Total (18)		
I/You		
Other		
I (3)		
You (3)		
He/she (3)		
one/this/that (3)		
It (3)		
We/they (3)		
Object Pronoun Total (15)		
Me (3)		
You (3)		
Him/Her/them/us (3)		
One/this/that (3)		
It (3)		
Possessive Pronoun Total (12)		
My/mine (3)		
Your/yours (3)		
Our (3)		
Their/theirs/his/her/hers (3)		

Verb tenses:		
Regular Past (4)		
Irregular Past (4)		
Progressive (3)		
Future Total (2)		
Going to/gonna (1)		
Will (1)		
Prepositions Total (4)		
Questions Asked by the child Total (18):		
Who/Where/When (3)		
What/Which (3)		
Why/How (3)		
Y/N (3)		
One word Qs (3)		
Questions marked only by intonation (3)		
Sentence Forms:		
N+N (2)		
N+V (2)		
N+other (2)		
Coordination (4)		
Subordination (4)		
Declarative Sentences (4)		
Imperative Sentences (4)		
Wh-/How infinitive phrases (2)		
S+V agreement (6)		
Conjunctions (2)		
and (1)		
or (1)		
Modal Auxiliary Verbs Total (3):		
can/could (1)		
shall/should (1)		
may/might (1)		
will/would (1)		
Copula Verbs Total (4)		
am (1)		
is (1)		
are (1)		
was/were (1)		
Infinitive Phrase (4)		
General All Purpose Verbs Total (8)		
Other Verbs ≤5, 6-10, 11-15, >15		

Nouns ≤5, 6-10, 11-15, >15		
Conversation turns (8)		
Longest Sentence (8)		
Clarifications to comments (2)		
Clarifications to questions (2)		
Reporting main ideas for Story (3)		
Reporting main ideas for Picture Vignette (3)		

Appendix C: OSEL Pragmatic-Semantic Profile

Name of the Child: Name of the Examiner:
Date of Birth: Date of Testing:

Observation of Spontaneous Expressive Language (OSEL) Pragmatic-Semantic Profile

In addition to the child's morpho-syntactic profile, the Observation Scale of Expressive Language (OSEL) provides an opportunity to gain insight into a child's pragmatic language. The Pragmatic-Semantic Profile is divided into four different domains: Communication, Orientation to the Speaker, Narrative, and Semantic and Other Skills. Code these items WITHOUT *reference to developmental level, estimated language skills, or chronological age unless specified otherwise.*

A. Communication

The Communication domain focuses on the verbal interaction between the child and examiner with regard to the child's ability to a) flexibly take on different roles within conversations (responding and initiating) and to b) communicate for various reasons, such as to make requests, share observations and experiences, and gain information. The Communication domain should be coded based on frequency within the spontaneous language sample and not solely on the best examples.

☐ **1. a. *Verbal requests to get needs met***
This code focuses on the child's ability to verbally request to get needs met. Examples include, but are not limited to, needing assistance, or wanting to obtain objects. Do not include requests to discontinue any task or conversation.
- 0 : Frequently uses language to verbally request to get needs met.
- 1 : Uses language to verbally request but exhibits some instances in which the skill would have been expected and was not used or was not used in the amount that would be expected for expressive language level
- 2 : Occasionally uses language to verbally request but consistently exhibits instances in which the skill would have been expected and was not used or was not used in the amount that would be expected for expressive language level.
- 3 : The child rarely or never requests verbally.

☐ **1. b. *Coordination of verbal and nonverbal requests to get needs met***
This code focuses on the child's ability to combine nonverbal and verbal requests (e.g. eye contact and/or gestures with vocalizations). Examples include, but are not limited to, needing assistance, or wanting to obtain objects. Do not include requests to discontinue any task or conversation.
- 0 : Frequently uses nonverbal and verbal behaviors to request to get needs met.
- 1 : Coordinates verbal and nonverbal behaviors to request but exhibits some instances in which the skill would have been expected and was not used or was not used in the amount that would be expected for expressive language level.
- 2 : Occasionally coordinates verbal and nonverbal behaviors to request but consistently exhibits instances in which the skill would have been

expected and was not used or was not used in the amount that would be expected for expressive language level.

3: The child rarely or never coordinates verbal and nonverbal behaviors to request.

☐ **1. c. *Purely nonverbal requests to get needs met***

Code this if the majority of nonverbal requests are not combined with verbal requests (e.g. if most requests consist of eye contact and/or gestures without vocalizations). If the child combines nonverbal and verbal requests frequently, code 8. Examples include, but are not limited to, needing assistance, or wanting to obtain objects. Do not include requests to discontinue any task or conversation.

0 : Frequently requests nonverbally to get needs met.

1 : Requests nonverbally but exhibits some instances in which the skill would have been expected and was not used or was not used in the amount that would be expected for expressive language level

2 : Occasionally requests nonverbally but consistently exhibits instances in which the skill would have been expected and was not used or was not used in the amount that would be expected for expressive language level.

3 : The child rarely or never requests nonverbally.

8 : The majority of child's requests are verbal, with or without nonverbal behaviors.

☐ **2. a. *Asks for information about thoughts, feelings, or experiences***

The focus of this item is on the child's spontaneous expression of interest in the examiner's ideas, knowledge, experiences, or reactions.

0: Asks the examiner about his/her thoughts, feelings, or experiences that are not related to preoccupations or circumscribed interests on several occasions.

1: Occasionally (at least one clear example) asks the examiner about his/her thoughts, feelings, or experiences that are not related preoccupations or circumscribed interests.

2: Responds appropriately to examiner's comments about his/her thoughts, feelings, and experiences, but does not spontaneously inquire about them and/or only asks information related to preoccupations or circumscribed interests.

3: Does not respond to examiner's comments about his/her thoughts, feelings, and experiences or express interest in them, even about preoccupations or circumscribed interests.

☐ **2. b*. Asks for information about non-personal facts***

The focus of this item is on the child's spontaneous expression of interest in the OSEL materials or about non-personal facts (e.g. weather, furniture in the room, outside noises, camera).

0: Asks the examiner about non-personal facts that are not related to preoccupations or circumscribed interests on several occasions.

1: Occasionally (at least one clear example) asks the examiner about non-personal facts that are not related to preoccupations or circumscribed interests.

2: Responds appropriately to examiner's comments about non-personal facts, but does not spontaneously inquire about them and/or only asks information related to preoccupations or circumscribed interests.
3: Does not respond to examiner's comments about non-personal facts or express interest in them, even about preoccupations or circumscribed interests.

☐ **3. a. *Comments or offers information about thoughts, feelings, or experiences***
The focus of this item is on the child's spontaneous, appropriate offering of personal information, new to the examiner. It does not have to occur in context or be part of a sustained interaction. It can occur as an elaboration or response to questions, but must include new information not specified by the question. It can be related to the child's interests, but should not be related solely to preoccupations or circumscribed interests. If a child meets criteria for a "0," "1," or "2" code and comments about his/her own preoccupations/circumscribed interests, still code "0," "1," or "2."
0: Spontaneously offers information about his/her own thoughts, feelings, or experiences on several occasions.
1: Occasionally offers information spontaneously about his/her own thoughts, feelings, or experiences.
2: Only offers information about facts or general knowledge (not including preoccupations or circumscribed interests).
3: Rarely or never offers information spontaneously, except about circumscribed interests or preoccupations.

☐ **3. b. *Comments or offers information about non-personal facts***
The focus of this item is on the child's spontaneous, appropriate offering of information about non-personal facts (e.g. OSEL materials, weather, furniture in the room). *It does not have to occur in context or be part of a sustained interaction. It can occur as an elaboration or response to questions, but must include new information not specified by the question. It can be related to the child's interests, but should not be related solely to unusual preoccupations or unusually intense circumscribed interests.*
0: Spontaneously offers information about non-personal facts on several occasions.
1: Occasionally offers information spontaneously about non-personal facts.
2: Only offers information about related to preoccupations or circumscribed interests.
3: Rarely or never offers information spontaneously, even about circumscribed interests or preoccupations.

B. Orientation to the Speaker

The Orientation to the Speaker domain addresses the quality of the child's conversational skills, especially whether the child is able to carry on back-and-forth conversations and the extent to which conversations are reciprocal.

☐ **4. *Maintains a conversation***

This code is focused on the child's ability to build on to what the examiner says in order to continue a conversation.

0: The child is able to build a conversation, offering information and asking about the examiner's remarks. This rating requires that much of the child's speech provides both a response and some additional talking that builds on what has just been said and allows a response from the examiner. The conversation flows and requires no or minimal effort on the part of the conversational partner to keep it going over multiple turns.

1: The child is occasionally able to continue conversations in a way that the child is able to respond to the examiner's questions AND provide leads for the examiner to follow. However, the child exhibits some instances in which the skill would have been expected and was not used or was not used in the amount that would be expected for expressive language level.

2: Code a "2" if the child is only able to talk about his/her interests or only responds to the examiner's questions AND does not add any information spontaneously AND does not provide leads for the examiner to follow.

3: The child rarely or never carries on conversations with the examiner, even about favorite topics, and does not consistently respond to questions. The child may follow his/her own train of thought rather than participate in an interchange; may have some spontaneous offering of information or comments, but little sense of reciprocity.

☐ **5. *Preoccupation with specific interests***

0: The child is able to have conversations about multiple topics beyond his/her preferred interests and is able to flexibly move to different topics without redirecting the conversation back to special interests.

1: The child is able to talk about things outside of his/her special interests but occasionally changes the topic of a conversation or initiates a conversation about specific interests more frequently than most children of the same language level.

2: The child is occasionally able to talk about things outside of his/her special interest but often changes the topic of a conversation or initiates a conversation about specific interests more frequently than a child of the same language level.

3: The child talks about his/her special interests and on rare occasions is able to have a conversation about other topics.

8: Code 8 if the child is not able to hold a conversation with the examiner (i.e. received a code of 2 or 3 on the preceding Conversation item.)

☐ **6.** ***a. Interrupts the examiner***

This code focuses on whether the child interrupts the examiner frequently, which may make it difficult for the conversation to be truly reciprocal.

0: The child rarely interrupts the examiner.
1: The child occasionally interrupts the examiner.
2: The child frequently interrupts the examiner but it is not difficult for the examiner to give an instruction, describe an event or make a statement that requires several sentences.
3: The child frequently interrupts the examiner such that it is difficult for the examiner to give an instruction, describe an event or make a statement that requires several sentences.
8: Code 8 if the child is not able to hold a conversation with the examiner (i.e. received a code of 2 or 3 on the preceding Conversation item.)

☐ **6. b.** ***Dominates conversations***

This code focuses on the balance of conversational turns between the child and the examiner and whether conversations are dominated or controlled by the child. In a balanced, reciprocal conversation, the examiner should be able to interrupt or redirect the child. In an unbalanced conversation, the child may provide excessive amounts of description and/or detail into which it is difficult for the examiner to insert himself or herself and which makes the conversation one-sided.

0: It is not difficult for the examiner to add to the conversation or change topics. The child does not excessively direct the interchange with the examiner to an unusual degree.
1: Occasionally it is difficult for the examiner to add to the conversation or change topics. The child may control the conversation by offering excessive details and asking repeated questions occasionally.
2: The child frequently tries to direct the conversation by telling others what to say asking frequent repeated questions, or adding excessive details. On multiple occasions, the examiner may have difficulty interrupting the child, but when interrupted, the child is able to yield the floor to the examiner for a brief period of time.
3: The conversation is mostly dominated and controlled by the child such that the majority of conversation is one sided and rarely reciprocal. Code a "3" if the child is unable to yield the floor to the examiner even when explicitly interrupted.
8: Code 8 if the child is not able to hold a conversation with the examiner (i.e. received a code of 2 or 3 on the preceding Conversation item.)

7. Repairs/Request clarification

This code focuses on the child's ability to repair or request clarifications for the unfamiliar words that are used in the OSEL by the examiner. In order to receive the full credit, the child needs to ask a specific question to clarify the words that the examiner mentions (e.g. what is "Usan?"). Any clarifications that occur in other contexts can also be coded here.

0: The child effectively repairs examiner's unclear or incorrect questions/comments or requests clarification if he/she does not understand the examiner. Must include at least 1 clarification of an examiner's comment AND 1 clarification of an examiner's question.

1: The child effectively repairs examiner's incorrect questions/comments or requests clarification if he/she does not understand the examiner on at least one occasion.

2: The child attempts to repair examiner's incorrect questions/comments or request clarification if he/she does not understand the examiner on at least one occasion, but this is not clear or completely effective (e.g. what?).

3: The child does not repair or request clarification.

C. Narrative Skills

The Narrative Skills domain focuses on the child's ability to tell and re-tell stories during the OSEL with the pictures and props provided, as well as on the child's reports of events and stories during conversation.

8. ***Reports main ideas***

Main ideas should include ALL MAIN elements for each story during Telling a Picture Story and Picture Description (See Coding Sheet). DO NOT INCLUDE MAIN IDEAS REPORTED DURING "RETELL THE STORY: WHERE ARE MY FRENCH FRIES?" FOR CODE 0 or 1. If the examiner had to present different pictures because the child did not show interests in the initially presented picture story or vignette, code the best examples (except for code 0).

0: The child is able to spontaneously state the main ideas of the story correctly in all picture stories/vignettes that were presented *initially*; should mention 3 elements for BOTH *initially* presented tasks.

1: The child spontaneously states at least 2 elements for any pictures presented for BOTH tasks OR some contextual cues or prompting is required by the examiner to get the main ideas AND/OR there is one clear misunderstanding of a main idea in addition to correct reporting of other main ideas.

2: The child correctly describes at least 1 idea from all stories OR 2 ideas for 1 story only OR the child is able to state 3 main ideas of the story correctly only during the Retell the Story: Where are my French Fries? task.

3: The child is unable to state the main idea for any of the stories presented in the OSEL nor does he/she report events throughout the assessment that include main ideas.

☐ ***9. Reports sequence of events/story***

This can be coded throughout different tasks, Telling a Picture Story, Picture Description, and Conversation. DO NOT INCLUDE REPORTING SEQUENCE OF EVENTS/STORY REPORTED DURING "RETELL THE STORY: WHERE ARE MY FRENCH FRIES?" FOR CODE 0 or 1. If the examiner had to present different pictures because the child did not show interests in the initially presented picture story or vignette, code the best examples (except for code 0).

0: The child is spontaneously able to appropriately sequence the stories presented *initially* and sequence ideas during conversation so that the examiner can follow along.

1: The child correctly sequences most stories presented or sequences in conversation (at least 1 clear example) but some prompting is necessary.

2: The child is able to appropriately sequence in at least one story or one conversation with examiner's prompting, but confuses sequence in at least one other example OR only able to appropriately sequence the story during the Retell the Story: Where are my French Fries? task.

3: The child is unable to appropriately sequence stories presented or provides multiple examples of confused sequences.

☐ ***10. Comments on characters' emotional and/or mental states***

0: The child spontaneously and correctly comments about several different emotional (e.g. sad, happy, angry) and/or mental (e.g. confused, surprised) states of characters presented in the story tasks and/or comments on the emotional and/or mental states of others during interactions and conversations with the examiner.

1: The child makes some spontaneous comments about an emotional and/or mental state of the characters or others in the story tasks or in conversation (i.e. at least 1 clear example).

2: Code "2" if the child is only able to comment on the characters' facial expression(s) such as crying, smiling and/or actions related to emotional states (e.g. running away, hiding) AND/OR the child incorrectly identifies emotional and/or mental states.

3: The child does not identify emotional nor mental states unless prompted.

☐ ***11. Synthesizes cause-and-effect information***

This item focuses on the child's descriptions of cause-and-effect relationships, using information from the pictures or tasks. The child has an opportunity to do this in story tasks presented in the OSEL such as Telling a Picture Story and Picture Description as well as during conversation in which an event or personal narrative is reported. DO NOT INCLUDE SYNTHESIZING CAUSE-AND-EFFECT INFORMATION DURING "RETELL THE STORY: WHERE ARE MY FRENCH FRIES?" FOR CODE 0 or 1. If the examiner had to present different pictures because the child did not show interests in the initially presented picture story or vignette, code the best examples (except for code 0).

0: The child spontaneously conveys cause-and-effect relationships correctly in more than one picture story and vignette presented *initially,* or in conversation.

1: The child is able to spontaneously portray at least one cause-and-effect relationship in any of the pictures presented. Some prompting may be required to understand the plot of the story.

2: The child makes comments about a story or picture, but may list information without apparent relevance to a plot OR show some misunderstanding of the cause-and-effect relationships OR the child is able to synthesize cause-and-effect information only during the Retell the Story: Where are my French Fries? task.

3: The child does not provide any comments/plot about the stories or during conversations.

D. Semantic and Other Aspects

☐ ***12. Stereotyped/Idiosyncratic use of words or phrases***

Coding for this item includes delayed echolalia or other highly repetitive utterances with consistent intonation patterns, as well as the use of words or phrases that are inappropriately formal. These words or phrases can be intended meaningfully and can be appropriate to conversation at some level. The focus of the item is on the stereotyped or idiosyncratic quality of the phrasing, unusual use of words or formation of utterances, and/or their arbitrary association with a particular meaning.

0: Rarely or never uses stereotyped or odd words

1: Use of words or phrases tends to be more repetitive or formal than that of most individuals at the same level of expressive language, but not obviously odd, OR occasional stereotyped utterances or odd use of words or phrases, with substantial spontaneous language, as well.

2: Often uses stereotyped utterances or odd words or phrases with other language.

3: Mostly uses stereotyped utterances or odd words or phrases without other language.

13. *Unspecific language and/or semantic errors*
This item focuses on the child's ability to communicate content. That is, is the child's vocabulary sufficient in size and variety to convey specific messages, or does the child tend to use unspecific language (e.g. says "that thing" instead of naming an object or uses general purpose verbs, such as make or have, instead of specific verbs)? Frequent groping for a word, false starts (not referring to dysfluency of speech), as well as semantic errors (e.g. spoon for knife) may also be indicative of word finding difficulties and limited vocabulary skills.

0: The child typically uses appropriately specific language without groping for words so that it is usually clear what he/she is talking about.
1: The child sometimes has difficulties finding words AND/OR sometimes makes false starts but usually finds the appropriate words eventually.
2: The child sometimes uses unspecific language so that the examiner is not always positive to what or whom the child is referring without some use of contextual cues, or requests for clarification.
3: The child frequently uses unspecific language so that the examiner is not clear about to what or whom the child is referring, such that the examiner frequently needs to guess about intended messages even after requests for clarification.

14. Immediate echolalia
This item pertains to the child's immediate repetition of the last statement, series of statements, or last few words of the examiner. When coding, do not include repetitions that are a lead-in to a response to the examiner or that are used as a memory device in specific tasks.

0: Rarely or never repeats others' speech.
1: Occasional echoing.
2: Echoing words and phrases regularly with much spontaneous language as well.
3: Echoing words and phrases consists of a significant proportion of utterances.

15. Impolite or inappropriate language
This item focuses on a child's use of language that seems inappropriate for the social situation, including language that is rude or cheeky (i.e. "you have a zit") that seems to indicate a lack of awareness for social cues or situations.

0: The child does not use any impolite or inappropriate language.
1: The child is sometimes impolite or inappropriate.
2: The child is frequently impolite or inappropriate but it does not interfere with the interaction.
3: The child is repeatedly impolite or inappropriate such that it interferes with the interaction.

☐ **16. Level of support required for conversation**.

The focus of this item is on whether the child is able to have conversations without the use of objects to support the interaction or if materials are needed for him/her to carry on a conversation. Conversations relating to any of the OSEL materials may be included, but tasks often eliciting conversation around the materials include Picture Story, Picture Vignette, and Conversation. Conversations do not need to be initiated by the participant to be coded here.

0: The child is able to carry on several conversations with the examiner. These may include conversations with materials present, as well as conversations that are initiated in the presence of materials, but extend beyond the objects that are physically present (e.g., during Picture Story, the participant points to the swimming pool and then tells the examiner about his/her past trip to a swimming pool). However, there must be at least ONE clear example of a conversation about a topic that is not part of his/her special interests AND that is unrelated to the OSEL materials.

1: The child is able to carry on several conversations, but all conversations are related to materials that are physically present. These conversations may extend beyond the materials (e.g., during Picture Story, the participant points to the swimming pool and then tells the examiner about his/her past trip to a swimming pool), but are prompted by the presence of the materials.

2: The child is able to carry on several conversations, but only with materials present (e.g., during the Camping Trip activity, the child indicates that s/he likes camping and s/he wants to eat grapes, but the conversation does not extend beyond the objects that are physically present) OR all conversations that occur without the support of materials are related to circumscribed interests or highly specific interests.

3: The child is able to carry on a conversation with materials present, but only in ONE situation (e.g., Conversation Task) OR s/he is able to respond to the examiner's initiations when materials are present, but does not build on the conversation.

8: Conversations too limited to judge OR No conversations (i.e. item 4 scored a 2 or 3).

☐ **17. Intonation/volume/rhythm/rate**.

The focus of this item is on speech abnormalities related to intonation, volume, rhythm, and rate. Code this item relative to the child's expressive language level. Abnormal speech patterns typically associated with general language delay should be assigned a rating of 0. Odd non-speech sounds are not coded here.

0: Appropriately varying intonation, reasonable volume, and normal rate of speech, with regular rhythm coordinated with breathing.

1: Little variation in pitch and tone; rather flat or exaggerated intonation, but not obviously peculiar, OR slightly unusual volume, AND/OR speech that tends to be somewhat unusually slow, fast, or jerky.

2: Speech that is clearly abnormal for ANY of the following reasons: slow and halting; inappropriately rapid; jerky and irregular in rhythm (other than ordinary stutter/stammer); odd intonation or inappropriate pitch and stress; markedly flat and tone-less ("mechanical"); consistently abnormal volume.

3: Speech that is difficult to understand because of one or more speech abnormalities as specified above.

☐ **18. Intelligibility**

The focus of this time is on the intelligibility of the child's speech. The examiner may experience difficulties understanding the child's speech due to articulation problems, stutter, stammer or other fluency disorder.

0: No articulation difficulties, stutter, stammer, or other fluency disorder are noted.

1: Mild articulation difficulties, stutter, stammer, and/or other fluency disorder are noted but the examiner rarely has difficulty understanding the child's speech.

2: Moderate articulation difficulties, stutter, stammer, and/or other fluency disorder are noted and the examiner may have difficulty understanding the child's speech.

3: Severe articulation difficulties, stutter, stammer, and/or other fluency disorder are noted which such that the examiner may have difficulty understanding the majority of the child's speech.

References

Alarcon, M., Canton, R.M., Liu, J., Gilliam, T.C., Geschwind, D.H., Autism Genetic Research Exchange Consortium. (2002). Evidence for a language quantitative trait locus on chromosome 7q in multiplex autism families. *American Journal of Human Genetics, 70,* 60-71.

Barrett, S., Beck, J.C., Bernier, R., Bisson, E., Braun, T.A., Casavant, T.L., Childress D., et al (1999) An autosomal genomic screen for autism: Collaborative linkage study of autism. *American Journal of Medical Genetics, 88*, 609-615

Bartolucci, G., Pierce, S. J., & Streiner, D. L. (1980). Cross-sectional studies of grammatical morphemes in autistic and mentally retarded children. *Journal of Autism and Developmental Disorders, 10*(1), 39-50.

Bauer, D. J., Goldfield, B. A., & Reznick, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary development. *Applied Psycholinguistics* , *23*, 313-335.

Bishop, D.V.M. (2002). Putting language genes in perspective. *Trends in Genetics, 18*(2), 57-59.

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*(7), 917-929.

Bradford, Y., Haines, J., Hutcheson, H., Gardiner, M., Braun, T., Sheffield, V., et al. (2001). Incorporating language phenotypes strengthens evidence of linkage to autism. *American Journal of Medical Genetics, 105*(6), 539-47.

Brown, R. (1973). *A first language: The early stages*. Oxford England: Harvard U. Press.

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

Condouris, K., Meyer, E., & Tager-Flusberg, H. (2003). The relationship between standardized measures of language and measures of spontaneous speech in children with autism. *American Journal of Speech-Language Pathology, 12,* 3-15.

Farrell, A. D., Mariotto, M. J., Conger, A. J., Curran, J. P., & Wallander, J. L. (1979). Self-Ratings and judges' ratings of heterosexual social anxiety and skill: A generallizability study. *Journal of Consulting and Clinical Psychology, 47*, 164-175.

Fenson, L., Pethick, S.J., Renda, C., Cox, J.L., Dale, P.S., & Reznick, J.S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics, 21,* 95-115.

Galsworthy, M. J., Dionne, G., Dale, P. S., & Plomin, R. (2000). Sex differences in early verbal and non-verbal cognitive development. *Developmental Science, 3*, 206-215.

Goffman, L., & Leonard, J. (2000). Growth of language skills in preschool children with specific language impairment: Implications for assessment and intervention. *American Journal of Speech-Language Pathology, 9,* 151-161.

Gillberg, C., & Steffenburg, S. (1987). Outcome and prognostic factors in infantile autism and similar conditions: A population-based study of 46 cases followed through puberty. *Journal of Autism and Developmental Disorders, 17*(2), 273-287.

Harris, G., Chabris, C., Clark, J., Urban, T., Aharon, I., Steele, S., et al., (2006). Brain activation during semantic processing in autism spectrum disorders via functional magnetic resonance imaging. *Brain and Cognition, 61,* 54-68.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry, 45*(2), 212-229.

Kasari, C., Gulsrud, A., Wong, C. Kwon, S., & Locke, J. (2010). Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of Autism and Developmental Disorders, 40,* 1045-56.

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

Leonard, L. (1998). *Children with specific language impairment.* Cambridge, MA: The MIT Press.

Lord, C. & Corsello, C. (2004). Diagnostic instruments in autism spectrum disorders. In F. Volkmar, A. Klin, R. Paul, & D. Cohen (Eds.). *Handbook of Autism and Pervasive Developmental Disorders* (pp. 1-23). New York:Wiley.

Lord, C., Risi, S., Lambrecht, L., Cook, E., & Leventhal, B. (2000). The Autism Diagnostic Observation Schedule-Generic: a standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders, 30*(3), 205-223.

Lord, C., Risi, S., & Pickles, A. (2004). Trajectory of language development in autistic spectrum disorders. In S. F. Warren (Ed.), *Developmental Language Disorders:*

From Phenotypes to Etiologies. (pp. 7-29). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Lord, C., Schulman, C., & DiLavore, P. (2004). Regression and word loss in autistic spectrum disorders. *Journal of Child Psychology and Psychiatry, 45,* 936-955.

Muthen, L. K., Muthen, B. O. (1998). *M-plus User's Guide.* Los Angelos, CA: Muthen and Muthen.

Pettit, G. S., McClaskey, C. L., Brown, M. M., & Dodge, K. A. (1987). The generalizability of laboratory assessments of children's socially competent behavior in specific situations. *Behavioral Assessments, 9,* 81-96.

Rescorla. L., Roberts, J., & Dahlsgaard, K.(1997). Late talkers at 2: Outcome at age 3. *Journal of Speech and Hearing Research. 40*, 556–566.

Reynell, J. & Gruber, C. (1990) *Reynell Developmental Language Scales*. Los Angeles: Western Psychological Services.

Rogers, S. J. (2005). Evidence-based practices for language development in young children with autism. In T. Charman & W. Stone (Eds.), *Social and Communication Development in Autism Spectrum Disorders* (pp. 143–179). New York: Guilford.

Rutter, M., Le Couteur, A., Lord, C., MacDonald, H., Rios, P., & Folstein, S. (1988). Diagnosis and subclassification of autism: Concepts and instrument development. In E. Schopler & G. B. Mesibov (Eds.), *Diagnostic and Assessment Issues in Autism* (pp. 239-260). New York: Plenum Press.

Scarborough, H., Rescorla, L., Tager-Flusberg, H., Fowler, A., & Sudhalter, V. (1991). The relation of utterance length to grammatical complexity in normal and language disordered groups. *Applied Psycholinguistics, 12,* 23-45.

Semel, E., Wiig, E.H. & Secord, W.A. (2003). *Clinical Evaluation of language fundamentals-Fourth Edition (CELF-4).* San Antonio: The Psychological Corporation.

Smith, T., Groen, A. & Wynn, J. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal of Mental Retardation, 105*(4), 269-85.

Spence, S. J., Cantor, R. M., Chung, L., Kim, S., Geschwind, D. H., & Alarco´n G. (2006). Stratification Based on Language-Related Endophenotypes in Autism: Attempt to Replicate Reported Linkage. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)*, 141B, 591–598.

Tager-Flusberg, H., & Calkins, S. (1990). Does imitation facilitate the acquisition of grammar? Evidence from a study of autistic, down's syndrome and normal children. *Journal of Child Language, 17*(3), 591-606.

Tager-Flusberg, H., Rogers, S., Cooper, J., Landa, R., Lord, C., Paul, R., et al. (2009). Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. *Journal of Speech, Language, and Hearing Research, 52*(3), 643-652.

Tager-Flusberg, H., Skwerer, D.P., Joseph, R.M. (2006). Model syndromes for investigating social cognitive and affective neuroscience: a comparison of Autism and Williams syndrome. *Social Cognitive and Affective Neuroscience, 1*(3):175-82.

Tomblin, J.B., Zhang, X., Weiss, A., Catts, H., & Weismer, S.E. (2004). Dimensions of individual differences in communication skills among primary grade children. In M.L. Rice and S.F. Warren (Eds.), *Developmental Language Disorders: From Phenotypes to Etiologies* (pp. 53-76). Mahwah, NJ: Lawrence Erlbaum Associates.

Venter, A., Lord, C., & Schopler, E. (1992). A followd-up study of high-functioning autistic children. *Journal of Child Psychology and Psychiatry, 33*(3), 489-507.

Walenski, M., Tager-Flusberg, H., & Ullman, M. T. (2006). Language in autism. In J. L. R. Rubenstein (Ed.), *Understanding Autism: From Basic Neuroscience to Treatment.* (pp. 175-203). Boca Raton, FL US: CRC Press.

Ward, A., Stoker, H. W., & Murray-Ward, M. (1996). *Educational Measurement: Theories and Applications.* Lanham, MA, US: University Press of America, Inc.

Chapter V

Conclusion

Over the past few decades, diagnostic instruments designed to capture the early signs of autism in toddlers and young preschoolers have contributed to the identification of very young children with autism spectrum disorders (ASD). Effective and appropriate assessment of early signs of autism is associated with early provisions of services and treatments for young children with ASD. However, even though efforts to describe autism symptoms in young children have grown dramatically, they have been continually compromised by lower sensitivity and specificity of diagnostic instruments for toddlers and young preschoolers with ASD compared to older children with ASD. Continued advances in diagnostic practices and descriptive capabilities are needed to more accurately differentiate children with ASD from other developmental disorders (e.g., language delays, intellectual disabilities) at young ages.

The first two studies of this dissertation suggest ways to maximize our ability to validly differentiate young children with ASD from those with developmental disorders using existing gold standard diagnostic instruments. The information gained using the methods suggested in these studies for the early detection of ASD could help clinicians to effectively decide whether the child should be followed up in future assessments and enter into treatments. The ability to validly identify and describe early features of ASD will also contribute to a more accurate and effective stratification of samples in research studies, including those examining genetic and neurobiological etiology of ASD.

The assessment of language impairments is also a crucial part of the identification of clearly defined behaviors that are necessary for provisions of services and treatments for children with ASD. However, there have not generally been standardized instruments that measure spontaneous expressive language of children with ASD in a relatively naturalistic setting. Most instruments currently used for the assessment of language development focus on measuring pre-determined responses by asking a child to answer specific questions or fill in blanks or label pictures or objects. Therefore, the third study of this dissertation focuses on developing a new measure for children with ASD and other communication disorders from 2 to 12 years of age for the valid description of spontaneous language use in a standardized, but naturalistic, setting.

The results from these studies have important implications for treatment outcome, genetic, and neuroimaging studies. First, the first two studies have expanded the valid use of diagnostic instruments to children as young as 12 months of age. Second, valid phenotyping using empirically validated methods and measures developed in these studies can provide uniform measurement approach in treatment studies to monitor changes in autism symptoms and provide ways to measure language skills for genetic and neuroimaging studies. Valid phenotyping using these empirically validated methods and measures may also inform programming intervention goals for children with ASD.

Conceptualizations of ASD are highly dependent on how behaviors are measured by different instruments. Therefore, as the standard diagnostic and language instruments for children with ASD become more refined, we will be able to improve our understanding of the behavioral manifestations of ASD. The three studies that comprise this dissertation reflect progress toward a more valid and effective description of autism symptoms and of deficits in spontaneous expressive language for children from 2 to 12 years of age. Further research on these topics will inform our use and refinement of new measurement techniques and instruments described herein. Future studies in this area will also extend our understanding of the early behavioral manifestations and language deficits in children with ASD and other communication disorders.