TOWARD A BETTER UNDERSTANDING AND IMPROVED VALIDITY OF AUTISM SYMPTOM MEASURES ACROSS THE LIFESPAN

by

Vanessa Hus Bal

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in The University of Michigan
2014

Doctoral Committee:

Emeritus Professor Albert Cain, Co-Chair
Professor Catherine Lord, Weill-Cornell Medical College, Co-Chair
Associate Professor Donna M. Martin
Associate Professor Christopher S. Monk

DEDICATION

To my husband, Brian, who has showered me with love and support throughout all of my academic endeavors. And to my advisor, Cathy Lord, whose confidence in me inspired me to accomplish things I wouldn't have otherwise thought I could.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institute of Mental Health (R01MH081873 and RC1MH089721) and the Simons Foundation awarded to Catherine Lord, as well as a Simons Foundation graduate fellowship, Autism Speaks Dennis Weatherstone Predoctoral Fellowship, Naomi Lohr Fellowship, and Rackham Doctoral Research Grant awarded to Vanessa Hus Bal.

I am indebted to Cathy Lord, my advisor and mentor, who has been an unprecedented model of a true clinical scientist, which I will always look to for inspiration throughout my career. Under her mentorship, the first and second studies (Chapter II and III) were published in the Journal of Child Psychology & Psychiatry and the Journal of Autism & Developmental Disorders, respectively, in 2013. The third study (Chapter IV) has been accepted pending minor revisions by the Journal of Autism & Developmental Disorders.

I gratefully acknowledge the many individuals and families who participated in this research; it is their stories that give this work meaning and that fuel my passion for research. I would like to express my deepest gratitude to Al Cain for his support, which has not wavered since the time I first met him as an undergraduate student with aspirations to become a Clinical Psychologist. I also want to thank my "UMACC" colleagues, particularly Somer Bishop, Liz Buvinger, Themba Carr, Marisela Huerta and Katherine Gotham, and Sophy Kim, who have provided both mentorship and friendship throughout the years. I am also thankful for the many staff in North Carolina, Illinois,

Michigan and New York, who were involved in data collection and curation. In particular, I am grateful to Shanping Qiu for her good-natured willingness to fulfill last minute data requests; Nicole Saghy for attention to details of the SSC data; Emma Salzman for her support of the Weatherstone study throughout the UMACC-CADB transition; and Andrew Pickles for ongoing statistical consultation. I am also thankful to the Septaphiles, with whom I have had the honor of sharing the past 5 years. Finally, I would like to express my thanks to Chris Monk and Donna Martin for serving on my committee and providing valuable critiques to improve this work.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

ABSTRACT

Autism Spectrum Disorders (ASDs) are defined by impairments in social-communication skills and restricted, repetitive and stereotyped patterns of behaviors and interests. The presentation of symptoms is affected by a variety of factors not specific to ASD, such as developmental characteristics (e.g., age, language level, and cognitive ability) and co-occurring dimensions of behavior (e.g., internalizing and externalizing behaviors). With the shift toward a dimensional diagnostic system put forth by the DSM-5, research is needed to explore the influence of non-ASD-specific factors on scores from widely-used ASD diagnostic instruments. This research will inform the development of new measures that take into account the interaction between ASD symptoms and non-ASD-specific dimensions of behavior in order to provide more appropriate quantitative measures of ASD symptoms.

This three-study dissertation seeks to expand the valid use of pre-existing ASD diagnostic measures with individuals across a range of ages. Study One provides a systematic look at the influences of developmental characteristics and non-ASD-specific dimensions of behavior on interpretation of scores from the Social Responsiveness Scale (SRS), a parent questionnaire widely used as an index of ASD severity. The second two studies focus on the Autism Diagnostic Observation Schedule (ADOS). Study Two standardizes Social Affect and Restricted, Repetitive Behavior domain scores to provide separate measures of social-communication and repetitive behavior severity that are less influenced by developmental level. Study Three revises the diagnostic algorithm and

calibrates scores for ADOS Module 4, used with verbally fluent older adolescents and adults. This increases Module 4's comparability to other modules used with younger or more language impaired individuals.

Overall, the results of these studies provide a more in-depth understanding of how ASD diagnostic measures are influenced by developmental characteristics and non-ASD-specific dimensions of behavior. This knowledge can inform use of these measures as quantitative indices of ASD symptom severity. In clinical settings they may be used to monitor treatment progress. Application in the research domain may facilitate exploration of links between biological mechanisms and behavior and predictors of adult functioning, which will hopefully inform development of targeted interventions to promote positive outcomes for individuals with ASD.

CHAPTER I

Introduction

First described by Leo Kanner in 1943, the category of *Infantile Autism* did not appear in the Diagnostic and Statistical Manual of Mental Disorders until 1980 (DSM-III; American Psychiatric Association [APA]). At that time, autism was defined by four criteria, including a lack of responsiveness to others, language delay, unusual patterns of speech and "bizarre" responses to the environment, each which must occur before 30 months of age. Infantile autism was also differentiated from schizophrenia by the absence of positive symptoms (e.g., delusions, hallucinations). In subsequent revisions of the DSM, *Infantile Autism* was replaced by *Autistic Disorder (AD)* and listed under the category of "Pervasive Developmental Disorders" (PDD). Diagnostic criteria were expanded to include more concrete and observable symptoms in three domains of behavior: qualitative impairments in social interaction, qualitative impairments in communication, and presence of restricted, repetitive and stereotyped patterns of behaviors, interests and activities (DSM-IV-TR; APA, 2000). *AD* was distinguished from other PDDs of *Asperger's Disorder (AS), Pervasive Developmental Disorder - Not Otherwise Specified (PDD-NOS), Child Disintegrative Disorder (CDD)* and *Rett Syndrome*. Criteria for differentiation of these syndromes were based on factors such as age of onset of symptoms, presence or absence of language delay, and occurrence and timing of developmental regression.

More recently, however, the DSM-5 (APA, 2013) has taken on a more dimensional approach that has yielded considerable changes to the classification of this family of disorders. A single category of *Autism Spectrum Disorder (ASD)* now subsumes the previously differentiated syndromes. In addition, the triad of defining symptoms has been collapsed into two domains characterizing ASDs: *Deficits in Social Communication and Social Interaction* and *Restricted, Repetitive Patterns of Behavior, Interests or Activities* (APA, 2011). These changes have been implemented in response to a large body of research demonstrating that distinctions made among the different syndromes are unreliable across sites and frequently associated with child characteristics, such as IQ or language level, rather than the differential presentation of ASD symptoms (e.g., Lord et al., 2012). Research has also indicated that communication and social behaviors comprise a single construct (e.g., Gotham, Risi, Pickles, & Lord, 2007; Lord, Rutter, DiLavore, & Risi, 1999; Robertson, Tanguay, L'Ecuyer, Sims, & Waltrip, 1999). Acknowledging the vast variability in manifestation of symptoms observed across individuals, DSM-5 also provides guidelines for indications of severity levels for both behavioral domains (from requiring support to requiring very substantial support) and diagnostic specifiers to indicate other variations that may influence symptom presentation (e.g., intellectual impairment, language impairment, known medical or genetic condition or environmental factor).

Early epidemiological studies suggested that ASDs were rare. Based on 18 international studies published between 1966 and 1993, Fombonne (2005) reported a median prevalence for AD of 4.7 per 10,000 children. Examination of studies published between 1994 and 2004 revealed a clear increase in prevalence of AD, with a median of

12.7 per 10,000 children; this number increased to 36.4/10,000 when combining estimated prevalence of AD, AS and PDD-NOS. In a review of 27 studies from 9 different countries published between 2000 and 2010, Fombonne and colleagues (2011) reported an even higher prevalence, estimating that approximately 70/10,000, or 1 in 143 children, were diagnosed with an ASD. In a more recently published epidemiological study, based on data collected in 2008, the prevalence of ASDs among 8-year-olds in the United States was estimated to be 1 in 88 using the DSM-IV criteria (Autism and Developmental Disabilities Monitoring Network [ADDMN] & Center for Disease Control and Prevention [CDC], 2012). The ADDMN & CDC highlight that this estimate reflects a 23% increase from their 2006 estimate of 1 in 110 children, and a 78% increase from 2002, when they estimated the US ASD prevalence to be 1 in 150. Notably, retrospective review of cases in the ADDMN & CDC sample suggested that ASD prevalence estimates may be somewhat lower when DSM-5 criteria were applied (1 in 100 in 2008, 1 in 135 in 2006; Maenner et al., 2014). This would suggest a slightly greater increase between surveys (26%). Changes in conceptualization and diagnostic criteria for ASDs, as well as increased awareness and availability of services are likely to play a role in the rising rates, though the possibility that a true increase in ASD *incidence* exists cannot yet be ruled out (Fombonne, Quirke, & Hagen, 2011).

Irrespective of the reasons for these apparent increases, these studies suggest that a substantial number of people in the US and around the world are diagnosed with an ASD and need services. Ganz (2007) estimated that the societal per capita lifetime incremental cost of autism (i.e., cost due exclusively to having autism) was approximately $3.2 million. This estimate included both direct medical and non-medical

costs (e.g., intervention or care for the child or adult with ASD), as well as indirect costs, such as productivity losses for individuals with ASD and their parents. A substantial proportion of the incremental societal cost was spent after the age of 21, with adult care and loss of productivity accounting for approximately 21% and 30.7% of incremental societal cost, respectively (Ganz, 2007).

Considering concerns regarding the rise in autism prevalence and the substantial costs associated with this disorder, it is not surprising that there has been an increase in autism research and funding over the last several decades (Amaral, Dawson, & Geschwind, 2011). Of 741 projects funded between 1997 and 2006, Singh and colleagues (2009) report that 65% were concentrated in basic science, the majority of which focused on neurobiology or genetics. These studies have implicated numerous genetic and chromosomal variants as playing a role in the risk for ASDs or related features (see State & Levitt, 2011). Abnormalities in both structural and functional neurobiological measures have also been identified in groups of individuals with ASD (see Stigler, McDonald, Anand, Saykin, & McDougle, 2011). Nonetheless, a biological marker that can be used to reliably diagnose ASD on an individual basis has not yet been identified.

Until reliable biomarkers can be identified, diagnosis of ASD will continue to be based on behavioral assessment. As we take on a more dimensional diagnostic system, continued research is needed to quantify the range of behaviors and deficits that characterize ASD. This will allow us to more precisely study how ASD symptoms interact with developmental level and other dimensions of behavior that commonly co-occur with ASD, such as hyperactivity and aggression to produce functional impairments (Lord & Jones, 2012). To this end, an in-depth understanding of the strengths and

limitations of the instruments used to describe ASD symptoms in diagnostic assessments across the lifespan is needed.

In the past two decades, a number of standardized screening and diagnostic tools have been published. These span a variety of methods, including *parent interview* (e.g., Autism Diagnostic Interview – Revised [ADI-R]; Rutter, Le Couteur, & Lord, 2003; Diagnostic Interview for Social and Communication Disorders [DISCO]; Wing, Leekam, Libby, Gould & Larcombe, 2002), *direct observation* (e.g., Autism Diagnostic Observation Schedule [ADOS]; Lord et al., 1999; Screening Tool for Autism Spectrum Disorders in Toddlers [STAT]; Stone, Coonrod, & Ousley, 2000) and *caregiver rating forms* (e.g., Social Responsiveness Scale [SRS]; Constantino & Gruber, 2005; Social Communication Questionnaire [SCQ]; Rutter, Bailey, Lord & Berument, 2003; Autism Spectrum Screening Questionnaire [ASSQ]; Ehlers, Gillberg & Wing, 1999). Together, these measures have greatly contributed to making autism one of the most reliably diagnosed child psychiatric disorders (Volkmar & Lord, 2007). However, previous research demonstrates that child characteristics, such as age, language level, cognitive ability and behavior problems, are strongly associated with raw scores from many of these ASD screening and diagnostic measures (e.g., Constantino, Hudziak, & Todd, 2003; Corsello et al., 2007; Gotham et al., 2007; Hus & Lord, 2013; Mayes & Calhoun, 2011). These findings suggest that elevated scores from these measures may reflect a variety of behaviors that are not ASD-specific and that children with severe ASD-related impairments may not be quantitatively distinct from children with co-morbid behavioral conditions. This has significant implications for the diagnostic accuracy of these measures, as well as the interpretation of scores as indicators of ASD severity. Strong

associations with developmental level make it difficult to use scores clinically to measure treatment-related gains; for example, if scores are positively correlated with chronological age, decreases reflecting symptom improvement may be masked by the tendency for scores to increase as the individual gets older. In research, such non-specific influences on ASD measures make it difficult to differentiate between true differences in ASD symptom severity and differences due to recruitment biases in age or cognitive level. This might be expected in genetic studies, where samples may be comprised of children spanning a wide range of ages and skill levels that were recruited at different sites and in different ways. Non-specific effects on scores may also hinder comparison over time in longitudinal studies assessing trajectories of development and predictors of outcome.

These associations also have considerable implications for basic research investigating links between behavioral measures and genetic risk factors or other biomarkers for ASD. Logically, we might question a measure that yields the same score for two very different behavioral phenotypes (e.g., a mild-tempered, nonverbal child with profound intellectual disability and a verbally fluent child with an above average IQ and significant behavior problems). Thus, we have to be cognizant of the possibility that measures strongly associated with developmental level or other non-ASD-specific dimensions of behavior may erroneously place different groups close together on a continuum thought to indicate ASD severity. If nonverbal children with profound intellectual disability are indistinguishable from verbal children with severe externalizing behaviors, the likelihood of identifying quantitative trait loci associated with ASD is reduced. Researchers may misinterpret associations between behavioral scores and

biological mechanisms as ASD-specific when they are actually markers for broader developmental delay or other behavioral dimensions.

As we take on a dimensional diagnostic nosology, there is a need for diagnostic tools that provide more valid measurements of ASD-specific symptom severity that are less influenced by developmental characteristics and non-ASD-specific dimensions of behavior, such as age or externalizing problems. Each of the three studies in this dissertation seeks to expand the use of pre-existing ASD diagnostic measures with individuals across a range of ages to more appropriately meet clinical and research needs. The first study in this three-paper dissertation aims to increase understanding of non-ASD-specific factors on scores from the Social Responsiveness Scale (SRS), a parent questionnaire frequently used in genetic studies as a screening or diagnostic instrument, as well as a continuous measure of ASD symptom severity. Although previous reports of the SRS have indicated strong associations with behavior problems (Bölte, Poustka, & Constantino, 2008; Charman et al., 2007; Constantino et al., 2003; Constantino, Przybeck, Friesen, & Todd, 2000; Kanne, Abbacchi, & Constantino, 2009), researchers using SRS scores as a severity measure have almost always failed to control for these non-ASD-specific factors in their analyses (e.g., Coon et al., 2010; Duvall et al., 2007). Associations between SRS scores and developmental characteristics, such as age or cognitive level, have been described as small or nonsignificant (e.g., Bölte et al., 2008; Charman et al., 2007; Constantino & Gruber, 2005). However, these studies have often focused on correlations in normative samples or small clinical samples of children with average intelligence; the influence of developmental level on SRS scores has not been systematically investigated in children with ASD representing the full range of

intellectual functioning (i.e., from profound delays to superior intelligence). For the first study, analyses examining the relationship between SRS scores and demographics, behavior problems, and developmental level were undertaken in a large dataset of 2,368 children with ASD and 1,913 unaffected siblings, all 4 to 18 years of age, who were recruited for participation in a genetic study of families with one child with ASD who has no first-, second-, or third-degree relatives with ASD (the Simons Simplex Collection; Fischbach & Lord, 2010). Chapter II of this dissertation describes the methods of this study in more detail. Results of this study draw attention to the need to carefully consider the influence of child characteristics on interpretation of SRS scores and the importance of statistically controlling for non-ASD-specific factors when using scores as indicators of ASD severity.

The objective of the second study of this dissertation is to expand the utility of an existing measure of ASD severity – the ADOS Calibrated Severity Score (ADOS-CSS; Gotham, Pickles, & Lord, 2009). The ADOS is a semi-structured observational assessment which provides different modules and algorithms for assessment and diagnosis of individuals of different ages and developmental levels. The ADOS-CSS is a 10-point severity metric derived from raw ADOS totals from Modules 1-3. Compared to raw totals, the ADOS-CSS are much less influenced by child characteristics, therefore enabling comparison of scores across modules and time to identify trajectories of ASD severity and facilitating use of scores as quantitative phenotypes in genetic and neurobiological studies (Gotham et al., 2009). However, the ADOS-CSS is based on the total ADOS score, which is comprised of scores from two separate domains: Social Affect (SA) and Restricted and Repetitive Behaviors (RRB). As such, the ADOS-CSS

does not distinguish a child with significant social-communication impairments and few repetitive behaviors from a child who exhibits relatively mild social-communication impairments and severely restricted interests. Using a dataset comprised of 1,807 assessments from 1,118 individuals with ASD ages 2 to 16 years (i.e., the same dataset used to derive the ADOS-CSS), the second study of this dissertation separately calibrated SA and RRB domain totals to provide distinct measures of severity for each of these behavioral domains. As intended, distributions of SA-CSS and RRB-CSS were more uniform across age and language groups and both calibrated metrics were less strongly influenced by participant developmental level than the raw domain totals. Chapter III of this dissertation provides a complete description of the methods and results of this study. If replicated, separately calibrated domain scores should provide a clearer picture of ASD severity by allowing separate examination of social-communication impairments and repetitive behaviors, which may have distinct developmental trajectories or respond differently to intervention and are likely to be differentially associated with biological mechanisms.

While the first two studies of this three-part dissertation focus on measures intended for use with preschool and school-aged children, the final study focuses on improvement of a diagnostic measure used with older adolescents and young adults. The US Department of Education Office of Special Education Programs (2011) reported that, in the fall of 2010, nearly 370,000 children ages 6 through 21 were being served under the special education classification of "Autism." Considering that not all individuals with ASD receive special education services, this likely provides a conservative estimate of the number of individuals with ASD who will exit high school and transition to "adult

life" over the next 10-15 years. In light of the considerable lifetime incremental costs of autism, a substantial proportion of which are incurred during adulthood due to loss of productivity and adult care (Ganz, 2007), there is a critical need for research focused on adolescents and adults with ASD (Interagency Autism Coordinating Committee, 2011). In particular, a better understanding of the strengths and difficulties of adults with ASD is needed to inform development of supports and services.

Recent studies have begun to explore predictors of outcome and the current needs of adolescents and adults with ASD. However, because there are currently no well-established standardized observational measures of ASD severity for adults, these studies have been limited to exploration of factors such as adaptive and cognitive functioning and change in diagnostic classification. Although some have investigated change in ASD symptoms using raw totals from measures such as the ADOS and ADI-R; as noted above, longitudinal comparisons of these measures are confounded by strong associations with developmental level (Gotham et al., 2009; Hus & Lord, 2013). Revisions to diagnostic algorithms and derivation of ADOS-CSS for ADOS modules 1-3, used with younger children and adolescents, are now available to facilitate comparisons. However, comparable changes have not yet been made to the ADOS Module 4, used with verbally fluent adolescents and young adults. This hinders the examination of developmental trajectories from early childhood through adulthood. Thus, the focus of the third dissertation study was to revise the ADOS Module 4 to increase comparability to currently used algorithms for children and younger adolescents. Chapter IV of this dissertation describes the analysis and results of Study 3 in greater detail. The revised Module 4 algorithm resulting from this study provides improved sensitivity while

maintaining or improving specificity. In addition, Module 4 calibrated severity scores derived in this study yield quantitative estimates of social-communication and restricted, repetitive behaviors that are relatively independent of participant characteristics, such as age and intellectual ability. These changes improve Module 4's comparability to previously revised ADOS modules used with younger and more language impaired children, which will hopefully facilitate efforts to increase understanding of the unique strengths and challenges of adults with ASD.

As a whole, this dissertation aims to enhance the validity of scores from currently widely-used ASD screening and diagnostic measures as indicators of ASD severity. More thorough knowledge of how non-ASD-specific child characteristics influence scores on the SRS will guide interpretation of studies using this measure as a quantitative trait in investigations of ASD biomarkers. Newly standardized ADOS domain scores providing separate metrics of severity for core ASD symptoms can be used by clinicians to inform treatment recommendations and monitor improvement over time, as well as researchers examining trajectories of ASD symptoms and mapping ASD symptoms to underlying biological mechanisms. Finally, revision of the Module 4 algorithm and derivation of calibrated severity scores will be useful to describe the ASD phenotype in adults and examine predictors of longer-term outcomes.

References

Amaral, D., Dawson, G., & Geschwind, D. H. (2011). Preface. In D. Amaral, G. Dawson, & D. H. Geschwind (Eds.), *Autism Spectrum Disorders* (1st ed., pp. 90–111). USA: Oxford University Press.

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

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.

Autism and Developmental Disabilities Monitoring Network & Center for Disease Control and Prevention. (2012). Prevalence of autism spectrum disorders — Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2008. *Morbidity and Mortality Weekly Report Surveillance Summaries*, *61*(3), 1–19.

Bölte, S., Poustka, F., & Constantino, J. N. (2008). Assessing autistic traits: Cross-cultural validation of the Social Responsiveness Scale (SRS). *Autism Research*, *1*(6), 354–363.

Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., & Pickles, A. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *The British Journal of Psychiatry*, *191*(6), 554–559.

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

Constantino, J. N., Hudziak, J. J., & Todd, R. D. (2003). Deficits in reciprocal social behavior in male twins: evidence for a genetically independent domain of psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*(4), 458–467.

Constantino, J. N., Przybeck, T., Friesen, D., & Todd, R. D. (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *Journal of Developmental and Behavioral Pediatrics*, *21*(1), 2–11.

Coon, H., Villalobos, M. E., Robison, R. J., Camp, N. J., Cannon, D. S., Allen-Brady, K., Miller, J. S., et al. (2010). Genome-wide linkage using the Social Responsiveness Scale in Utah autism pedigrees. *Molecular Autism*, *1*(1), 8.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Leventhal, B. L., & Lord, C. (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.

Duvall, J. A., Lu, A., Cantor, R. M., Todd, R. D., Constantino, J. N., & Geschwind, D. H. (2007). A quantitative trait locus analysis of social responsiveness in multiplex autism families. *American Journal of Psychiatry*, *164*(4), 656–662.

Ehlers, S., Gillberg, C., & Wing, L. (1999). A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *Journal of Autism and Developmental Disorders*, *29*(2), 129–141.

Fischbach, G. D., & Lord, C. (2010). The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron*, *68*(2), 192–195.

Fombonne, E. (2005). The changing epidemiology of autism. *Journal of Applied Research in Intellectual Disabilities*, *18*(4), 281–294.

Fombonne, E., Quirke, S., & Hagen, A. (2011). Epidemiology of pervasive developmental disorders. In D. Amaral & G. Dawson (Eds.), *Autism Spectrum Disorders* (1st ed., pp. 90–111). USA: Oxford University Press.

Ganz, M. L. (2007). The lifetime distribution of the incremental societal costs of autism. *Archives of Pediatrics & Adolescent Medicine*, *161*(4), 343–349.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(5), 693–705.

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

Hus, V., & Lord, C. (2013). Effects of child characteristics on the Autism Diagnostic Interview-Revised: Implications for use of scores as a measure of ASD severity. *Journal of Autism and Developmental Disorders*, 43(2), 371–381.

Interagency Autism Coordinating Committee. (2011, January 18). The 2011 Interagency Autism Coordinating Committee Strategic Plan for Autism Spectrum Disorder Research. Retrieved from http://iacc.hhs.gov/strategic-plan/2011/index.shtml

Kanne, S. M., Abbacchi, A. M., & Constantino, J. N. (2009). Multi-informant ratings of psychiatric symptom severity in children with autism spectrum disorders: The importance of environmental context. *Journal of Autism and Developmental Disorders*, *39*(6), 856–864.

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

Lord, C., & Jones, R. M. (2012). Annual Research Review: Re-thinking the classification of autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *53*(5), 490–509.

Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D. M., Ousley, O., et al. (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*, *69*(3), 306–313.

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

Maenner MJ, Rice CE, Arneson CL, & et al. (2014). Potential impact of DSM-5 criteria on autism spectrum disorder prevalence estimates. *Journal of the American Medical Association - Psychiatry*. Advanced Online Publication. doi:10.1001/jamapsychiatry.2013.3893

Mayes, S. D., & Calhoun, S. L. (2011). Impact of IQ, age, SES, gender, and race on autistic symptoms. *Research in Autism Spectrum Disorders*, *5*(2), 749–757.

Robertson, J. M., Tanguay, P. E., L'Ecuyer, S., Sims, A., & Waltrip, C. (1999). Domains of social communication handicap in autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *38*(6), 738–745.

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

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

Singh, J., Illes, J., Lazzeroni, L., & Hallmayer, J. (2009). Trends in US autism research funding. *Journal of Autism and Developmental Disorders*, *39*(5), 788–795.

State, M. W., & Levitt, P. (2011). The conundrums of understanding genetic risks for autism spectrum disorders. *Nature Neuroscience*, *14*(12), 1499–1506.

Stigler, K. A., McDonald, B. C., Anand, A., Saykin, A. J., & McDougle, C. J. (2011). Structural and functional magnetic resonance imaging of autism spectrum disorders. *Brain Research*, *1380*(0), 146–161.

Stone, W. L., Coonrod, E. E., & Ousley, O. Y. (2000). Brief report: Screening Tool for Autism in Two-year-olds (STAT): Development and preliminary data. *Journal of autism and developmental disorders*, *30*(6), 607–612.

U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS. (2011, July 15). Table 1-3. Number of students ages 6 through 21 served under IDEA, Part B, by disability category and state: Fall 2010. Retrieved from https://www.ideadata.org/TABLES34TH/AR_1-1.pdf

Volkmar, F. R., & Lord, C. (2007). Diagnosis and definition of autism and other pervasive developmental disorders. In F. R. Volkmar (Ed.), *Autism and pervasive developmental disorders* (2nd ed., pp. 1–31). New York: Cambridge University Press.

Wing, L., Leekam, S. R., Libby, S. J., Gould, J., & Larcombe, M. (2002). The Diagnostic Interview for Social and Communication Disorders: Background, inter-rater reliability and clinical use. *Journal of Child Psychology and Psychiatry*, *43*(3), 307–325.

CHAPTER II

Factors influencing scores on the Social Responsiveness Scale

Autism Spectrum Disorders (ASD) are characterized by a range of symptoms which are heterogeneous in nature and severity. A single measure that captured ASD-symptom severity would be useful as a quantitative phenotype in genetic and neurobiological studies. However, in addition to heterogeneity across individuals, measurement of ASD severity is complicated by common co-occurrence of non-ASD-specific conditions (e.g., intellectual disability) and behaviors, such as difficulties with attention or hyperactivity, as well as age-related variation in symptom presentation. Research has demonstrated that raw totals from many ASD diagnostic and screening measures are influenced by non-ASD-specific child characteristics, such as age and language level (e.g., Corsello et al., 2007; Gotham, Pickles, & Lord, 2009; Mayes & Calhoun, 2010). For example, age and language explained 22% of variance in scores from the Autism Diagnostic Interview-Revised (ADI-R; Rutter et al., 2003), a diagnostic parent-interview often used as a measure of ASD severity (Hus & Lord, 2013). Recent discussions regarding developmental screening and assessments have called for better understanding of factors influencing parent report (Aylward, 2009; Warren et al., 2012). Often, parent-factors such as education level and frame of reference are acknowledged as limitations and weighed against the relative benefits of efficiency and cost-effectiveness of questionnaires compared to interview or observational measures requiring more time and highly-trained clinicians. For example, the Social

Responsiveness Scale (SRS; Constantino & Todd, 2005), a parent-completed questionnaire which was originally proposed as a continuously distributed, quantitative measure of autism-related severity in the general population (Constantino et al., 2000; 2003), is commonly used as an estimate of ASD severity in genetic and neurobiological studies.

Although the SRS is frequently referred to as a measure of "social impairment," many SRS items describe other core features of ASD, including communication deficits and repetitive behaviors (Constantino et al., 2000), as well as symptoms not exclusively related to ASD diagnostic criteria (Grzadzinski et al., 2011). Informants complete all 65 SRS items, irrespective of the child's age or language level. Without explicit instructions, it is unclear how parents rate items that are not applicable to their child (e.g., items assessing conversation for a nonverbal child). Considering that scores from the ADI-R are affected by child characteristics, despite having subsets of items for children of different ages and language abilities and being administered and scored by a trained clinician, it seems likely that scores on the parent-rated SRS would be similarly influenced. However, in spite of their implications for interpretability of scores, particularly when being used as indicators of ASD-specific severity, studies have not systematically examined how non-ASD-specific child characteristics affect the use of SRS scores as a quantitative measure. The goal of this study is to provide a better understanding of how factors that affect other ASD-symptom measures influence interpretation of SRS scores.

Underscoring the concern regarding effects of non-ASD-specific factors on ASD-symptom measures, several studies have shown strong associations between the SRS and

measures of behavior problems in clinical samples of children with ASD and other psychiatric diagnoses (Bölte, Poustka, & Constantino, 2008, Charman et al., 2007; Constantino et al., 2000; Kanne, Abbacchi & Constantino, 2009). The strength of these associations was similar for children with ASD and children with other non-ASD diagnoses (Constantino et al., 2000). In an epidemiological sample of twins, Constantino, Hudziak, and Todd (2003) reported that scores from the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001), a parent-report measure of psychiatric symptoms, explained 43-52% of the variance in SRS scores, though they emphasized that an additional 44% of variance was independent of behaviors captured on the CBCL. Similarly, Charman and colleagues (2007) reported decreased specificity of the SRS, and two other ASD screening instruments, for children with elevated behavior problems. Although many children with ASD may have additional behavior problems (Kanne et al., 2009), it is possible that associations between the SRS and measures of behavior problems reflect non-specific difficulties rather than (or in addition to) ASD-related variation in behavior. If this were true, children with severe ASD-related impairments may not be quantitatively distinct from children with co-morbid behavioral conditions, and labeling the SRS as a measure of autism severity could be misleading. Instead, SRS scores may be more appropriately interpreted as reflecting a broad range of impairments beyond ASD. This is of particular concern, considering that the SRS is widely used to describe the severity of ASD symptoms and/or of ASD-related social impairment in both clinical and research settings (e.g., Constantino et al., 2006, Duvall et al., 2007; Kanne et al., 2009).

Fewer studies have examined the relationship between SRS scores and factors that influence other measures of ASD-symptoms, such as age, language and cognitive level. Although there is some evidence that SRS scores may be influenced by these child characteristics, this is not widely acknowledged, possibly because the focus of these studies has not been to systematically examine the effects of child characteristics on SRS scores. For example, in a small clinical sample, when children were grouped by language level, nonverbal children with autism had higher scores and their distribution was clearly differentiated from that of verbal children with autism (Constantino et al., 2000). In a larger study of families of children with ASD (Constantino et al., 2010), there was a modest effect of nonverbal status (or parent-reported intellectual disability) on gender-normed SRS-T for children with ASD.

With regard to age, two studies reported that SRS scores were not significantly correlated with age in normative or clinical samples (Bölte et al., 2008; Constantino & Gruber, 2005). However, factor loadings for SRS items differed when subsets of 4-7-year-old and 8-14-year-old school children were analyzed separately (Constantino et al., 2000). Three studies including children with ASD and non-ASD diagnoses indicated nonsignificant associations with IQ (Charman et al., 2007; Constantino, et al., 2003; Constantino et al., 2006), but three additional studies reported negative correlations between SRS and FSIQ or NVIQ (Boltë, et al., 2008; Constantino et al., 2000; Constantino et al., 2007). Moreover, two of these studies (Boltë, et al., 2008, Constantino et al., 2000) reported that correlations were stronger for children with ASD (r=-.18 to -.42) than non-ASD clinical controls (r=-.04 to -.08). These inconsistencies are difficult to

interpret, perhaps because of small sample sizes (ranging from 37 to 127 in all but Boltë, et al., 2008) that have primarily included children with average intelligence.

In sum, studies consistently suggest a relationship between behavior problems and SRS scores; however, in spite of their implications for interpretability of scores, particularly when being used as indicators of ASD-specific severity, this is rarely acknowledged by researchers using the SRS as a quantitative measure. Moreover, the effects of age, language level and IQ have been documented for other measures, but thorough understanding of the influence of these child characteristics on SRS scores has been obscured by small sample sizes and a lack of systematic analyses with ASD samples. Such understanding has critical implications for interpretation of SRS scores as a quantitative measure of ASD-symptoms. The present study seeks to address such limitations by investigating these relationships in a large sample of probands with ASD and their unaffected siblings. Based on previous studies, it is hypothesized that more behavior problems and greater expressive language impairment will be associated with higher SRS scores for probands and siblings, and that NVIQ will be negatively associated with SRS scores in probands (IQs are not available for siblings). Consistent with other parent-rating measures (e.g., Corsello et al., 2007), it is predicted that SRS scores will be higher with increasing age.

Method

Participants

Participants were 2,368 probands and 1,913 unaffected siblings evaluated at 12 university-based centers from 2007-2011 as part of the Simons Simplex Collection (SSC), a genetic study of families with one child with ASD who does not have first-,

second- or third-degree relatives with ASD. All probands met Collaborative Programs of Excellence in Autism (CPEA) criteria for a diagnosis of Autism, ASD, or Asperger Disorder (see Appendix A.1). All siblings screened negative for ASD or indication of the broader phenotype. Detailed study procedures are included in Appendix A.2. Families were predominantly White (78%) and well-educated (61% maternal education of Bachelor's degree or higher). Sample demographics are provided in Table 2.1. Parents gave informed consent, approved by Institutional Review Boards at each university.

Measures

Autism Symptoms.

The SRS (Constantino & Gruber, 2005) is a parent-completed questionnaire; items describe a child's behavior in the past 6 months, yielding a raw total (SRS-Raw) and gender-normed T-score (SRS-T; intended to correct gender differences observed in normative samples). Though originally proposed as a continuously distributed, quantitative measure of ASD-severity, recent work has found bimodal distributions within affected and unaffected family members of children with ASD (Constantino et al., 2010; Virkud et al., 2009). The manual recommends use of SRS-Raw in research for comparability to early studies of the SRS, though several recent studies use SRS-T (e.g., Constantino et al., 2010).

The Autism Diagnostic Observation Schedule (ADOS; Lord et al., 1999) Calibrated Severity Score (CSS) was chosen as an ASD-severity measure that is less influenced by child characteristics than raw totals (Gotham et al., 2009). This 10-point metric (higher scores reflecting greater ASD-severity) was derived from raw totals based on participants' ages and language levels. The ADI-R (Rutter et al., 2003) Current

Behavior Algorithm total (ADI-Current; see Hus & Lord, 2013) was used as a parent-report measure of current ASD-symptoms.

Social Development.

The Vineland Adaptive Behavior Scales, Second Edition (Vineland-II; Sparrow, et al., 2005) is a parent interview. Standard scores from the Socialization domain (VSOC) were used as a measure of social development available for probands and siblings to allow comparison between groups.

Behavior Problems.

Two forms of the CBCL (for children ages 18 months to 5 years and 6 to 18 years) each yield T-scores for Internalizing (CBCL-I) and Externalizing (CBCL-E) domains and five overlapping Syndrome Scales (Anxious-Depressed, Withdrawn-Depressed, Somatic Complaints, Attention Problems, Aggressive Behavior). CBCL-I and CBCL-E were used as estimates of behavior problems; Syndrome Scales were used for post-hoc analyses.

Developmental Level.

Proband and sibling chronological ages in years were used as a continuous predictor for regression analyses. The Vineland-II Expressive Communication standard score (VEC) was chosen to provide a continuous indicator of expressive language abilities available for probands and siblings. For probands, ADOS Module was used as a categorical indicator of expressive language; Module-1 (single words or nonverbal) =18.4%, Module-2 (simple phrases)=22.8%, Module-3 (complex sentences)=58.8%. NVIQ was used to indicate proband cognitive level. VIQ was not included due to

multicollinearity with NVIQ and because expressive language level was included separately.

Only demographics, SRS, Vineland-II and CBCL were available for siblings.

Data Analysis

Preliminary gender comparisons of SRS-Raw and SRS-T-Scores were conducted using SPSS 17.0 T-TEST. Pearson correlations were run between measures of ASD-symptoms and social development.

Linear regression models were analyzed separately for probands and siblings using SPSS REGRESSION. In Model-A, SRS-Raw was the dependent variable and predictors were entered in the following blocks to allow examination of the relative contribution of each set of variables: Demographics (gender=female vs. male; race=white vs. non-white; maternal education=graduate/bachelor degree vs. some college or less), Social Development (VSOC), Behavior Problems (CBCL-I, CBCL-E), and Developmental Level (age, VEC). All variables were centered at the mean. To examine effects of language level and age in the ASD sample, Model-B replaced the age-standardized VSOC and VEC with ADOS-CSS and ADOS-Module. To explore the relationship between parent-report measures of ASD-symptoms, ADI-Current was added but entered last (due to its strong associations with age and language; Hus & Lord, 2013). Thus, Model B, predicting proband SRS-Raw, included: Demographics, ASD-symptoms (ADOS-CSS), Behavior Problems, Developmental Level (age, Module-1 vs. Module-3, Module-2 vs. Module-3, NVIQ), ADI-Current.

Post-hoc Analyses.

Post-hoc analyses were conducted to better understand associations between SRS-

Raw and CBCL. First, to examine whether behavior problems influence social development, a regression predicting VSOC (Model-C), was fit with the following variables: Demographics, Behavior Problems, Developmental Level (Age, VEC) and ASD-Symptoms (SRS-Raw). Next, to explore how the profile of differences between CBCL Syndrome Scales related to differences in SRS-Raw, differences between probands and siblings from the same family were computed for VSOC, CBCL Syndrome Scales, age and VEC and used to predict proband-sibling differences in SRS-Raw (Model-D). Finally, to investigate whether externalizing behaviors significantly predicted SRS-Raw, a regression predicting SRS-Raw was run excluding CBCL-I from the predictors (Model-E); Model-E was otherwise identical to Model-B. More detailed explanation and justification for these analyses are provided in Appendix A.3.

For all regression models, Cohen's f^2 was computed to assess the effect of each block of predictors while controlling for all other variables; f^2 of .02, .15, and .35 reflect small, medium, and large effect sizes, respectively (Cohen, 1988). Regressions predicting SRS-Raw and SRS-T were nearly identical, therefore only analyses for SRS-Raw are reported below.

To visually demonstrate the effects of behavior problems on SRS-Raw, children were divided at VSOC=70 (two standard deviations below the standard mean of 100) into "low" or "high" groups; within each group, children were further divided into "low" or "high" groups at the CBCL-E clinical-concern cut-off of 64. SPSS ONEWAY and post-hoc Tukey tests were used to compare children across the four VSOC/CBCL-E groups. To investigate the effects of controlling for CBCL scores, residuals from a model

including SRS-Raw as the dependent variable and CBCL-I and CBCL-E as the predictors were compared across the four groups.

The powmr.exe program (Dunlap, Xin and Myers, 2004) was used to compute power. Power estimates were adequate (at or above .9) for all models fit. For example, for Model-A, fit for siblings (8 predictors, N=1894, small effect estimated at R=.14, which is equivalent to an f^2 of .02), the power estimate is .998. For Model-B, fit for probands (13 predictors, N=2322), the power to detect a small effect was estimated to be .999.

Given the large sample and multiple comparisons, significance level was set at $p<.001$ for all analyses.

Results

Preliminary Analyses

As shown in Table 2.1, male siblings had higher SRS-Raw than female siblings; $t(1747.36)=5.03$, $p<.001$, but sibling SRS-T did not differ by gender. Male siblings also had somewhat lower VEC than females; $t(1905)=-3.47$, $p<.001$. In contrast to the sibling results, male and female probands did not differ on SRS-Raw, but male probands had lower SRS-T than females; $t(382.75)=-10.11$, $p<.001$. Male probands also had higher NVIQ; $t(2366)=4.94$, $p<.001$ than females. Table 2.2 shows correlations between SRS-Raw and measures of ASD symptoms and social development.

Predictors of SRS-Raw

As shown in Table 2.3, Model-A for probands and siblings explained 46% and 33% of variance in SRS-Raw, respectively. For both, more behavior problems and social impairment (i.e., higher CBCL-I and CBCL-E, lower VSOC) predicted higher SRS-Raw.

Additionally, higher SRS-Raw were associated with greater language impairment (i.e., lower VEC) for both groups, as was being male and younger for siblings only; these effects were small, but significant.

In Model B, ADOS-CSS was significant, but explained only 1% of variance in proband SRS-Raw (Block 2; Table 2.4). Behavior problems (Block 3) and developmental level (Block 4) had medium to large effects on SRS-Raw. In the final model including all predictors, more behavior problems, higher age and lower NVIQ were associated with higher SRS-Raw; ADI-Current explained an additional 9% of variance in SRS-Raw after controlling for previous factors (Block 5).

Post-hoc Analyses

Given that associations between SRS-Raw and CBCL-I and CBCL-E were equally large or larger than relationships with social development (VSOC) and ASD-symptoms (ADOS-CSS, ADI-Current), it was of interest to more closely examine the relationship between SRS-Raw and behavior problems. A summary of post-hoc analyses is provided below (details are described in the online supplement).

First, one must consider the possibility that the association between SRS-Raw and behavior problems reflects true influences of behavior problems on social skills in probands and siblings. If true, CBCL-I and CBCL-E should be significant predictors of VSOC, a standardized measure of social development. As shown in Table 2.5 (Model-C), CBCL scores explained only 2-3% of variance in VSOC.

Associations between SRS-Raw and CBCL scores could also be explained by ASD-specific variation in CBCL scales containing items that appear to describe core ASD-symptoms (e.g., CBCL-Withdrawn/Depressed). If true, proband-sibling differences

in scores on these scales should be related to proband-sibling differences in SRS-Raw, whereas differences in other CBCL scales (e.g., CBCL-Attention) should not. In Model-D (Table 2.6), the best predictors of proband-sibling differences in SRS-Raw were differences in CBCL-Attention and CBCL-Withdrawn/Depressed. Associations between SRS-Raw and CBCL-Attention, CBCL-Withdrawn/Depressed and VSOC were of similar magnitude.

Next, the significance of externalizing behaviors as a predictor of SRS-Raw was tested in the absence of CBCL-I. As shown in Table 2.7 (Model-E), the relationship between CBCL-E and SRS-Raw was significant and as strong as the relationship between ADI-Current and SRS-Raw, after controlling for all other factors.

Finally, as shown in Figure 2.1A, comparisons of children divided into groups according to low/high VSOC and low/high CBCL-E indicated significant differences in SRS-Raw, $F(3,2360)=257.29$ $p<.001$. Tukey tests revealed that, within VSOC groups, the high-CBCL-E group had higher SRS-Raw than the low-CBCL-E group (Mdiff=13.75 and Mdiff=20.31, $p<.001$, respectively). Additionally, the high-VSOC/high-CBCL-E group did not differ significantly from the low-VSOC/low-CBCL-E group (Mdiff=2.16, $p=.54$), indicating that children whose parents reported relatively good social skills and high levels of externalizing behaviors had comparable SRS-Raw to children whose parents reported relatively poor social skills and low levels of externalizing behaviors. As shown in Figure 2.1B, when CBCL scores were controlled, SRS-Residual scores differed significantly across the four groups; $F(3,2360)=165.49$, $p<.001$; not surprisingly, the effects of behavior problems were diminished. Within the low-VSOC group, the low-CBCL-E group now had somewhat higher SRS-Residuals than the high-CBCL-E group

(Mdiff=4.10, p=.02). Within the high-VSOC group, low- vs. high-CBCL-E groups did not differ (Mdiff=-.29, p=.99). Additionally, the low-VSOC/low-CBCL-E group had significantly higher SRS-Residual scores than the high-VSOC/high-CBCL-E group (Mdiff=20.03, p<.001).

Discussion

In the present study, for both probands and siblings, parent-reported behavior problems (CBCL) were strongly predictive of higher SRS-Raw, explaining similar, and often higher, proportions of variance in SRS-Raw than measures of social development (Vineland-II). For probands, SRS-Raw were also higher for older children and children with less language and lower NVIQ. When children were divided into four groups based on low or high levels of parent-reported social impairment and behavior problems, children with more externalizing behaviors had higher SRS-Raw than children with low externalizing behaviors, in spite of similar parent-reported social skills. Perhaps most significant was the finding that children with more impaired social skills and fewer externalizing behaviors had comparable SRS-Raw to children with relatively better social skills and more externalizing behaviors. In other words, the SRS-Raw of children with good social skills but high levels of behavior problems were indistinguishable from SRS-Raw of children with poor social skills and fewer behavior problems. It was possible to minimize these effects by using SRS-Residuals from the regression model controlling for CBCL scores. These findings demonstrate that SRS-Raw are strongly influenced by non-ASD-specific child characteristics, such as internalizing and externalizing behavior problems and developmental level, highlighting the need to exercise caution when using the SRS as a continuous measure of ASD-severity or social deficits.

Associations between behavior problems, age, language and SRS-Raw are not surprising and have been reported for several other diagnostic measures. These factors contribute to the phenotypic heterogeneity in ASD. The extent to which elevated scores on measures of behavior problems indicate distinct, co-morbid disorders or reflect secondary impairments related to ASD is unclear (Constantino, 2011; Georgiades et al., 2010). One possibility is that these associations could be limited to parent report questionnaires. The weaker relationship between SRS-Raw and ADOS-CSS compared to that observed between SRS-Raw and ADI-Current scores highlights that method variance (i.e., clinician observation vs. parent report) may be an important factor in the measurement of ASD symptoms. However, the differential relationships between associations between SRS and VSOC for siblings and probands suggest that findings cannot be entirely attributed to parent-report bias.

Another possibility is that these associations reflect a high prevalence of behavior problems in children with ASD. Nonetheless, if the relationship between CBCL and SRS scores was explained by core ASD-features, we would not necessarily expect to find the same relationship between SRS and CBCL scores in siblings. In this study, the association between sibling SRS and CBCL was of similar magnitude to that observed for probands. Additionally, while CBCL scores explained 20-26% of variance in SRS-Raw, they explained only 2-3% of variance in sibling and proband Vineland-II Social scores in this study and 2-4% of variance in proband ADOS-CSS and ADI-Current scores in a related study (Hus & Lord, 2013).

When examining the association between SRS and CBCL scores more closely, the strongest predictors of proband-sibling differences in SRS-Raw were differences in

CBCL-Withdrawn/Depressed and CBCL-Attention scores. While the Withdrawn/Depressed scale may reflect some ASD-symptoms, the CBCL-Attention scale does not include items describing core-ASD-features. Grzadzinski and colleagues (2011) reported that children with ADHD who score highly on SRS items related to DSM-IV criteria for ASD also score highly on items not specifically related to ASD criteria. Moreover, when internalizing symptoms were excluded from the model, externalizing behaviors had a medium-sized effect on SRS-Raw. This further suggests that the association with behavior problems cannot be entirely explained by items which may be capturing ASD-symptoms. Although the externalizing domain is comprised of items measuring aggression, attention problems and rule-breaking behaviors which may frequently co-occur with ASD, they are not part of the core-ASD-symptoms as defined by diagnostic criteria.

Taken together, these results indicate that SRS scores are highly influenced by behavior problems. It is not clear whether this is because items intended to capture social impairments (e.g., poor eye contact, difficulty with peers) lack diagnostic specificity, or whether parents interpret questions as describing qualitatively different behaviors than the ASD-symptoms that items were intended to assess (Veenstra-VanderWeele & Warren, 2011). Thus, it may be appropriate to interpret SRS scores as reflecting parents' perception of their child's overall level of impairment (which may be influenced by developmental difficulties and behavior problems, as well as ASD-symptoms), rather than as a measure of severity of core-ASD-features. This is particularly important for researchers using SRS scores as a quantitative phenotype (e.g., Duvall et al., 2007) because biological mechanisms associated with these scores may actually be markers for

general impairment rather than social or ASD-specific impairments. Constantino and colleagues highlighted this idea, saying, "Endophenotypes can be misleading…if they do not represent truly independent subdomains of autism" (2004, p. 719). As shown in Figure 2.1B, one way to increase the probability that associations between SRS scores and biological mechanisms are due to ASD-related behaviors is to statistically control for non-ASD-specific influences (e.g., CBCL).

It is also important to note that while gender-normed T-scores are available to correct gender differences observed in normative samples (Constantino & Gruber, 2005), the effects of gender were minimal in our unaffected sibling sample and there were no gender differences in SRS-Raw for probands. While SRS-T "corrects" for sibling gender differences, the same adjustment results in female probands having higher (i.e., worse) SRS-T than male probands. The only other gender difference observed for probands was that females were more cognitively impaired than males. In a recent study, gender was reportedly a significant predictor of SRS-T for both probands and unaffected siblings (Constantino et al., 2010). However, the absence of gender differences on other proband measures of ASD-symptoms suggests that using gender-normed T-scores in clinical populations may exaggerate difficulties in females. Alternatively, failing to control for gender differences by using SRS-Raw could overestimate impairments in unaffected males and erroneously lead to conclusions that unaffected males are more impaired than unaffected females. Additional data are needed to better understand whether raw or T-scores are more appropriate and to inform development of standard expectations for reporting SRS scores in different sample types. Nevertheless, the effects of gender on SRS-Raw were small compared to associations with behavior problems, age and

language level. Thus, standardizing SRS-Raw to account for behavior problems and developmental level may be more crucial to interpretation of the SRS than gender-based T-scores.

Limitations

In Model A, the Vineland-II was used as a measure of social development available for both probands and siblings (because siblings were not administered the full battery of tests). The Vineland-II has demonstrated acceptable reliability and validity in normative samples, and score profiles in children with ASD reflect expected impairments in the Communication and Socialization domains (Sparrow et al., 2005; Kanne et al., 2011), suggesting that this was an appropriate measure of social impairment for both groups. However, the restricted range of VSOC scores in this sibling sample may have contributed to the relatively weak association between SRS-Raw and VSOC for siblings. It is possible that rater contrast effects (i.e., parents comparing the proband and unaffected sibling) affected post-hoc analyses of how differences in CBCL scores predicted differences in SRS-Raw. However, if this were the case, we would expect differences in SRS-Raw to predict differences on all scales or only scales assessing potentially ASD-specific symptoms (e.g., CBCL-Social Problems), which was not seen here.

Finally, many factors known to influence parental ratings of behavior were not measured, such as parent stress levels or previous knowledge of their child's diagnosis. Although demographics such as maternal education and race did not emerge as significant predictors, the limited variability in our sample with respect to these characteristics limits our ability to interpret such findings. Attention to how informant

characteristics influence scores on parent-report may provide important insight about ways to improve questionnaires for use as indicators of ASD severity. Notably, these issues are not limited to the SRS and should be considered for other parent-report measures being used as measures of ASD-severity. Moreover, how child and informant characteristics influence sensitivity and specificity of the SRS should be examined as they have been for other screening instruments (e.g., Corsello et al., 2007); this was not feasible in this study given the stringent inclusion/exclusion criteria used for the SSC.

Conclusions

SRS scores were strongly associated with behavior problems for children with ASD and their unaffected siblings. Effects of age and expressive language level were smaller, but significant. When used as a quantitative phenotype measure in samples of children with ASD, the SRS may exaggerate impairments in children who are older or have greater behavior problems or cognitive delays. These results caution against interpretation of SRS scores as a measure of social impairment or ASD-specific severity without careful consideration of the effects of behavior problems, age, language and cognitive level. Future studies utilizing clinical samples of children with non-ASD diagnoses are needed to explore how these factors influence the SRS' sensitivity and specificity, as well as to inform standardization of scores for use as continuous measures of ASD-severity.

Table 2.1 Sample Demographics

	Probands				**Siblings**			
	males (n=2056)		**females (n=312)**		**males (n=890)**		**females (n=1023)**	
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Age (years)	8.74	(3.32)	8.90	(3.60)	9.49	3.71623	9.46	(3.65)
SRS-Raw	97.56	(26.82)	99.32	(27.24)	**20.53**	**(15.44)**	**17.22**	**(13.02)**
SRS T-score	**80.56**	**(12.83)**	**89.63**	**(15.05)**	43.70	(7.39)	44.26	(7.19)
VSOC	71.54	(12.57)	70.15	(12.71)	*101.77*	*(11.93)*	*103.19*	*(11.29)*
CBCL-E	*56.39*	*(10.7)*	*57.79*	*(10.2)*	46.84	(9.81)	46.19	(9.42)
CBCL-I	60.35	(9.47)	59.96	(9.98)	*48.29*	*(10.19)*	*47.32*	*(9.96)*
VEC	*10.23*	*(3.06)*	*9.71*	*(3.03)*	**16.02**	**(2.37)**	**16.39**	**(2.34)**
ADI-Current	17.04	(7.30)	17.59	(7.80)				
ADOS-CSS	7.43	(1.68)	7.43	(1.73)				
NVIQ	**85.81**	**(25.70)**	**78.12**	**(25.18)**				

Bold=p<.001, *Italics*=p<.05 male vs. female*;* Ns vary due to missing data; VSOC=Vineland-II Social Standard Score; CBCL=Child Behavior Checklist; I=Internalizing; E=Externalizing; VEC=Vineland-II Expressive Communication V-Score; ADOS-CSS=ADOS Calibrated Severity Score; NVIQ=NonverbalIQ.

Table 2.2 Correlations between SRS-Raw and child measures

	Age	VSOC	CBCL-I	CBCL-E	VEC	ADI-C	ADOS-CSS	NVIQ
Probands	.14	-.50	.48	.42	-.38	.52	.10	-.27
Siblings	-.08	-.27	.47	.43	-.22			

Ns vary due to missing data; All correlations are significant ($p \leq .001$); VSOC=Vineland-II Social Domain Standard Score; CBCL=Child Behavior Checkist-Internalizing T-Score; CBCL-E Child Behavior Checklist Externalizing T-Score; VEC=Vineland-II Expressive V-Scale Score; ADI-Current=Autism Diagnostic Interview-Revised Current Behavior Algorithm Total; ADOS-CSS=ADOS Calibrated Severity Score; NVIQ=Nonverbal IQ

Table 2.3 Model-A: Predictors of SRS-Raw for probands and siblings

	Probands							
	B	SE *B*	95% CI		r_{part}	R^2	ΔR^2	f^2
			Lower	Upper				
Constant	**97.81**	**.41**	**97.02**	**98.61**				
Demographics						**.01**	**.01**	.01
	-.16	1.21	-2.53	2.21	.00			
Race	*-2.19*	*1.01*	*-4.16*	*-.22*	*-.03*			
MatEduc	-.97	.85	-2.63	.69	-.02			
Social Development						**.25**	**.24**	.33
VSOC	**-.74**	**.05**	**-.84**	**-.64**	**-.22**			
Behavior Problems						**.45**	**.20**	.36
CBCL-E	**.44**	**.05**	**.35**	**.53**	**.14**			
CBCL-I	**.99**	**.05**	**.89**	**1.09**	**.29**			
Developmental Level						**.46**	**.01**	.01
Age	-.03	.13	-.29	.23	.00			
VEC	**-1.19**	**.20**	**-1.59**	**-.79**	**-.09**			
	Siblings							
	B	SE *B*	95% CI		r_{part}	R^2	ΔR^2	f^2
			Lower	Upper				
Constant	**18.74**	**.27**	**18.21**	**19.26**				
Demographics						**.03**	**.03**	.03
	-2.15	**.54**	**-3.21**	**-1.09**	**-.08**			
Race	*2.04*	*.67*	*.72*	*3.36*	*.06*			
MatEduc	*1.24*	*.56*	*.14*	*2.33*	*.04*			
Social Development						**.09**	**.06**	.07
VSOC	**-.14**	**.03**	**-.20**	**-.09**	**-.10**			
Behavior Problems						**.31**	**.22**	.32
CBCL-E	**.30**	**.03**	**.23**	**.36**	**.16**			
CBCL-I	**.48**	**.03**	**.42**	**.55**	**.27**			
Developmental Level						**.33**	**.02**	.04
Age	**-.53**	**.08**	**-.67**	**-.38**	**-.13**			
VEC	**-.64**	**.13**	**-.90**	**-.39**	**-.09**			

Bold=p<.001; *Italics*=p<.05; MatEduc=Maternal Education; VSOC=Vineland-II Social Standard Score; CBCL=Child Behavior Checklist; I=Internalizing; E=Externalizing; VEC=Vineland Expressive Communication V-Score

Table 2.4 Model-B: Predictors of SRS-Raw in probands only

SRS-Raw	Block 1					Block 2					Block 3				
	B	SE *B*	95% CI Lower	95% CI Upper	r_{part}	*B*	SE *B*	95% CI Lower	95% CI Upper	r_{part}	*B*	SE *B*	95% CI Lower	95% CI Upper	r_{part}
Constant	**97.56**	**.55**	**96.48**	**98.65**		**97.57**	**.55**	**96.49**	**98.65**		**97.64**	**.47**	**96.72**	**98.57**	
Demographics															
Gender	2.16	1.64	-1.06	5.37	.03	2.21	1.63	-.99	5.40	.03	1.53	1.40	-1.21	4.27	.02
Race	1.54	1.34	-1.09	4.16	.02	1.41	1.33	-1.20	4.02	.02	*2.24*	*1.14*	*.01*	*4.47*	*.03*
Mat Educ	**4.90**	**1.13**	**2.67**	**7.12**	**.09**	**5.09**	**1.13**	**2.88**	**7.31**	**.09**	*2.17*	*.97*	*.26*	*4.08*	*.04*
ASD-Symptoms															
ADOS-CSS						**1.69**	**.33**	**1.05**	**2.34**	**.11**	**2.08**	**.28**	**1.53**	**2.63**	**.13**
Behavior Problems															
CBCL-E											**.59**	**.05**	**.49**	**.69**	**.20**
CBCL-I											**.99**	**.06**	**.87**	**1.10**	**.30**
R^2		**.009**					**.021**					**.283**			
R^2 change		**.009**					**.011**					**.263**			
Effect size (f^2)		.009					.011					.367			

SRS-Raw	Block 4					Block 5				
	B	SE *B*	95% CI Lower	95% CI Upper	r_{part}	*B*	SE *B*	95% CI Lower	95% CI Upper	r_{part}
Constant	**97.75**	**.43**	**96.90**	**98.15**		**97.73**	**.40**	**96.95**	**98.51**	
Demographics										
Gender	-.59	1.29	-3.11	1.94	-.01	-.46	1.18	-2.78	1.86	-.01
Race	-.57	1.06	-2.64	1.51	-.01	.54	.98	-1.37	2.45	.01
Mat Educ	.33	.89	-1.42	2.08	.01	.02	.82	-1.59	1.63	.00
ASD-Symptoms										
ADOS-CSS	**1.73**	**.26**	**1.22**	**2.24**	**.11**	**.79**	**.25**	**.30**	**1.27**	**.05**
Behavior Problems										
CBCL-E	**.61**	**.05**	**.52**	**.71**	**.20**	**.52**	**.05**	**.43**	**.61**	**.17**
CBCL-I	**1.06**	**.05**	**.95**	**1.17**	**.31**	**.87**	**.05**	**.77**	**.97**	**.25**
Developmental Level										
Age	**1.58**	**.15**	**1.29**	**1.87**	**.17**	**1.50**	**.13**	**1.24**	**1.77**	**.17**
M1vsM3	**16.39**	**1.65**	**13.16**	**19.62**	**.16**	*3.33*	*1.64*	*.10*	*6.55*	*.03*
M2vsM3	**6.81**	**1.23**	**4.40**	**9.21**	**.09**	*2.74*	*1.14*	*.50*	*4.98*	*.04*
Nonverbal IQ	**-.12**	**.02**	**-.17**	**-.07**	**-.08**	**-.09**	**.02**	**-.14**	**-.05**	**-.06**
ADI-Current						**1.37**	**.07**	**1.24**	**1.50**	**.30**
R^2			**.401**					**.494**		
R^2 change			**.118**					**.093**		
Effect size (f^2)			.197					.184		

Bold=p<.001; *Italics*=p<.05; MatEduc=Maternal Education; ADOS-CSS=Autism Diagnostic Observation Schedule Calibrated Severity Score; CBCL=Child Behavior Checklist; I=Internalizing; E=Externalizing; M1=ADOS Module 1; M2=ADOS Module 2; M3=ADOS Module 3; ADI-Current=Autism Diagnostic Interview-Revised Current Behavior Algorithm Total

Table 2.5 Model-C: Predictors of Vineland Social standard scores for probands and siblings

VSOC	Probands							
	B	SE *B*	95% CI Lower	95% CI Upper	r_{part}	R^2	ΔR^2	f^2
Constant	**82.11**	**.76**	**80.61**	**83.60**				
Demographics						**.03**	**.03**	.03
Gender	.35	.47	-.57	1.26	.01			
Race	*-1.01*	*.39*	*-1.77*	*-.25*	*-.03*			
Mat Educ	**-.30**	**.33**	**-.94**	**.34**	**-.01**			
Behavior Problems						**.06**	**.02**	.02
CBCL-E	**-.13**	**.02**	**-.16**	**-.09**	**-.09**			
CBCL-I	**.09**	**.02**	**.05**	**.13**	**.05**			
Developmental Level						**.60**	**.54**	1.35
Age	**-.60**	**.05**	**-.70**	**-.50**	**-.15**			
VEC	**2.48**	**.06**	**2.36**	**2.60**	**.52**			
ASD-Symptoms						**.63**	**.03**	.09
SRS-Raw	**-.11**	**.01**	**-.13**	**-.10**	**-.18**			
	Probands							
	B	SE *B*	95% CI Lower	95% CI Upper	r_{part}	R^2	ΔR^2	f^2
Constant	**104.43**	**.43**	**103.58**	**105.28**				
Demographics						**.03**	**.03**	.03
Gender	.28	.46	-.63	1.19	.01			
Race	**-2.77**	**.57**	**-3.90**	**-1.64**	**-.10**			
Mat Educ	.09	.48	-.86	1.03	.00			
Behavior Problems						**.06**	**.03**	.03
CBCL-E	**-.10**	**.03**	**-.16**	**-.04**	**-.07**			
CBCL-I	.01	.03	-.05	.07	.01			
Developmental Level						**.25**	**.20**	.27
Age	**.25**	**.07**	**.12**	**.38**	**.08**			
VEC	**2.11**	**.10**	**1.91**	**2.30**	**.41**			
ASD-Symptoms						**.26**	**.01**	.01
SRS-Raw	**-.11**	**.02**	**-.14**	**-.07**	**-.11**			

Bold=p<.001; *Italics*=p<.05; VSOC=Vineland Socialization Standard Score; MatEduc=Maternal Education; CBCL=Child Behavior Checklist; I=Internalizing; E=Externalizing; VEC=Vineland Expressive Communication V-Scale Score; ASD=Autism Spectrum Disorder; SRS-Raw= Social Responsiveness Scale Raw Total

Table 2.6 Model-D: Predictors of P1-S1 SRS-Differences

SRS-Raw Difference	All participants							
	B	SE *B*	95% CI Lower	95% CI Upper	r_{part}	R^2	ΔR^2	f^2
Constant	**36.53**	**1.49**	**33.62**	**39.45**				
Demographics						**0.00**	**0.00**	0.00
Gender	.32	1.03	-1.70	2.35	.01			
Race	-.50	.65	-1.77	.77	-.01			
Mat Educ	.00	.53	-1.04	1.04	.00			
Social Development						**0.19**	**0.19**	0.24
VSOC	**-.53**	**.04**	**-.61**	**-.45**	**-.22**			
Behavior Problems - Both						**0.44**	**0.25**	0.45
Anxious/Depressed	**.27**	**.06**	**.15**	**.39**	**.07**			
Withdrawn/Depressed	**.74**	**.06**	**.63**	**.85**	**.23**			
Somatic Complaints	.09	.07	-.04	.23	.02			
Attention	**.74**	**.05**	**.64**	**.84**	**.25**			
Aggressive Behavior	**.30**	**.06**	**.17**	**.42**	**.08**			
Age/Language						**0.45**	**0.01**	0.01
Age	**-.72**	**.18**	**-1.07**	**-.36**	**-.07**			
VEC	-.01	.16	-.33	.31	.00			

SRS-Raw Difference	Only 6-18 year olds							
	B	SE *B*	95% CI Lower	95% CI Upper	r_{part}	R^2	ΔR^2	f^2
Constant	**31.90**	**1.99**	**28.00**	**35.79**				
Demographics						*0.01*	*0.08*	0.08
Gender	.30	1.22	-2.10	2.70	.01			
Race	-1.10	.78	-2.62	.43	-.03			
Mat Educ	.63	.62	-.59	1.86	.02			
Social Development						**0.18**	**0.17**	0.21
VSOC	**-.46**	**.05**	**-.55**	**-.36**	**-.20**			
Behavior Problems - Both						**0.44**	**0.26**	0.47
Anxious/Depressed	-.02	.08	-.17	.14	.00			
Withdrawn/Depressed	**.73**	**.07**	**.59**	**.88**	**.21**			
Somatic Complaints	.00	.08	-.16	.16	.00			
Attention	**.51**	**.07**	**.38**	**.65**	**.16**			
Aggressive Behavior	*.19*	*.09*	*.01*	*.37*	*.04*			
Behavior Problems - 6-18 only						**0.47**	**0.03**	0.06
Social Problems	**.52**	**.09**	**.34**	**.71**	**.12**			
Thought Problems	**.52**	**.08**	**.36**	**.68**	**.14**			
Rule Breaking	-.15	.11	-.37	.07	-.03			
Age/Language						**0.48**	**0.01**	0.02
Age	**-.93**	**.21**	**-1.34**	**-.52**	**-.09**			
VEC	-.24	.21	-.65	.17	-.03			

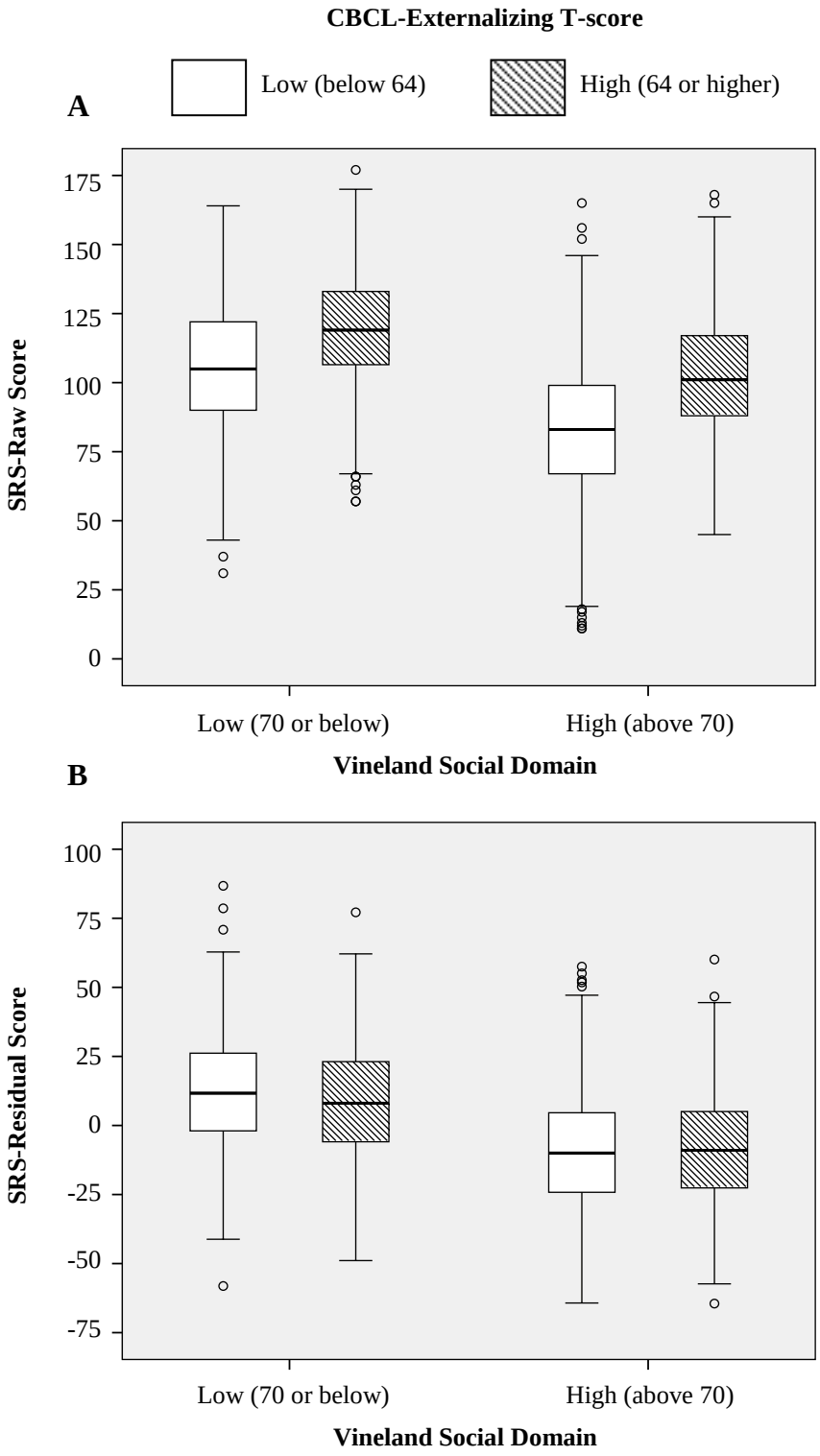
Bold=p<.001; *Italics*=p<.05; MatEduc=Maternal Education; VSOC=Vineland-II Social Domain Standard Score; CBCL=Child Behavior Checklist; I=Internalizing; E=Externalizing; VEC=Vineland Expressive Communication V-Scale Score

Table II.7 Model E: Predictors of proband SRS-Raw (without CBCL-I)

SRS-Raw	*B*	SE *B*	95% CI		r_{part}	R^2	ΔR^2	f^2
			Lower	Upper				
Constant	**97.68**	**.42**	**96.86**	**98.51**				
Demographics						**0.01**	**0.01**	0.01
Gender	-.96	1.26	-3.43	1.50	-.01			
Race	1.00	1.03	-1.03	3.03	.02			
Mat Educ	.40	.87	-1.31	2.11	.01			
ASD-Symptoms						**0.02**	**0.01**	0.01
ADOS-CSS	*.53*	*.26*	*.03*	*1.04*	*.03*			
Behavior Problems						**0.20**	**0.18**	0.22
CBCL-E	**.93**	**.04**	**.84**	**1.01**	**.35**			
Developmental Level						**0.31**	**0.11**	0.16
Age	**1.75**	**.14**	**1.47**	**2.03**	**.19**			
M1vsM3	1.23	1.74	-2.18	4.64	.01			
M2vsM3	1.08	1.21	-1.29	3.45	.01			
Nonverbal IQ	*-.05*	*.02*	*-.10*	*-.01*	*-.03*			
ADI-Current	**1.57**	**.07**	**1.43**	**1.70**	**.35**	**0.43**	**0.13**	0.22

Bold=p<.001; *Italics*=p<.05; MatEduc=Maternal Education; VSOC=Vineland-II Social Domain Standard Score; CBCL=Child Behavior Checklist; E=Externalizing; M1=ADOS Module 1; M2=ADOS Module 2; M3=ADOS Module 3; ADI-Current=Autism Diagnostic Interview-Revised Current Behavior Algorithm Total

Figure 2.1 SRS-Raw and SRS-Residual by Vineland-Social and CBCL Groups



APPENDIX A

1. Collaborative Programs of Excellence in Autism (CPEA) Diagnostic Criteria

CPEA Diagnoses of Autism, Asperger Disorder or ASD are determined hierarchically and based upon scores from the ADI-R, ADOS, and the BEC diagnosis; criteria for Aspergers also include consideration of child age, IQ and language milestones. These criteria were established by the CPEA to ensure standardized diagnostic classification across sites and adopted by the SSC under the same rationale. The criteria used in the SSC were slightly modified from those described by Lainhart and colleagues (2006) and are as follows:

A. For a CPEA diagnosis of "Autism" an individual must meet the following:

   i. **ADI-R classification of "Autism."** This is based on meeting published cut-offs on the ADI-R diagnostic algorithm (LeCouteur et al., 2003) in the domains of Reciprocal Social Interaction (RSI), Communication, Restricted, Repetitive & Stereotyped Patterns (RRB) of Behaviors, and Age of Onset.

   ii. **ADOS classification of "Autism" or "Autism Spectrum."** This is based on meeting published cut-offs on the revised diagnostic algorithms for Modules 1-3 (Gotham et al., 2009) and cut-offs on the originally published diagnostic algorithms for Module 4 (Lord et al., 2000).

   iii. **BEC Diagnosis of "Autism," "Autism Spectrum" or "Aspergers."**

B. For a CPEA diagnosis of "Aspergers" an individual must *not* meet criteria for "Autism" and must also meet the following:

   i. **Chronological age of 5 years or older**

   ii. **Verbal IQ of 80 or above**

   iii. **Age of First Words (from the ADI-R) is 24 months or younger**

   iv. **Age of First Phrases (from the ADI-R) is 33 months or younger**

   v. **ADI-R classification is not "Autism"**

   vi. **ADI-R RSI domain score is 10 or higher**

   vii. **ADI-R RRB domain score is 2 or higher**

   viii. **ADOS classification of "Autism" or "Autism Spectrum" or ADOS Social+Communication Total (based on originally published algorithms) is 4 or higher.**

C. For a CPEA diagnosis of "Autism Spectrum Disorder" an individual must *not* previous criteria for "Autism" or "Aspergers" and must also meet the following:

   i. **ADI-R classification of "Autism Spectrum."** This is based on CPEA criteria (Lainhart et al., 2006; Risi et al., 2006), which requires one of the following:

      i. Meeting cut-offs on the Social *and* Communication domains

      ii. Meeting cut-offs on *either* the RSI *or* Communication domain and score within 2 points of the cut-off on the other

      iii. Score within 1 point on *both* the RSI *and* Communication domains.

ii. **ADOS classification of "Autism Spectrum."** This is based on revised diagnostic algorithms for Modules 1-3 and originally published diagnostic algorithms for Module 4.

D. If none of the above criteria are met, the participant receives a CPEA diagnosis of "Nonspectrum"

2. Study Procedures

Participants were drawn from 2,442 families evaluated as part of the Simons Simplex Collection (SSC). Eighty-two percent of probands had at least one unaffected sibling enrolled in the study. Probands were excluded from analyses if they were missing more than 10 SRS items (n=10) or administered an ADOS Module 4 (because Calibrated Severity Scores (ADOS-CSS) are not available for Module 4; n=64). Eighty-two siblings were not administered the SRS. This yielded a total of 2,368 probands and 1,913 siblings.

A list of measures completed for the proband and sibling is available at: http://sfari.org/sfari-initiatives/simons-simplex-collection/ssc-instruments. Additional data regarding the relationship of SRS-Raw to other measures not included in the present analyses (e.g., Social Communication Questionnaire – Lifetime Version, Aberrant Behavior Checklist, etc.) is available from authors upon request. Order of assessments sometimes varied, but child questionnaires (e.g., SRS, CBCL, etc.) were generally completed prior to the assessment visit and ADOS and ADI-R were often completed simultaneously by two different research-reliable clinicians. Senior clinicians reviewed all available information (including observation of the proband in–person or via video) to specify a Best Estimate Clinical (BEC) diagnosis (see Lord et al., 2012).

Inclusion criteria for probands were as follows:

A. Between 4 and 18 years of age
B. Meet Collaborative Programs of Excellence in Autism (CPEA) criteria for a diagnosis of Autism, Asperger Disorder or Autism Spectrum Disorder (described below)
C. Receive a BEC diagnosis of an autism spectrum disorder
D. Have a nonverbal mental age of at least 18 months

Families were excluded if any of the following were met:

A. Proband had significant hearing, vision or motor problems that may affect interpretation of behavioral data
B. Proband had Fragile X syndrome, Tuberous sclerosis, Down syndrome or a significant early medical history (e.g., very low birth rate)
C. If any known 1st, 2nd or 3rd degree relative had ASD
D. If the sibling had substantial language or psychological problems related to ASD*

*Siblings were screened using the Vineland-II and parent-questionnaires (see http://sfari.org/sfari-initiatives/simons-simplex-collection/ssc-instruments) and excluded if they had substantial language or psychological problems related to ASD. Although questionnaires' clinical cut-offs were used to flag siblings for follow-up, SSC inclusion/exclusion was based on clinicians' impressions of the sibling (i.e., a child with an elevated SRS score may be included in the SSC if the clinician's impression based on observation and parent report was that s/he did not have ASD).

Additional information regarding the collection is provided by Fischbach & Lord (2011) and can be obtained from http://sfari.org/sfari-initiatives/simons-simplex-collection.

3. Post-hoc Analyses

Post-hoc analyses were conducted to better understand associations between SRS-Raw and behavior problems.

Model C: Predictors of Vineland Social standard scores for probands and siblings.

First, we considered whether the relationship might actually reflect the impact of behavior problems on social development (i.e., children with more behavior problems have higher SRS-Raw because they have poorer social skills), which could explain the SRS-CBCL association observed for both proband and siblings. If this were true, Vineland-II Social domain scores (VSOC), a standardized measure of social development, should also be significantly associated with CBCL-I and CBCL-E. To assess this, we examined predictors of VSOC, with SRS-Raw entered last to assess their significance as a predictor after all other variables of interest were controlled. As shown in Table 2.5, while CBCL scores were significant predictors of VSOC, this relationship was much weaker (f^2=.024) than that between the CBCL and SRS-Raw. Together, age and expressive communication explained over half of the variance in proband VSOC, whereas CBCL and SRS-Raw each accounted for approximately 2-3%. Siblings showed similar, albeit weaker, associations with developmental level and nearly identical relationships with CBCL and SRS-Raw. The minimal relationship between VSOC and CBCL suggest that the high association between CBCL and SRS-Raw does not simply reflect the impact of behavior problems on social skills.

Model D: Predictors of P1-S1 SRS-Differences.

Next, how the profile of differences between proband's and siblings' CBCL Syndrome Scale scores related to differences in SRS-Raw was examined. This is more informative than simply saying that high CBCL scores predict high SRS-Raw in each group separately because it shows that the variation in SRS-Raw is related to variation in particular CBCL scales. Relationships between proband-sibling differences on Syndrome Scales containing items related to core symptoms of ASD (e.g., CBCL-Withdrawn/Depressed) may reflect ASD-specific variation. However, if there were similar relationships between proband-sibling differences on the SRS and Syndrome Scales that do not have items which overlap with core ASD symptoms (e.g., CBCL-Attention), this would caution against such an interpretation. To explore this question, differences between probands and siblings from the same family were computed for SRS-Raw, VSOC, the CBCL Syndrome Scales, age, and VEC. The first analysis included 1,864 families with both a proband and a sibling where each had the five Syndrome Scales that overlap across the two CBCL forms (Anxious-Depressed, Withdrawn-Depressed, Somatic Complaints, Attention Problems, Aggressive Behavior). To allow examination of all eight Syndrome Scales (previous 5 plus Social Problems, Thought Problems, Rule-Breaking Behavior), the second analysis included only 6 to 18 year olds (n=1,171). As shown in Table 2.6, even after controlling for demographics and VSOC, greater proband-sibling differences on several CBCL scales predicted greater proband-sibling differences in SRS-Raw. Notably, relationships between differences in SRS-Raw and differences in CBCL-Attention and CBCL-Withdrawn were of similar magnitude to the relationship between SRS-Raw and VSOC.

Model E: Predictors of proband SRS-Raw (without CBCL-I).

Finally, although the Attention Problems subscale does not include items assessing what are likely to overlap with core ASD symptoms, a sizeable proportion of the association with CBCL scores was being accounted for by questions that may reflect difficulties associated with core ASD symptoms (i.e., from the Withdrawn/Depressed subscale, which is part of the internalizing score). Given the strong correlation between CBCL-I and CBCL-E ($r=.53$, $p<.001$), it was of interest to see whether externalizing behaviors would be a significant predictor of SRS-Raw in the absence of CBCL-I. To assess this, a regression (Model-E) predicting SRS-Raw was fit with the following predictors: Demographics (gender, race, maternal education), ASD-Symptoms (ADOS-CSS), Behavior Problems (CBCL-E), Developmental Level (Age, ADOS Module, NVIQ), and ADI-Current. As shown in Table 2.7, the relationship between CBCL-E and SRS-Raw was significant and as strong as the relationship between ADI-Current and SRS-Raw, after controlling for all other factors. While age remained the only significant developmental predictor in the final model, language and NVIQ had been significant prior to inclusion of ADI-Current. Their lack of significance in the final model likely reflects the strong association between these factors and ADI-Current demonstrated by Hus & Lord, 2013.

References

Achenbach, T.M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles*. ASEBA Burlington, VT.

Aylward, G. P. (2009). Developmental Screening and Assessment: What Are We Thinking? Journal of Developmental & Behavioral Pediatrics, 30(2), 169–173.

Bölte, S., Poustka, F., & Constantino, J.N. (2008). Assessing autistic traits: Cross-cultural validation of the Social Responsiveness Scale (SRS). *Autism Research, 1*(6), 354-363.

Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., et al. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *The British Journal of Psychiatry*, *191*(6), 554-559.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Psychology Press.

Constantino, J.N. (2011). The quantitative nature of autistic social impairment. *Pediatric Research*, *69*(5), 55R-62R.

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

Constantino, J.N., & Todd, R.D. (2003). Autistic traits in the general population: A twin study. *Archives of General Psychiatry*, *60*(5), 524-530.

Constantino, J.N., & Todd, R.D. (2005). Intergenerational transmission of subthreshold autistic traits in the general population. *Biological Psychiatry*, *57*(6), 655-660.

Constantino, J.N., Davis, S., Todd, R.D., Schindler, M.K., Gross, M., Brophy, S.L., et al. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised, *Journal of Autism and Developmental Disorders*, *33*(4), 427-33.

Constantino, J.N., Gruber, C.P., Davis, S., Hayes, S., Passanante, N., & Przybeck, T. (2004). The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry*, *45*(4), 719-726.

Constantino, J.N., Hudziak, J.J., & Todd, R.D. (2003). Deficits in reciprocal social behavior in male twins: evidence for a genetically independent domain of psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*(4), 458-467.

Constantino, J.N., Lajonchere, C., Lutz, M., Gray, T., Abbacchi, A., McKenna, K., et al. (2006). Autistic social impairment in the siblings of children with pervasive developmental disorders. *American Journal of Psychiatry*, *163*(2), 294-296.

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

Constantino, J.N., Przybeck, T., Friesen, D., & Todd, R.D. (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *Journal of Developmental and Behavioral Pediatrics*, *21*(1), 2-11.

Constantino, J.N., Zhang, Y., Frazier, T., Abbacchi, A.M., & Law, P. (2010). Sibling recurrence and the genetic epidemiology of autism. *American Journal of Psychiatry*, *167*, 1349-1356.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E.H., Leventhal, B.L., 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.

Duvall, J.A., Lu, A., Cantor, R.M., Todd, R.D., Constantino, J.N., & Geschwind, D.H. (2007). A quantitative trait locus analysis of social responsiveness in multiplex autism families. *American Journal of Psychiatry*, *164*(4), 656-662.

Dunlap, W.P., Xin, X, & Myers, L. (2004). Computing aspects of power for multiple regression. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 695.

Fischbach, G. D., & Lord, C. (2010). The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron*, *68*(2), 192-195.

Georgiades, S., Szatmari, P., Duku, E., Zwaigenbaum, L., Bryson, S., Roberts, W., et al. (2010). Phenotypic overlap between core diagnostic features and emotional/behavioral problems in preschool children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *41*, 1321-1329.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(5), 693-705.

Grzadzinski, R., Di Martino, A., Brady, E., Mairena, M. A., O'Neale, M., Petkova, E., Lord, C., et al. (2011). Examining autistic traits in children with ADHD: Does the autism spectrum extend to ADHD? *Journal of Autism and Developmental Disorders*, *41*(9), 1178–1191.

Hus, V., & Lord, C. (2013). Effects of child characteristics on the Autism Diagnostic Interview-Revised: Implications for use of scores as a measure of ASD severity. *Journal of Autism and Developmental Disorders*, 43(2), 371–381.

Kanne, S.M., Abbacchi, A.M., & Constantino, J.N. (2009). Multi-informant Ratings of Psychiatric Symptom Severity in Children with Autism Spectrum Disorders: The

Importance of Environmental Context. *Journal of Autism and Developmental Disorders*, *39*(6), 856-864.

Lainhart, J. E., Bigler, E.D., Bocian, M., Coon, H., Dinh, E., Dawson, G., et al. (2006). Head circumference and height in autism: A study by the collaborative program of excellence in autism. *American Journal of Medical Genetics Part A*, *140A*(21), 2257-2274.

Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D. M., Ousley, O., et al. (2012). A Multisite Study of the Clinical Diagnosis of Different Autism Spectrum Disorders. *Archives of General Psychiatry*, *69*(3), 306-313.

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

Mayes, S.D., & Calhoun, S.L. (2010). Impact of IQ, age, SES, gender, and race on autistic symptoms. *Research in Autism Spectrum Disorders*, *5*(2), 749-757.

Rea, L.M., & Parker, R.A. (1992). Designing and conducting survey research: A comprehensive Guide. *San Francisco: Joyssey-Bass*.

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

Sparrow, S.S., Cicchetti, D.V., & Balla, D.A. (2005). *Vineland Adaptive Behavior Scales, Second Edition*. Circle Pines, MN: America.

Veenstra-VanderWeele, J., & Warren, Z. (2011). Social communication deficits in the general population: How far out does the autism spectrum go? *Journal of the American Academy of Child and Adolescent Psychiatry*, *50*(4), 326–328.

Virkud, Y.V., Todd, R.D., Abbacchi, A.M., Zhang, Y., & Constantino, J.N. (2009). Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, *150*(3), 328–334.

Warren, Z., Vehorn, A., Dohrmann, E., Nicholson, A., Sutcliffe, J. S., & Veenstra-Vanderweele, J. (2012). Accuracy of phenotyping children with autism based on parent report: what specifically do we gain phenotyping "rapidly"? *Autism Research*, *5*(1), 31–38.

CHAPTER III

Standardizing ADOS domain scores: Separating severity of Social Affect and Restricted & Repetitive Behaviors

The search to elucidate underlying biological mechanisms which cause or increase risk for autism spectrum disorders (ASD) has been made more complicated by the marked phenotypic heterogeneity associated with this developmental disorder (State & Levitt, 2011). Diagnostic criteria focus on the presence or absence of specific behaviors or impairments in three domains: *Communication, Reciprocal Social Interaction,* and *Restricted and Repetitive Stereotyped Behaviors and Interests* (American Psychiatric Association [APA], 2000; World Health Organization [WHO], 1992). However, ASD symptoms within each domain vary considerably in type and severity, depending upon an individual's age, language level, and IQ.

Current nosology attempts to capture some of this variation through categorical diagnoses (e.g., Autistic Disorder, Asperger's Disorder and Pervasive Developmental Disorder, Not Otherwise Specified; APA, 2000). However, research has demonstrated that differentiations made between ASD subgroups are often not reliable across different sites (Lord et al., 2012). In addition, in several studies, items reflecting social and communication impairments comprised a single factor on ASD diagnostic instruments (e.g., Frazier et al., 2012; Gotham, Risi, Pickles, & Lord, 2007). In light of these findings, proposals for DSM-5 and ICD-11 call for subgroups to be subsumed into a single category of ASD defined by two behavioral domains: *Social/Communication Deficits* and

Fixated or *Restricted Interests and Repetitive Behaviors* (APA, 2011, WHO, 2012). Several initial studies support these proposed changes (Frazier et al., 2012; Huerta et al., 2012, Mandy, Charman, & Skuse, 2012, though see Mattila et al., 2011 and McPartland, Recihow & Volkmar, 2012). To further capture the heterogeneity, criteria for assessing severity within each domain are recommended.

As these changes are implemented, many of the currently used ASD diagnostic instruments will need to be revised to more accurately reflect new DSM-5 and ICD-11 criteria, both to inform diagnosis and to describe severity of symptoms within each behavioral domain. For example, the diagnostic algorithm of the Autism Diagnostic Interview – Revised (ADI-R; Rutter, Le Couteur, & Lord, 2003), a widely-used parent interview in autism research, is divided into three domains reflecting the current DSM-IV and ICD-10 criteria for Autistic Disorder, whereas the Social Responsiveness Scale (SRS; Constantino & Gruber, 2005), a caregiver questionnaire, relies on a single total score for diagnostic classification. In contrast, the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord, et al., 2012), a clinician-administered observational assessment, has recently revised diagnostic algorithms that comprise two behavioral domains (referred to as Social Affect [SA] and Restricted and Repetitive Behaviors [RRB]) and provide cut-offs for ASD classification (Gotham et al., 2007). In addition, total scores from the revised ADOS algorithms have been standardized to provide a continuous measure of overall ASD symptom severity that is less influenced by child characteristics, such as age and language skills, than raw totals (Calibrated Severity Scores [CSS]; Gotham, Pickles, & Lord, 2009). These scores can be used to compare ASD symptom severity across individuals of different developmental levels. As such,

they provide a "purer" metric of overall ASD severity than raw totals from the ADI-R and SRS, for which studies have demonstrated strong influences of child characteristics, such as age, language level, and non-ASD specific behavior problems (e.g., Constantino, Hudziak, & Todd, 2003; Hus, Bishop, Gotham, Huerta & Lord, 2013; Hus & Lord, 2013).

Although the ADOS-CSS may provide some advantages over these other measures of general ASD severity, the nature of the symptoms underlying an individual's CSS may vary greatly. For example, an ADOS-CSS of 10, indicating the highest level of severity, may be assigned to a child with very significant social-communication impairments who exhibits few repetitive behaviors during the ADOS. The same score may also be assigned to a child who has moderate levels of impairments in both domains or very high levels of repetitive behaviors and more subtle social-communication difficulties. Social-communication difficulties often pertain to a "lack" of typical behaviors that are pervasive across social contexts, such as reduced use of gestures or eye contact or reduced frequency of appropriate social responses, making them more easily-observable during brief interactions. In comparison, RRBs are often characterized by the presence of an abnormal behavior, such as hand flapping, sensory examination of materials or excessive references to a particular topic. Because RRBs may only occur in particular conditions (e.g., hand flapping when a child is very excited or prolonged discussion of a topic only if it is raised), it is more difficult to assess them in a short period of time. Therefore, it is important to acknowledge that, when assessing and comparing symptom severity in different domains, the ADOS as a source of information, particularly about RRBs, is limited by both time and context. While the presence of

RRBs during this brief observation may be clinically significant, the absence of these behaviors in this time-limited, standardized context must be interpreted more cautiously. Nevertheless, research has suggested that both social-communication and repetitive behaviors measured by the ADOS are surprisingly good predictors of diagnosis (e.g., Lord, Risi, DiLavore, Shulman, Thurm, & Pickles, 2006).

Separate calibration of these distinct domains is needed to provide a clearer picture of ASD severity. For example, calibrated domain scores would allow for examination of two dimensions (SA and RRB), which may have distinct developmental trajectories or respond differently to intervention. In large samples, researchers could use estimates of social-communication and repetitive behavior severity to increase phenotypic homogeneity by clustering individuals according to similar levels of severity in each domain (e.g., high SA and RRB; high SA and low RRB, etc.). In smaller studies that cannot afford the loss of power resulting from sample stratification, researchers might use continuous scores to statistically control for differences in one domain while focusing on the other. Separately calibrated domain scores may also be useful in genetic and neurobiological studies seeking to draw associations between biological mechanisms and specific behavioral domains, many of which currently rely on raw domain totals (e.g., Dichter, Richey, Rittenberg, Sabatino, & Bodfish, 2011). While some studies have controlled for effects of age or IQ in individual samples (e.g., Di Martino et al., 2011), use of calibrated scores may facilitate comparisons across samples comprised of individuals of varying developmental levels.

The goal of the current study was to separately calibrate raw totals from the ADOS SA and RRB domains to reduce the effects of child characteristics and increase

the utility of these scores as continuous measures of social-communication and of repetitive behavior symptom severity.

Methods

Participants

For comparability, the same sample used to standardize the overall ADOS total (see Gotham et al., 2009) was also employed to calibrate separate severity metrics for the Social Affect (SA) and Restricted, Repetitive Behavior (RRB) domains. Briefly, this included data from 1,415 individuals ranging in age from 2 to 16 years. With repeated assessments for 25% of the sample, data from 2,195 ADOSes with contemporaneous best estimate clinical diagnoses were available for analysis. Of these assessments, 1,786 cases were given an autism spectrum disorder diagnosis (ASD; 1,187 Autistic Disorder, 599 Other-ASD) and 409 had a Non-ASD diagnosis. Non-ASD diagnoses included language disorders (27%), nonspecific intellectual disability (20%), Down syndrome (14%), oppositional defiant disorder or ADD/ADHD (13%), mood or anxiety disorders (8%), Fetal Alcohol Spectrum Disorders (7%), other genetic or physical disabilities, such as Fragile X or mild cerebral palsy (6%) and early developmental delays (5%).

Individuals were consecutive referrals to specialty clinics in Ann Arbor, Michigan and Chicago, Illinois, and participants in research studies conducted through the University of North Carolina - Chapel Hill, University of Chicago, and University of Michigan. All participants provided informed consent and all procedures related to this project were approved by institutional review boards at the University of Chicago or University of Michigan. Sample characteristics are provided in Table 3.1.

Procedure

The ADOS was conducted as part of a clinical or research evaluation (see Gotham et al., 2009 for more detailed procedures). All ADOSes were administered and scored by a clinical psychologist or trainee who met standard requirements for research reliability. The Pre-Linguistic ADOS (PL-ADOS; DiLavore, Lord, & Rutter, 1995) was given in 418 (19%) assessments and a pilot version of the ADOS-Toddler (Luyster et al., 2009) was given in 82 assessments (4%). For both measures, scores from items identical to those in the Module 1 algorithms were used. Verbal and/or nonverbal IQ scores were available for 2009 (92%) of participants. These were derived from a developmental hierarchy of cognitive measures (see Lord et al., 2006), most frequently the Mullen Scales of Early Learning (Mullen, 1995) and the Differential Ability Scales (Elliott, 1990). Best estimate clinical diagnoses were made by a supervising clinical psychologist and/or a child psychiatrist after review of all assessment data (including, at a minimum, the ADOS and cognitive scores).

Standardization of raw totals.

Calibration of each domain began by following a similar procedure to that described for standardization of overall ADOS totals (Gotham et al., 2009). Only assessments from individuals with ASD were used for raw domain total standardization. This included all assessments with a corresponding best estimate clinical diagnosis of autism or ASD, as well as data from 13 individuals who had ADOS data with a contemporaneous Non-ASD diagnosis but who were later diagnosed with ASD (total n=1,807 assessments from 1,118 individuals). Participants were first divided into the 18 age/language groups used for the calibration of the overall raw totals. SA and RRB scores

were compared separately for each 1-year chronological age group within a given cell to ensure that distributions of the domain scores were comparable. Some of the 18 cells were then collapsed due to comparable distributions (likely due to the reduced range of scores in each domain compared to the overall totals). This resulted in 12 age/language cells (See Figure 4.1; note that the raw total-to-calibrated score mapping for the RRB domain could have been further collapsed into two Module 2 cells, 2-3 year olds and above 4 years; however, these were left expanded across 4 cells so that both domains would have the same number of cells).

In the overall total calibration, ADOS diagnostic classifications were used to anchor raw totals to ranges of severity scores. That is, raw totals corresponding to an ADOS classification of "Autism" were mapped on to CSS of 6-10, "ASD" to CSS of 4-5 and "Nonspectrum" to CSS of 1-3. This was done to make the metric more generalizable to other samples, as we cannot assume that the datasets used for calibration in all developmental cells were representative of the heterogeneous ASD population. Next, the range of raw totals assigned to each point on the 10-point severity scale was determined by the percentiles of available data within that classification range (Gotham et al., 2009). Because there are not separate SA cut-offs for "Autism" and "ASD" classifications, the same percentiles used for mapping raw ADOS totals (i.e. SA+RRB) to the 10-point scale were used to inform the mapping of raw SA totals to SA-CSS within each of the 12 age/language cells. Raw total-to-calibrated score mappings were then adjusted so that, for each of the 5 diagnostic algorithm groups (Gotham et al., 2007), sensitivity for individuals receiving an ADOS classification of "Autism" and an SA-CSS greater than or equal to 6 was, if possible, at or above 90%. Within algorithm groups, the lowest

individual cell sensitivity was .89 for Module 2, 2-3 year olds. A goal of 80% sensitivity across algorithm groups was set for individuals with an "Autism Spectrum" ADOS classification and an SA-CSS of 4 or higher. Sensitivity for individual developmental cells within algorithm groups was sometimes lower in groups with few participants; however, considering cells with greater than 20 participants, only Module 3, 3-5 year olds (n=59) fell just below this threshold, with a sensitivity of .78. Finally, adjustments were made to ensure that specificity (individuals with a "Nonspectrum" ADOS classification and SA-CSS less than or equal to 3) was, if possible, at least 80% for each algorithm. Within algorithms, only the Module 2, 5-6 year old cell fell below this threshold, with a specificity of .76.

Because the RRB domain is limited to a range of 9 points (0-8), it was not possible to use all 10 points in the severity metric for this domain. However, given concerns that SA- and RRB-CSS scores may be misinterpreted if they are not on a comparable scale, it was decided to maintain the full 10-point range and have some points on the severity scale for which no raw scores were assigned. Thus, as with SA-CSS, percentiles from mapping of raw overall totals were used to inform mapping of raw RRB totals to the calibrated metric. This resulted in the raw RRB totals mapping on to CSS values of 5-10. These distributions were skewed compared to the overall and SA-CSS scales and reflect the trade-off in using the ADOS as a measure of RRBs: while a lack of RRBs is difficult to interpret, the presence of RRBs during this brief observation is more meaningful as an indication of greater severity. Given the lower sensitivity of repetitive behaviors in the limited context in which they may be observed during the ADOS, a goal of 80% sensitivity was set for individuals receiving an ADOS

classification of "Autism" and RRB calibrated scores of 6 or greater; Module 3, 2-5 year olds fell just below this threshold with a sensitivity of 77%. No sensitivity threshold was set for individuals with an "Autism Spectrum" classification. A goal of 80% specificity was set for scores less than or equal to 6. For individual cells with greater than 20 participants, the lowest specificity was 79% for Module 3, 6-16 year olds.

Table 2 shows the mappings of raw SA and RRB totals to the 10-point severity scale for each of the 12 calibration cells.

Associations between participant characteristics, raw domain totals and calibrated domain scores.

Following procedures in Gotham et al., 2009, separate linear regression analyses were conducted using the sample of participants with ASD who had contemporaneous demographic data (N=1,369) to examine the influences of child characteristics on raw domain totals and calibrated domain scores. The child's verbal and nonverbal IQs and mental ages were entered into the first block, followed by child chronological age, gender, maternal education and race in the second block. Only model R^2 are reported because interpretation of the meaning of these individual coefficients is limited by multicollinearity. Next, significant predictors were entered into Forward Stepwise models to assess the relative contributions of these variables in predicting raw domain totals and calibrated domain scores. (Results from analyses including Non-ASD participants are available from authors. Consistent with the results for the participants with ASD, when applied to the entire clinically-referred sample, standardized severity scores were less influenced by participant characteristics than were raw domain totals.)

Results

Comparison of Raw Domain Totals and Calibrated Domain Scores by Calibration Cell

As shown in Table 3 and Figures 4.2a and 4.2c, distributions of raw SA and RRB domain totals varied significantly by age/language group. Across algorithms reflecting different language levels, individuals with less language had higher scores than those who were more verbally fluent. Within algorithm groups, older children and adolescents tended to have higher scores than toddlers and young children. In contrast, calibrated SA and RRB domain scores were more comparable across calibration cells, though not uniform (see Table 3 and Figures 4.2b and4.2d). Notably, children who were verbally fluent (i.e., Module 3) have a wider distribution of RRB-CSS scores compared to children of other language levels. This reflects the somewhat larger proportion of verbally fluent children (8.5-12.9%) that did not have repetitive behaviors during the ADOS (i.e., received a RRB-CSS of 1).

As noted above, ADOS *classifications*, which are based on raw overall totals (SA+RRB) were used to anchor the raw total-to-overall severity score mappings for the domains to specific calibrated score ranges (e.g., "Autism" to CSS of 6-10). Using percentiles from the raw total-to-overall CSS mapping to inform raw domain totals-to-domain severity score mappings, mean SA-CSS and RRB-CSS also distinguished between individuals grouped by clinicians' best estimate clinical diagnoses (i.e., Autism vs. Other-ASD vs. Non-ASD diagnoses; SA-CSS: $F(2,2192)=974.43$, $p\leq.001$; RRB-CSS: $F(2,2192)=421.35$, $p\leq.001$). Nonetheless, there was marked overlap in the distribution of scores across the three diagnostic groups (see Figure 4.3a and 4.3b).

Correlations Between Domain Calibrations and Overall Calibrated Severity Score

In the ASD sample, associations between SA-CSS and RRB-CSS were significant, but weak (r=.25; Cohen, 1988). Although correlations between each of the domain calibrated scores and the overall CSS were both strong, the association between SA-CSS and CSS (*r*=.89) was greater than that observed for RRB-CSS and CSS (*r*=.57). This is a reflection that the overall total from which the CSS is derived is comprised of a greater proportion of items from the SA domain than the RRB domain.

Predictors of SA-Raw and SA-CSS

The final model including all predictors explained a total of 45% of variance in the SA-Raw total. Verbal IQ and maternal education (mothers with graduate/professional degrees vs. all others) emerged as significant predictors of SA-Raw. In contrast, the same model accounted for only 13.1% of the variance in the SA-CSS, with verbal IQ and nonverbal IQ both making small, but significant contributions to the calibrated SA score. Thus, although there is still a significant association between SA-CSS and the child's cognitive level, the calibrated SA scores are markedly less influenced by child cognitive level than SA-Raw.

Next, verbal IQ, nonverbal IQ, and maternal education were entered into a Forward Stepwise model to assess the relative contributions of each of these variables in predicting SA-Raw. As shown in Table 4, verbal IQ accounted for the majority of variance (43%) and the contributions of nonverbal IQ and maternal education were minimal (0.3% and 0.2%, respectively). In the Forward model predicting SA-CSS, verbal IQ accounted for 10.5% of variance while nonverbal IQ explained an additional 0.4%; maternal education was excluded by the model, indicating that it was not significant (see

Table 4). These results reflect a reduction in the influence of verbal IQ from a large effect on SA-Raw (R=.66) to a small-to-medium effect on SA-CSS (R=.33; Cohen, 1988; McCarthy et al., 1991). It is noteworthy that verbal and nonverbal IQ were highly correlated (r=.76) and when verbal IQ was removed as a predictor, nonverbal IQ accounted for 21.8% of variance in SA-Raw and only 4.3% in SA-CSS; both models excluded maternal education as a predictor.

Predictors of RRB-Raw and RRB-CSS

Child characteristics such as IQ explained much less variance in raw RRB totals (i.e., 15.3%). Significant predictors included verbal IQ, nonverbal IQ, and race (African American vs. all others). In the Forward Stepwise Model, verbal IQ, nonverbal IQ and race each remained significant predictors of RRB-Raw (see Table 4). Verbal IQ accounted for the majority of variance (11.7%) and nonverbal IQ and race each made small contributions (1.4% and 1.1%, respectively). Again, if verbal IQ was excluded from the models, nonverbal IQ explained 11.4% and race explained 0.8% of variance in RRB-Raw.)

Calibrated RRB scores reduced the influence of child characteristics; in the end, child characteristics explained only 5.5% of the variance, with verbal IQ, nonverbal IQ and race emerging as small, but significant predictors of RRB-CSS. In the Forward Model predicting RRB-CSS, nonverbal IQ explained 3.5% of the variance in RRB-CSS; verbal IQ and race accounted for an additional 0.5 and 0.6%, respectively.

Case Summaries

Four children with ASD diagnoses were chosen to demonstrate the utility of the newly calibrated domain scores for separately examining the severity of social and repetitive behaviors over time (see Table 5 for child characteristics at first and last assessments). Each child's SA-CSS and RRB-CSS are plotted by age in Figure 4.4. Overall CSS scores are also provided; in many cases the overall CSS and SA-CSS follow similar, if not identical, trajectories, again reflecting that the overall total from which the CSS is derived is comprised of a greater proportion of items from the SA domain than the RRB domain.

Case 1. "Bianca," a Caucasian female, was diagnosed with autism at 4 years of age when she was first seen as a clinical referral (see Gotham et al., 2009). Her overall CSS suggests that her symptom severity was relatively stable across early childhood, followed by a gradual decrease in severity throughout late childhood and early adolescence. Her SA-CSS follows a similar trajectory, reflecting persistent difficulties with eye contact and unusual social overtures accompanied by an increase in use of gestures and shared enjoyment with the examiner. In contrast, her RRB-CSS follows a quite different pattern, with a RRB-CSS of 10 at Bianca's first assessment (reflecting her exhibition of sensory-seeking behaviors, delayed echolalia, repetitive asking of questions and repeated lining up of toys). This was followed by a considerable decrease in severity at age 5 and a year of relative stability, during which time she demonstrated some repetitive speech and mild preoccupations with a particular musician, but no hand and finger mannerisms. Although Bianca did not demonstrate repetitive behaviors when she was assessed at 8 years old, in early adolescence, she again exhibited clear hand and

finger mannerisms and engaged in somewhat repetitive speech (though recall that there is not a RRB-CSS of 2-4, so the fluctuation in severity later childhood may appear greater than it actually was).

Case 2. "Joey," a Caucasian male, was first seen as a clinical referral at 2 years, 10 months of age, at which time he received a diagnosis of PDD-NOS. When first seen, he exhibited severe social-communication symptoms (i.e., an SA-CSS of 10 demonstrating poor eye contact and very limited social overtures), but mild repetitive behaviors (RRB-CSS of 5 reflecting very brief repetitive behaviors) during the ADOS. In his subsequent assessments, there was an apparent increase in repetitive behaviors due to his use of stereotyped language (e.g., "That's all folks!"), accompanied by an improvement in the social affect domain (i.e., improvements in eye contact and more frequent and appropriate overtures). At age 7 years, 7 months, Joey's SA-CSS of 3 and RRB-CSS of 7 suggested milder severity of social-communication symptoms compared to repetitive behaviors. His overall CSS followed a similar trajectory to his SA-CSS, showing a steady decrease in severity across early childhood, and did not reflect the apparent increase in repetitive behaviors during this same period.

Case 3. "Carolyn," a Caucasian female, was first seen as part of a clinical research project just after her second birthday. At this time, she received a diagnosis of PDD-NOS and her SA-CSS of 4 suggested milder severity of social-communication impairments during the ADOS (e.g., strengths in shared enjoyment and facial expressions, but difficulties using coordinated eye gaze) compared to her RRB-CSS of 9 (reflecting hand and finger, as well as whole-body mannerisms, a preoccupation with cars and brief peering at objects). However, over the next eight years, there was a steady

increase in deficits in SA, resulting in an SA-CSS of 10 by the time she was 10 years old; while she continued to express some shared enjoyment with the examiner, her use of facial expressions was more limited and deficits in eye contact persisted. Her overall CSS also follows this pattern. In contrast, during the period in which she had the most dramatic increases in SA-CSS, the severity of Carolyn's repetitive behaviors remained relatively stable. Over time, she continued to exhibit hand and finger and whole-body mannerisms (e.g., twirling and jumping), and brief visual sensory interests. She also demonstrated unusual preoccupations (e.g., with time), as well as ritualistic behaviors, such as placing objects in toy trucks in a particular way.

Case 4. "Matthew," an African American male, was seen at age 4 years as part of a clinical research study, at which time he received a diagnosis of autism. During his first ADOS, Matthew exhibited more severe social-communication symptoms (SA-CSS=8) than repetitive behaviors (RRB-CSS=5). Separate examination of his SA-CSS and RRB-CSS suggest relatively stable severity in both domains across early childhood, marked by persistent difficulties in nonverbal social communication (e.g., facial expressions and eye contact), initiation of overtures, brief sensory interests and possible hand and finger mannerisms. At 11 years of age, Matthew showed an apparent decrease in severity of social-communication symptoms (a greater range of facial expressions and more reciprocal social communication) and a worsening of repetitive behaviors, including clear hand and finger mannerisms, excessive references to Batman and wrestling, repetitive stereotyped questions, and listing of his classmates when asked the names of his friends. In his case, the overall CSS showed a gradual worsening of symptom severity between

ages 4 and 11, failing to account for the possible divergence of trajectories in social-communication skills and repetitive behaviors in later childhood.

Discussion

ADOS calibrated domain totals achieved the goal of significantly reducing associations with child characteristics compared to raw SA and RRB totals. For SA-Raw domain scores, 45% of variance was explained by child characteristics not specific to ASD, with verbal IQ and maternal education emerging as significant predictors. For the SA-CSS, verbal IQ remained the only significant predictor, accounting for just under 11% of variance in the calibrated SA score. Similarly, approximately 12% of variance in RRB-Raw Total was explained by verbal IQ, with nonverbal IQ and race collectively accounting for an additional 3%. For the RRB-CSS, nonverbal IQ, verbal IQ and race remained significant predictors, but explained less than 5% of variance. Thus, though the effects of child characteristics were not completely eliminated, the calibrated domain scores provided a measure of ASD severity that was significantly less influenced by child characteristics, particularly verbal IQ, than were raw totals.

It is interesting to note that associations between IQ and RRB Raw were much smaller compared to the relationship between IQ and SA-Raw. A similar difference in associations with developmental level was noted for Social+Nonverbal Communication vs. Repetitive Behavior raw domain totals on the Autism Diagnostic Interview-Revised (Hus & Lord, 2013*)*. The restricted range of RRB-Raw scores may explain the weaker associations. Nevertheless, in spite of relatively smaller influences of developmental level on RRB-Raw, it is important to calibrate RRB-CSS in order to provide a comparable

severity metric for both ADOS domains. Most important, the RRB-CSS reduced the influence of developmental level on RRB totals even further.

It is also noteworthy that there was marked overlap in the distributions of domain calibrated scores across diagnostic groups. On one hand, the overlap of the Non-ASD group with the Autism and Other-ASD groups may reflect recruitment biases in our Non-ASD sample, some of whom were referred for assessment of ASD, but who received a clinical Non-ASD diagnosis. On the other hand, the overlap between the Autism and Other-ASD group could reflect that the calibrated scores are capturing the heterogeneity of symptom severity that characterizes ASDs. Moreover, the overlap with the Non-ASD group highlights that some social-communication and repetitive behaviors captured on the ADOS are not specific to ASD.

The newly standardized SA-CSS and RRB-CSS provide useful measures of autism symptom severity which are consistent with the two symptom domains defining ASD proposed for DSM-5. As we move toward a single classification of "autism spectrum disorder" in DSM-5, calibrated domain scores have the potential to play a role in the clinical specification of ASD severity. When DSM-5 criteria are finalized, assessing the degree to which the 10-point CSS scale indicating severity of ASD symptomatology relates to different DSM-5 levels of severity for each behavioral domain (currently proposed as "requiring support," "requiring substantial support," and "requiring very substantial support") will be an important step. If the scores can be mapped on to clinical levels of severity, they may be useful to inform the level of impairment in each behavioral domain; however, these scores will not be sufficient to make such clinical determinations, as they provide information about behaviors in a

limited context. Information collected from other modalities of assessment, such as caregiver interview or observation in other settings, will be needed to inform the appropriate level of severity to describe the level of support an individual requires.

It is also hoped that the calibration of severity metrics for social-communication deficits (SA-CSS) and repetitive behaviors (RRB-CSS) will bring us a step closer to parsing apart the phenotypic heterogeneity in ASD. Current studies frequently rely on totals from diagnostic instruments such as the SRS or ADI-R as estimates of ASD severity. Yet these totals are known to be greatly influenced by child characteristics such as age, language level, and non-ASD-specific behavioral problems (e.g., Constantino et al., 2003; Hus et al., 2013; Hus & Lord, 2013). Although the original ADOS calibrated severity metric was derived to reduce the effects of non-specific child characteristics (Gotham et al., 2009), it yields an estimate of overall severity that does not allow for separate examination of the variation in behavioral domains underlying these scores. In comparison, the SA-CSS and RRB-CSS provide more behavioral specificity than each of these general measures. Because potential biomarkers are frequently postulated to be related to specific domains of behavior (e.g., severity of RRBs), separate calibrated domain scores offer an important advance. Additionally, use of these calibrated domain scores in place of raw totals increases the likelihood that associations with genetic or neurobiological abnormalities are specific to ASD symptoms rather than associated with general developmental factors, such as age, IQ or language level.

Using these scores to separately examine distinct trajectories of social-communication and repetitive behaviors may also provide a more sensitive measure of intervention response over longer periods of time, enabling change in one domain to be

detected, even when behaviors in the second domain persist. Although children may become more familiar with particular tasks (e.g., participating in the birthday party routine) if they are administered the ADOS several times within a short period, because scores are based on spontaneous initiations and responses, rather than performance on tasks, scores and ADOS classifications do not demonstrate practice effects (Lord et al., 2012). Thus, the SA-CSS and RRB-CSS may provide a way to measure more global changes in behaviors in response to intervention, rather than improvements in very specific skills. Furthermore, different SA-CSS and RRB-CSS trajectory profiles may provide an additional method of stratification to increase phenotypic homogeneity in samples, which can be used to gain insight into biological mechanisms underlying specific developmental patterns.

Limitations

Domain calibrations were based on the large "convenience" sample that was used to create the overall ADOS CSS (Gotham et al., 2009). As these authors acknowledged, this sample is likely to be representative of other samples ascertained through North American clinical research centers over the past two decades. It is hoped that using ADOS classifications of (i.e., "Autism," "Autism Spectrum" and NonSpectrum"), rather than clinical best estimate diagnoses, to anchor overall severity scores and set thresholds for sensitivity and specificity of domain calibrated scores would circumvent, to some extent, recruitment effects in this sample (Gotham et al., 2009). However, it is possible that calibration using ADOSes from population studies or more recently ascertained samples may result in different mappings of raw totals and calibrated scores. Additionally, samples recruited outside of North America, or from other clinical

populations, may show a somewhat different distribution of scores. Here, the effects of maternal education and race observed on both overall raw totals and calibrated scores are likely to be an artifact of recruitment biases (Gotham et al., 2009), though the significance of these predictors may also have been influenced by the large sample size. Replication of the domain calibrations in independent samples is an important next step.

Given the restricted range of raw RRB totals, the RRB-CSS is not a full 10-point severity metric. Nonetheless, scores were mapped onto the 10-point scale to avoid confusion when using the calibrated domain scores together. That is, there was concern that a reduced RRB-CSS scale (e.g., of 1-6) may result in confusion when interpreting the meaning of an RRB-CSS score in comparison to a score on the overall-CSS scale (i.e., an assumption that an RRB-CSS of 6 would be equal to an overall-CSS of 6, when it actually would be more meaningful to interpret as similar to an overall-CSS of 10). The method of using the overall-CSS percentiles to inform mapping of domain raw scores to the 10-point calibrated scale allows comparability across the three scales, such that a given value on the overall-CSS, SA-CSS, and RRB-CSS corresponds to approximately the same percentile of raw score (for a child of that language level and age) for each. Such comparability also increases the clinical utility of this metric; for example, a child who has a high overall-CSS comprised of an SA-CSS of ’10 and an RRB-CSS of ‘6’ may need a different treatment approach than another child with the same overall-CSS reflecting an RRB-CSS of ‘10’ and an SA-CSS of ‘6’. When using scores to monitor change over time or in response to intervention, researchers and clinicians must bear in mind that there are not RRB-CSS values of 2, 3 or 4. Thus, changing from a score from RRB-CSS of 1, indicating that no repetitive behaviors were observed during the ADOS,

to 5 (reflecting mild severity), is not the same as a change in severity from an RRB score of 6 to 10. This distribution of scores reflects that, given the limited timeframe of the ADOS, the presence of repetitive behaviors is likely to be more meaningful than the absence of such. In order to ensure that a change is CSS for either domain is meaningful, the lower (or higher) score should be observed across several time points. In contrast, a significant increase or decrease during one particular session may suggest that other factors were influencing the child's behavior on that particular day.

Conclusions

ADOS domain calibrations provide separate estimates of severity of ASD-related social-communication deficits and repetitive behaviors that are relatively independent of child characteristics, such as age and language skills, compared to their respective raw totals. This improves their utility as continuous measures of ASD symptom severity that can be used to increase homogeneity of samples and identify links between specific behavioral domains and biological mechanisms, as well as to examine different trajectories of ASD symptoms over time.

Table 3.1 Sample descriptives

	Module 1, No Words			Module 2, Some Words			Module 2, Younger than 5			Module 2, 5 or Older			Module 3		
	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD	N	Mean	SD
ASD															
Age	551	4.22	2.21	395	4.41	1.99	197	3.78	.78	215	7.82	2.54	428	8.54	2.54
VIQ	522	26.84	14.71	361	52.63	21.80	164	80.80	20.93	199	55.14	19.49	386	95.53	22.97
NVIQ	515	53.16	21.40	358	69.74	21.67	161	92.53	22.82	201	76.60	23.43	383	96.22	22.32
VMA	528	.97	1.49	355	2.32	3.01	163	4.45	6.05	202	4.63	4.60	377	8.46	5.36
NVMA	516	1.98	.83	359	3.05	2.40	158	3.69	1.35	190	5.74	2.25	357	8.23	2.88
SA Raw	551	16.79	2.95	395	13.23	4.44	197	10.44	4.30	215	13.20	4.29	428	9.26	4.37
RRB Raw	551	4.67	2.11	395	4.07	2.07	197	3.90	2.11	215	4.68	2.10	428	2.71	1.87
Non-ASD															
Age	60	3.30	1.61	107	3.51	1.60	57	3.67	.62	44	8.00	2.55	141	8.95	2.47
VIQ	57	40.96	18.72	90	68.08	23.74	51	85.33	21.83	44	58.09	19.06	135	91.70	22.29
NVIQ	55	58.80	28.73	89	70.52	23.75	49	92.04	20.46	44	61.93	24.13	136	89.85	22.23
VMA	57	1.15	.47	87	2.30	.69	50	4.72	6.29	43	4.18	1.16	134	8.60	5.13
NVMA	55	1.72	.72	86	2.44	.71	46	3.47	.83	44	4.72	1.39	132	7.92	2.71
SA Raw	60	8.37	5.83	107	4.71	3.91	57	3.56	2.77	44	4.16	3.14	141	3.90	2.95
RRB Raw	60	1.88	1.88	107	1.40	1.49	57	1.49	1.43	44	1.64	1.64	141	.99	1.15

ASD=autism spectrum disorder (Autistic Disorder, Aspergers, PDD-NOS); VIQ=verbal IQ; NVIQ=nonverbal IQ; VMA=nonverbal mental age; NVMA=nonverbal mental age; SA Raw=Social Affect raw total; RRB Raw=Restricted, Repetitive Behaviors raw total; Non-ASD=non autism spectrum disorder diagnosis

Table 3.2 Mapping of ADOS raw domain totals onto calibrated severity scores

Domain	Calibrated Severity Score	Raw Domain Totals											
		Module 1; No Words			Module 1; Some Words			Module 2				Module 3	
		2 yrs	3 yrs	4-14 yrs	2-3 yrs	4 yrs	5-14 yrs	2-3 yrs	4 yrs	5-6 yrs	7-16 yrs	3-5 yrs	6-16 yrs
Social Affect Domain	1	0-3	0-3	0-2	0-1	0-1	0-1	0-1	0-1	0-1	0-1	0-2	0-1
	2	4-5	4-5	3-5	2-4	2-3	2-3	2-3	2	2-3	2	3	2
	3	6-8	6-9	6-9	5	4-5	4-5	4	3-4	4	3	4	3-4
	4	9	10	10	6-7	6-7	6-7	5	5-6	5	4-5	5	5
	5	10-13	11-12	11-12	8	8-9	8-9	6	7	6	6	6	6
	6	14-16	13-16	13-14	9-11	10-12	10-13	7-8	8-9	7-9	7-10	7-8	7
	7	17	17	15-16	12-13	13	14-15	9-10	10-11	10-11	11-13	9-10	8-9
	8	18	18	17-18	14-15	14-15	16	11	12-13	12-15	14-15	11-12	10-11
	9	19	19	19	16-17	16-17	17-18	12-14	14-15	16	16-17	13-14	12-14
	10	20	20	20	18-20	18-20	19-20	15-20	16-20	17-20	18-20	15-20	15-20
Restricted & Repetitive Behaviors Domain	1	0	0	0	0	0	0	0	0	0	0	0	0
	2												
	3												
	4												
	5	1	1	1-2	1	1	1	1	1	1	1	1	1
	6	2	2-3	3	2	2	2-3	2	2-3	2-3	2-3		
	7	3	4	4	3	3-4	4	3	4	4	4	2	2
	8	4	5	5-6	4	5	5	4	5	5	5	3	3
	9	5	6	7	5	6	6	5-6	6	6	6	4	4-5
	10	6-8	7-8	8	6-8	7-8	7-8	7-8	7-8	7-8	7-8	5-8	6-8

Table 3.3 Domain raw totals and calibrated severity score means and standard deviations by age/language cell (ASD assessments only)

Module	Age	N	SA-Raw		SA-CSS		RRB-Raw		RRB-CSS	
			Mean	SD	Mean	SD	Mean	SD	Mean	SD
Module 1 No Words	2y	203	16.38	3.85	7.36	2.11	3.75	2.01	7.49	2.21
	3y	141	16.88	2.88	7.45	1.75	4.76	2.00	7.77	1.75
	4-14y	216	16.76	2.59	7.75	1.46	5.36	2.03	7.82	1.65
Module 2 Some Words	2-3y	214	12.10	4.73	6.81	2.33	3.66	1.94	7.44	2.10
	4y	82	13.01	4.75	7.16	2.39	4.12	2.34	7.30	2.25
	5-14y	108	14.85	3.65	7.57	1.78	4.67	2.01	7.56	2.05
Module 2	2-3y	106	10.03	4.02	7.08	2.18	4.02	2.02	7.59	1.94
	4y	94	10.69	4.69	6.88	2.37	3.74	2.18	6.87	2.22
	5-6y	103	12.25	4.62	7.49	2.05	4.59	2.09	7.59	1.93
	7-16y	112	14.07	3.79	7.99	1.59	4.77	2.11	7.67	2.02
Module 3	3-5y	71	9.52	4.06	6.68	2.48	2.65	1.83	6.94	2.43
	6-16y	357	9.21	4.43	6.77	2.52	2.73	1.88	6.86	2.68
All Modules, All Ages		1807	12.98	4.99	7.21	2.17	3.96	2.18	7.39	2.19

Table 3.4 Forward stepwise linear regression models for domain raw totals and calibrated domain scores

	SA - Raw							SA - CSS					
	R^2	ΔF	*df*	*B*	SE *B*	β		R^2	ΔF	*df*	*B*	SE *B*	β
Step 1	.430	1079.07	1430				Step 1	.105	167.89	1430			
Constant				18.75	.20		Constant				8.45	.11	
Verbal IQ				-.10	.00	-.66	Verbal IQ				-.02	.00	-.32
Step 2	.433	7.41	1429				Step 2	.109	5.79	1429			
Constant				18.19	.29		Constant				8.18	.16	
Verbal IQ				-.11	.00	-.72	Verbal IQ				-.03	.00	-.40
Nonverbal IQ				.02	.01	.08	Nonverbal IQ				.01	.00	.09
Step 3	.435	5.51	1428				Step 3						
Constant				18.15	.29		Constant						
Verbal IQ				-.11	.00	-.73	Verbal IQ						
Nonverbal IQ				.01	.01	.08	Nonverbal IQ						
Mat Education				.56	.24	.05	Mat Education						

	RRB - Raw							RRB-CSS					
	R^2	ΔF	*df*	*B*	SE *B*	β		R^2	ΔF	*df*	*B*	SE *B*	β
Step 1	.117	208.86	1573				Step 1	.035	56.62	1573			
Constant				5.30	.10		Constant				8.49	.15	
Verbal IQ				-.02	.00	-.34	Nonverbal IQ				-.02	.00	-.19
Step 2	.131	25.64	1572				Step 2	.041	10.12	1572			
Constant				5.83	.15		Constant				8.68	.16	
Verbal IQ				-.01	.00	-.20	Nonverbal IQ				-.02	.00	-.21
Nonverbal IQ				-.01	.00	-.18	Race				-.50	.16	-.08
Step 3	.143	20.40	1571				Step 3	.045	7.47	1571			
Constant				6.07	.15		Constant				8.64	.16	
Verbal IQ				-.01	.00	-.23	Nonverbal IQ				-.01	.00	-.13
Nonverbal IQ				-.02	.00	-.19	Race				-.56	.16	-.09
Race				-.67	.15	-.11	Verbal IQ				-.01	.00	-.11

Table 3.5 Case summary characteristics

	First Assessment				Last Assessment			
	Age	VIQ	NVIQ	ADOS Module	Age	VIQ	NVIQ	ADOS Module
Bianca[a]	4.0	108	80	2	11.0	126	107	3
Joey	2.8	69	74	2	5.1	105	119	3
Carolyn	2.3	33	72	1	10.2	42	51	2
Matthew	4.0	31	63	1	11.0	58	88	3

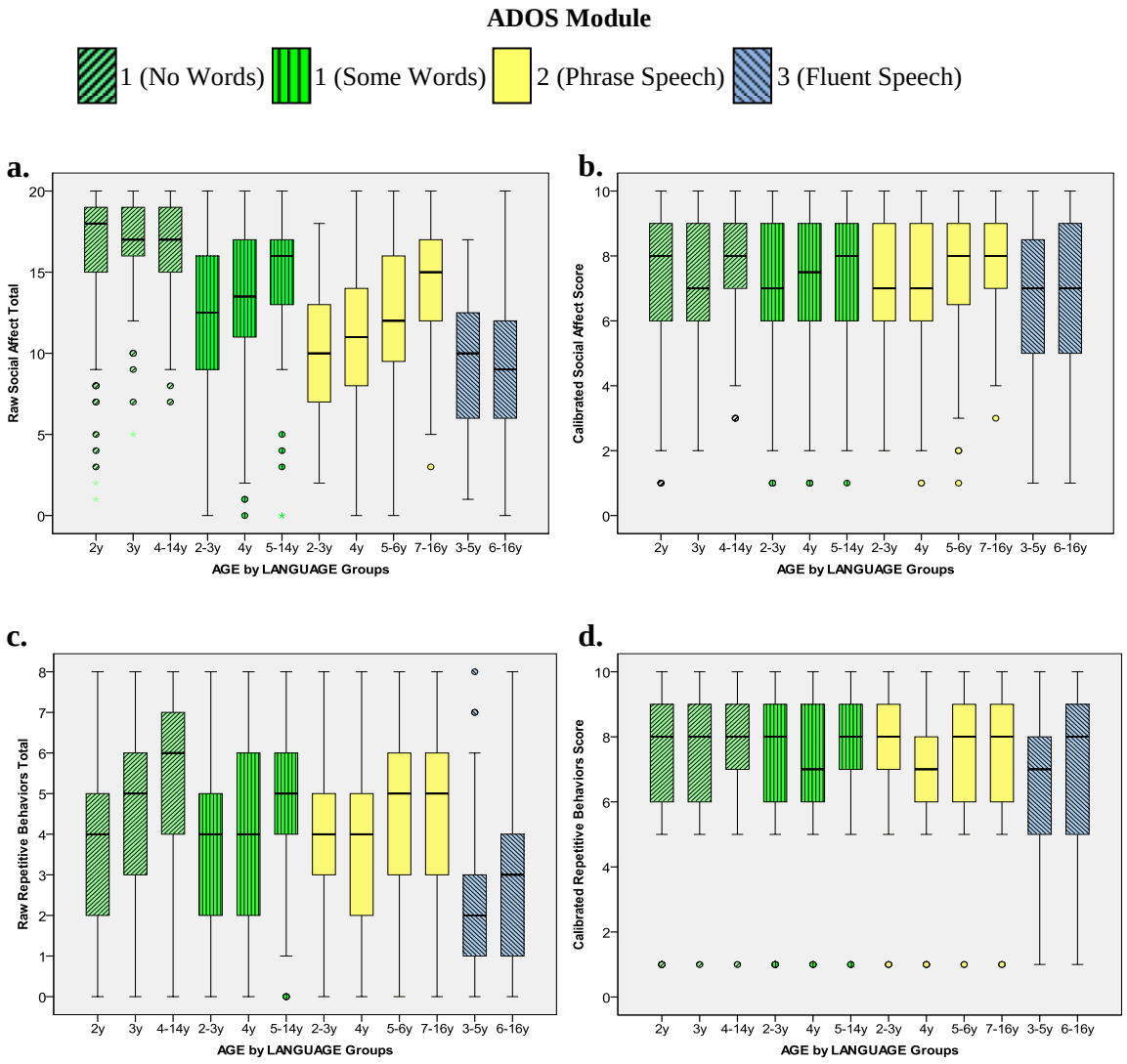
All ages in years; VIQ=verbal IQ; NVIQ=nonverbal IQ [a]Cognitive assessment was not completed at last assessment; IQs are from previous assessment at age 10

Figure 3.1 Age by language level calibration cells.

Age (in years)		2	3	4	5	6	7	8	9	10+
Mod 1	No Words	203	141	216						
	Single Words	214		82	108					
Mod2	Phrases	106		94	103		112			
Mod 3	Fluent	71				357				

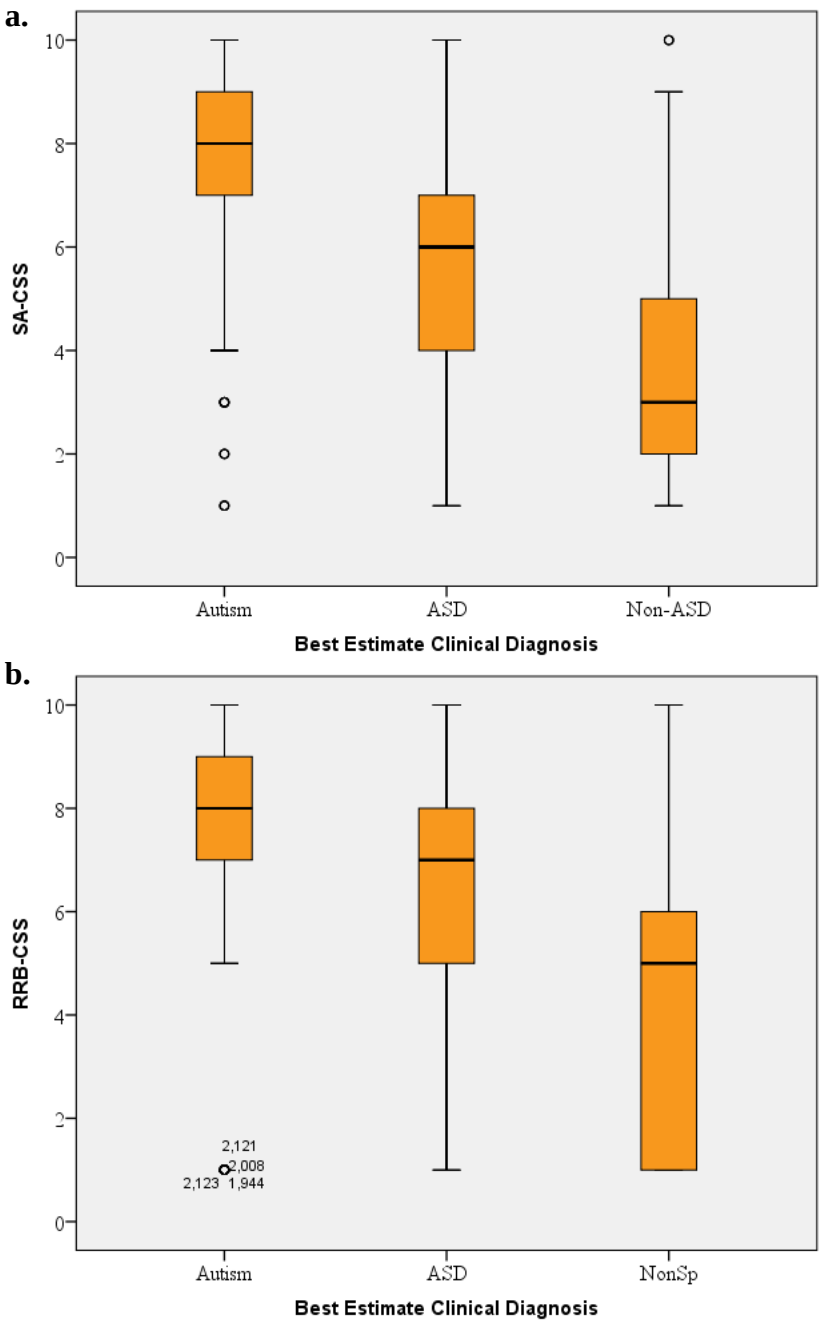
Numbers indicate number of participants per cell

Figure 3.2 Distributions of domain totals and calibrated severity scores



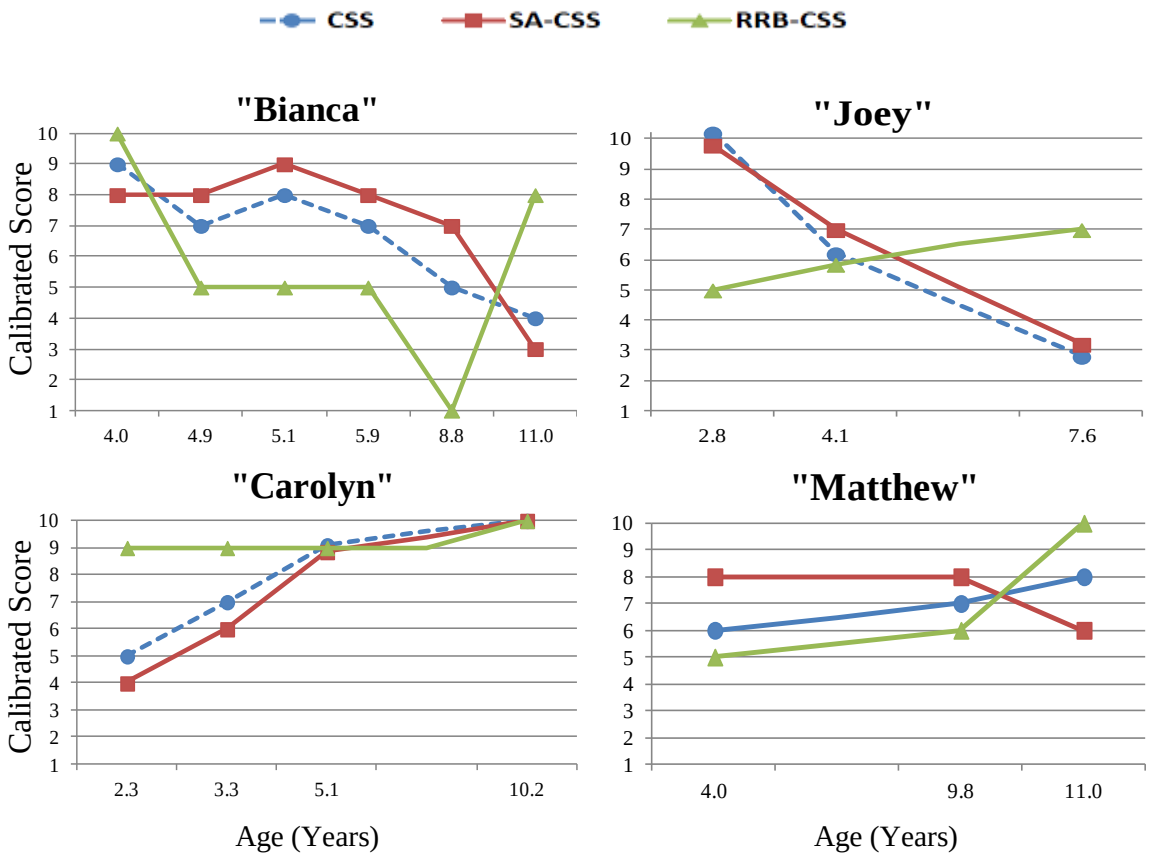
a (top, left) Distributions of raw Social Affect domain totals by age/language cells. **b** (top, right) Distributions of calibrated Social Affect domain scores by age/language cells. **c** (bottom, left) Distributions of raw Restricted and Repetitive Behavior domain totals by age/language cells. **d** (bottom, right) Distributions of calibrated Restricted and Repetitive Behavior domain scores by age/language cells

Figure 3.3 Distributions of domain calibrated severity scores by best estimate clinical diagnosis



a (top) Distributions of calibrated Social Affect domain scores by best estimate clinical diagnosis. **b** (bottom) Distributions of calibrated Restricted and Repetitive Behavior domain scores by best estimate clinical diagnosis

Figure 3.4 Case summaries of longitudinal domain severity scores



References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders. (4th ed., Text rev.)* Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders. (5th ed.)* Washington, DC: American Psychiatric Association.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Psychology Press.

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

Constantino, J. N., Hudziak, J. J., & Todd, R. D. (2003). Deficits in reciprocal social behavior in male twins: evidence for a genetically independent domain of psychopathology. *Journal of the American Academy of Child and Adolescent Psychiatry*, 42(4), 458–467.

Di Martino, A., Kelly, C., Grzadzinski, R., Zuo, X.-N., Mennes, M., Mairena, M. A., Lord, C., et al. (2011). Aberrant striatal functional connectivity in children with autism. *Biological Psychiatry*, 69(9), 847–856.

Dichter, G. S., Richey, J. A., Rittenberg, A. M., Sabatino, A., & Bodfish, J. W. (2011). Reward circuitry function in autism during face anticipation and outcomes. *Journal of Autism and Developmental Disorders*.

DiLavore, P., Lord, C., & Rutter, M. (1995). The Pre-Linguistic Autism Diagnostic Observation Schedule. *Journal of Autism and Developmental Disorders*, 25(4), 355–379.

Elliott, C. D. (1990). *Differential Ability Scales*. San Antonio, TX: Psychological Corporation.

Frazier, T. W., Youngstrom, E. A., Speer, L., Embacher, R., Law, P., Constantino, J., Findling, R. L., et al. (2012). Validation of proposed DSM-5 criteria for Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(1), 28–40.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 39(5), 693–705.

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

Huerta, M., Bishop, S. L., Duncan, A., Hus, V., & Lord, C. (2012). Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. *American Journal of Psychiatry, 169*(10), 1056–1064.

Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry, 54*(2), 216–224.

Hus, V., & Lord, C. (2013). Effects of child characteristics on the Autism Diagnostic Interview-Revised: Implications for use of scores as a measure of ASD severity. *Journal of Autism and Developmental Disorders, 43*(2), 371–381.

Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D. M., … Risi, S. (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. Archives of General Psychiatry, 69(3), 306–313.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2).* Torrance, CA: Western Psychological Services.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., Bishop, S., 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(9), 1305–1320.

Mandy, W. P. L., Charman, T., & Skuse, D. H. (2012). Testing the construct validity of proposed criteria for DSM-5 Autism Spectrum Disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(1), 41–50.

Mattila, M.-L., Kielinen, M., Linna, S.-L., Jussila, K., Ebeling, H., Bloigu, R., Joseph, R. M., et al. (2011). Autism Spectrum Disorders according to "DSM-IV-TR" and comparison with "DSM-5" draft criteria: An epidemiological study. *Journal of the American Academy of Child & Adolescent Psychiatry, 50*(6), 583–592.e11.

McCarthy, P. L., Cicchetti, D. V., Sznajderman, S. D., Forsyth, B. C., Baron, M. A., Fink, H. D., Czarkowski, N., et al. (1991). Demographic, clinical, and psychosocial predictors of the reliability of mothers' clinical judgments. *Pediatrics*, 88(5), 1041–1046.

McPartland, J. C., Reichow, B., & Volkmar, F. R. (2012). Sensitivity and specificity of proposed DSM-5 diagnostic criteria for Autism Spectrum Disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(4), 368–383.

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

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

State, M. W., & Levitt, P. (2011). The conundrums of understanding genetic risks for autism spectrum disorders. *Nature Neuroscience*, 14(12), 1499–1506.

World Health Organization. (2012). *The ICD-10 classification of mental and behavioral disorders: Clinical descriptions and diagnostic guidelines.* Geneva, Switzerland: World Health Organization.

World Health Organization. (2012). International Classification of Diseases 11th Revision: Beta draft. http://apps.who.int/classifications/icd11/browse/f/en. Accessed 22 September 2012.

CHAPTER IV

The Autism Diagnostic Observation Schedule, Module 4: Revised Algorithm and Standardized Severity Scores

Defined by impairments in social-communication and the presence of restricted, repetitive, stereotyped behaviors and interests (American Psychiatric Association [APA] 2013), autism has historically been considered a childhood disorder. However, studies of young adults with autism spectrum disorders (ASD) report variable outcomes (see Levy & Perry 2011). Only 3 to 25% of individuals with ASD reportedly achieve "optimal outcomes" (generally referring to cognitive and adaptive abilities falling within the "average" range and the ability of these individuals to function independently in the community; Helt et al. 2008; Levy & Perry 2011). Advances in early detection and intervention may contribute to higher rates of optimal outcomes in the future. Nonetheless, at present, the majority of individuals diagnosed with ASD require varying levels of life-long supports.

In the fall of 2010, 369,774 American children ages 6 through 21 received services under the special education classification of "Autism" (U.S. Department of Education, Office of Special Education Programs, Data Analysis System 2011). This provides a conservative estimate of the number of children who will be transitioning to adulthood over the next decade, as not all children diagnosed with ASD receive special education services. As a further reminder of this growing public health issue, the per capita lifetime incremental cost of autism is estimated at $3.2 million. Twenty-one

percent is attributed to care for the adult with ASD and 30.7% to loss of the individual with ASD's productivity during adult life (Ganz 2007).

These numbers underscore the pressing need for research to better understand the strengths and difficulties of adults with ASD, as well as factors that promote more positive outcomes (Interagency Autism Coordinating Committee 2011). Such information is critical to both develop and expand the available services and supports for adults, as well as to inform earlier intervention efforts and preparatory activities for the transition to adulthood that will promote positive long-term outcomes. With the help of longitudinal studies, we may begin to investigate trajectories in ASD that will inform prognosis when a child is diagnosed. Furthermore, a better understanding of the ASD phenotype is necessary for investigations seeking to link behavioral symptoms to differences in brain structure and function, which may contribute to the development of targeted interventions.

Recent studies have begun to explore predictors of outcome and the current needs of adolescents and adults with ASD (e.g., Farley et al. 2009; Howlin, Moss, Savage & Rutter 2013; see Henninger & Lounds-Taylor 2012 for review). Examination of development has been mostly limited to measures of cognitive and adaptive behavior, global ratings of outcome derived by authors and change in diagnostic classification on measures such as the *Autism Diagnostic Observation Schedule* (*ADOS*; Lord, Rutter, DiLavore, & Risi 1999) and *Autism Diagnostic Interview-Revised* (*ADI-R;* Rutter, LeCouteur, & Lord 2003). Although such analyses are informative, in order to fully understand the life course of the disorder, examination of trajectories in ASD symptom severity and how early ASD symptom severity predicts longer-term outcomes is needed.

Several studies have used raw totals from measures such as the *ADI-R* or the *Childhood Autism Rating Scale* (*CARS;* Schopler, Reichler, DeVellis, & Daly, 1980) to predict adult outcomes or investigate change in ASD symptoms (e.g., Anderson, Liang & Lord, 2013; Eaves & Ho, 2008; Fein et al., 2013; Gillespie-Lynch et al. 2012; Howlin et al. 2013; Sigman & McGovern 2005; Piven, Harper, Palmer & Arndt 1996; Shattuck et al. 2007; see Levy & Perry 2011 for review)). However, scores from these measures may be confounded by strong associations with developmental level (e.g., Hus & Lord 2012; Perry, Condillac & Freeman, 2002) and reporting biases (Hus, Taylor & Lord 2007; Jones et al. 2013). Increasing the availability of instruments based on standardized protocols of observation that are less influenced by these factors, and that can be used to explore trajectories in ASD symptoms, will allow for a more thorough investigation of factors that predict adult outcomes.

Many longitudinal studies of ASD have included the *ADOS*, making it possible to use *ADOS* scores to examine developmental trajectories of ASD symptoms. Diagnostic algorithms for *ADOS* Modules 1-3, used to assess children and adolescents of varying language levels, were recently revised (*ADOS-2*; Lord, Rutter, DiLavore, Risi, Gotham, & Bishop 2012). Changes afforded improved diagnostic validity and increased item overlap across modules, thereby facilitating comparisons of scores across childhood and early adolescence. Furthermore, the revised algorithms were divided into two domains, consistent with the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; APA 2013) ASD diagnostic criteria: *Social Affect* (which combines social and communication behaviors) and *Restricted and Repetitive Behaviors*. There is a single diagnostic cut-off for the combined domain total. Algorithms for the recently

published *ADOS-Toddler* module (Lord, Luyster, Gotham & Guthrie 2012) follow a similar structure to that of Modules 1-3, making it possible to examine symptom trajectories using the *ADOS-2* from a very young age.

Although the algorithms for Modules 1-3 were revised to be more comparable across modules, raw totals were significantly influenced by age and language level. Higher scores were associated with less language (i.e., Module 1>Module2>Module3). In addition, within Modules 1 and 2, older children tended to have higher scores than younger children (Gotham et al. 2007; Gotham, Pickles & Lord 2009). Intelligence, language and age are certainly important factors to consider when describing an individual's level of overall functional impairment. However, scores confounded by these issues make it difficult to examine the relative severity of autism-specific symptoms and cloud the interpretation of findings as specific to ASD. For example, an association between *ADOS-2* raw totals and a specific chromosomal abnormality may be misinterpreted as evidence of a causal mechanism for ASD when in fact it is a marker for more general cognitive impairment. To address this issue, Gotham and colleagues (2009) used a large sample of children with ASD to create the *ADOS-2* Calibrated Severity Metric (CSS), a standardized version of *ADOS-2* scores that is less strongly associated with age and language compared to raw *ADOS-2* totals. Importantly, the CSS is intended to be used as a marker of ASD symptom severity relative to age and language level (not as an indicator of functional impairment). The CSS allow *ADOS-2* scores to be used to investigate relationships between behavioral symptoms and underlying biological mechanisms and study stability and change in symptom severity over time and across individuals of different developmental levels. Clinicians may also use the CSS to

describe severity of a client's core symptoms. However, for an adequate description of functional impairment, one needs to move beyond the CSS to include information regarding language, adaptive functioning and other behavioral features.

The diagnostic algorithm for *ADOS* Module 4, designed for assessment of verbally fluent adults, has not yet been revised. Recent studies have supported the validity of the original Module 4 algorithm (e.g., Brugha et al. 2011). However, the lack of comparability with algorithms for the *ADOS-Toddler* Module and Modules 1-3, along with the absence of calibrated severity scores for Module 4, hinders comparisons of adult assessments to those conducted in childhood and early adolescence. Given that Module 4 is used with verbally fluent adults who are likely to demonstrate a wide range of abilities, revisions are particularly important to investigations seeking to understand how ASD severity interacts with other factors, such as verbal, cognitive, and adaptive ability, to predict functional outcomes. Thus, the purpose of this study is to revise the *ADOS* Module 4 algorithm to be more comparable to currently used algorithms for *ADOS-2* Modules 1-3 and to provide a calibrated score that can be used to quantify and compare the severity of core symptoms in adults with ASD.

Methods

Participants

The sample included data from 393 different *individual participants.* Some participants had repeated assessments, yielding a total of 437 *cases*. Each case was defined by an *ADOS* and best estimate clinical diagnosis; 29 participants provided data for multiple cases (M=2.52, SD=1.06, range=2-6) based on evaluations conducted at different points in time. Data were ascertained from three sources:

1) The majority of participants were obtained from the Center for Autism and the Developing Brain (CADB) Data Bank, which included 229 participants with ASD and 85 with non-ASD diagnoses. The CADB Data Bank includes consecutive referrals to specialty diagnostic clinics in White Plains, New York; Ann Arbor, Michigan; and Chicago, Illinois, and participants in research studies conducted through the Weill Cornell Medical College, University of Michigan, University of Chicago, and University of North Carolina - Chapel Hill.
2) Seventy-four participants with ASD were obtained from the Simons Simplex Collection (SSC) data repository. Families in the SSC were recruited to 12 university-based sites for participation in a genetic study of families with one child with ASD and no first-, second- or third-degree relatives with ASD or related conditions (see Chapter II Appendix A.2 for further description).
3) Forty-nine participants were recruited at the University of Michigan and CADB through the *Development and refinement of diagnostic measures for adults with ASD* project, a graduate fellowship funded by Autism Speaks. Adolescents and adults between 15-30 years of age with a previous or suspected diagnosis of an ASD were eligible for participation. Based on assessment, forty-four individuals received best estimate clinical diagnoses of ASD and five received non-ASD diagnoses. Additional eligibility requirements included: English as the primary language spoken in the home, no significant sensory impairments (e.g., deafness, blindness) that may interfere with completion or interpretation of standardized testing, and no

active psychosis or uncontrolled seizures. Participants were recruited from clinical referrals for diagnostic services, schools, and community organizations that provide services for adults with developmental disabilities.

Approximately 80% of the overall sample was male and 83% Caucasian. Inclusion/Exclusion criteria varied by research study. However, individuals with significant hearing, vision or motor problems that interfered with standardized testing or who were exhibiting active psychosis or uncontrolled seizures at the time of assessment were excluded from each study. Participants in the SSC were also required to meet Collaborative Programs for Excellence in Autism criteria for ASD (see Chapter II Appendix A.1) and were excluded if the individual had a diagnosis of Fragile X syndrome, tuberous sclerosis, Down syndrome or significant early medical history. Additionally, SSC participants could not have any first, second or third degree relatives with ASD or a sibling with substantial language or psychological problems related to ASD. Ages ranged from 9.92 to 62.25 years at the time of assessment (mean=21.56, standard deviation=8.62 years).

Of the 437 *cases*, 177 had clinical diagnoses of autism (40% of entire sample), 170 Other-ASD (i.e., PDD-NOS or Asperger's; 39%), and 90 Non-ASD diagnoses (21%). The Non-ASD sample was comprised of both clinical referrals and individuals recruited to research studies as controls. In addition to having first ruled-out an ASD diagnosis, 84% of non-ASD participants received a primary diagnosis of a non-ASD DSM-IV-TR disorder; 30% had a primary diagnosis of mood and/or anxiety disorders, 26% had non-specific intellectual disability, 14% had externalizing behavioral disorders (e.g., ADHD/ODD), 5% had Down syndrome or Fragile X, 4% had language disorders,

1% had Fetal Alcohol syndrome, 1% had Cerebral Palsy and 3% of cases had unspecified difficulties. The remaining 16% of Non-ASD sample did not meet criteria for a DSM-IV-TR diagnosis at the time of assessment; 64% of these individuals (n=9) had had a previous diagnosis of ASD and 36% (n=5) had had a previous Non-ASD diagnosis. There was no significant difference in *ADOS* totals between the 9 individuals with previous ASD diagnoses and the remaining non-ASD group (data available from authors upon request). Table 4.1 provides a more detailed sample description.

Measures

Autism Diagnostic Observation Schedule (ADOS).

The *ADOS* (Lord et al., 2000) is a standardized, semi-structured observational assessment used to assess communication, reciprocal social interaction, imagination/creativity and stereotyped behaviors and restricted interests to inform diagnosis of autism spectrum disorders. The *ADOS* is organized into four modules based on the individual's chronological age and expressive language level, ranging from preverbal to verbally fluent. Module 4 was designed for use with older adolescents and adults with fluent speech. The original Module 4 diagnostic algorithm provides separate cut-off values for the Communication and Social domains, as well as a cut-off for the sum of the two domains to provide instrument classifications of Autistic Disorder or Autism Spectrum Disorder. In the original validation study, internal consistency for all modules ranged from α=.74-.91 for the Communication and Social domains totals and α=.47 to .65 for the Stereotyped Behaviors and Restricted Interests domain totals. Interrater reliability of all Module 4 items exceeded 80% exact agreement, with $\kappa \geq .60$ for most items.

All *ADOSes* were administered and scored by a clinical psychologist or trainee (e.g., graduate student or research assistant) who met standard requirements for research reliability. Inter-rater reliability on the *ADOS* was monitored through joint administration and scoring by two different examiners or scoring of videotapes in 11% of cases. Exact item agreement was initially established at 80% and consistently exceeded 75%. Disagreements were resolved through discussion and consensus codes were used for analyses. Within this sample, 54 different examiners collected the data from the *ADOS* over 18 years (1994-2012).

Cognitive Assessment.

Intelligence quotients (IQ) were derived from a developmental hierarchy of cognitive measures; most frequently the *Wechsler Abbreviated Scale of Intelligence* (Wechsler, 1999) and the D*ifferential Ability Scales* (Elliott 1990; Elliott 2007). The *Peabody Picture Vocabulary Test* (Dunn & Dunn 2007) and *Ravens' Progressive Matrices* (Raven 1960) were also sometimes used to estimate verbal IQ and nonverbal IQ, respectively.

Procedure

The ADOS was conducted as part of a clinical or research evaluation. Most commonly, evaluations began with the collection of a developmental history using the *Autism Diagnostic Interview-Revised* (Rutter, LeCouteur & Lord 2003) or the *Social Communication Questionnaire* (Rutter, Bailey, Lord & Berument 2003). This was then followed by direct assessment of the individual consisting of psychometric testing and the *ADOS* (see Gotham et al. 2009 and Fishbach & Lord 2011 for more detailed procedures). At the clinics (UCDDC, UMACC, CADB), best estimate clinical diagnoses based on

DSM-IV-TR criteria (APA 2000) were made by a supervising clinical psychologist, child psychiatrist and/or advanced graduate student after review of all assessment information. For the SSC, best estimate clinical diagnoses were assigned by an experienced clinician after reviewing all available information and viewing the child in person or on video (see Lord et al. 2011). The ADI-R or SCQ was available for 86% of cases. Verbal and/or nonverbal IQ estimates were available for 361 participants (91%). Clinic-referred participants received oral feedback and a written report of results without financial compensation. Participants recruited solely for research purposes received a written summary of results and financial compensation. Institutional review boards at all sites approved all procedures related to this project.

Design and Analysis

Analyses used to revise the Module 4 algorithm followed a similar procedure to that described for derivation of new algorithms for Modules 1 through 3 (Gotham et al. 2007). Calibration of Module 4 raw totals also followed the procedure described for standardization of the overall (Gotham et al. 2009) and domain (Hus et al. 2013) totals for Modules 1-3. These will be described below.

Analysis of Original Module 4 Algorithm.

Item scores of 3 were collapsed with scores of 2. Domain total distributions were examined for floor or ceiling effects; variables contributing to the effects were identified. Correlations between Module 4 totals and participant characteristics (age and IQ) were examined to inform the need for different algorithms based on age or ability level. Item distributions were examined to select those items that best differentiated ASD vs. Non-ASD diagnoses. Two items, *Shared Enjoyment in Interaction* and *Self Injurious Behavior*

were not included in an early, prepublication version of the *ADOS* and were missing data for some participants (n=56 and n=49, respectively). Because these items were not included in either the original or the revised algorithm, these participants were maintained in the dataset. The items were treated as missing for item-level analyses.

Development and Analysis of New Module 4 Algorithm.

Social-communication items were labeled as "preferred" if no more than 20% of autism cases scored a zero and no more than 20% of Non-ASD cases scored a 2 or a 3. Inclusion criteria were not applied to RRB items. Although the presence of RRBs in a non-ASD participant may be clinically meaningful, the absence of RRBs during the *ADOS* is more difficult to interpret (i.e., while some *ADOS* "presses" may elicit RRBs from some individuals, the time-limited, standardized context of the *ADOS* may limit the number or intensity of RRBs exhibited by others). Exceptions were allowed for four items that were theoretically important and overlapped with items appearing on algorithms for other modules in order to promote conceptual uniformity that would enhance inter-module comparisons (Gotham et al. 2007). These items performed just outside one of the thresholds: *Unusual Eye Contact (15.8% Autism cases scored '0'; 32.2% Non-ASD cases scored a '2'), Emphatic Gestures (25.4%; 11.1%); Communication of Own Affect (33.9%; 8.8%); Amount of Reciprocal Social Communication (36.7%; 1.1%).* One item that met preferred item criteria, *Responsibility,* was excluded due to high rates of non-zero scores (i.e., '1' and '2') in the Non-ASD group, which resulted in reduced specificity when the item was included.

Preferred items were entered into ordinal probit item response models, run with Mplus Version 6.1 software. This method of exploratory factor analysis was chosen to

account for the ordinal nature of *ADOS* data. Factor loadings from promax oblique rotations were used to inform organization of items into domains. Root Mean Square Error Approximation (RMSEA) of 0.08 or less was used to indicate satisfactory fit (Browne & Cudeck 1993). Goodness-of-fit was assessed using confirmatory factor analysis. Comparative Fit Index above 0.9 was used to indicate a good fit (Skrondal & Rabe-Hesketh 2004). Logistic regression was used to examine the contributions of each domain to predict diagnosis.

Distributions of domain totals for the new algorithm were perused for floor and ceiling effects. Item correlations with the remainder of the domain (i.e., minus the item) and with participant characteristics (e.g., age and IQ) were examined. ROC curves were computed and the sensitivity and specificity of the revised algorithm was compared to that of the existing algorithm. Four participants were missing data from items used to compute the new algorithm total. These missing items were imputed using the average of the remaining domain items only for the purpose of computing the diagnostic algorithm.

Development and Analysis of Module 4 Overall Total and Domain Calibrated Severity Scores.

Only scores from individuals with ASD diagnoses were used for raw score standardization. This included all assessments with a corresponding best estimate clinical diagnosis of Autism or Other-ASD. Participants were divided into 1-year chronological age groups to ensure that distributions of the overall and domain scores were comparable and then collapsed into 8 age groups based upon similar distributions (see Figure 4.2). Eventually, all participants aged 9-29 were collapsed into a single group because of similar distributions and minimal correlations between raw scores and age. Participants

aged 40 and above were excluded from standardization due to differences in distribution compared to younger ages and too few older participants to calibrate separately.

As described by Gotham et al. 2009, for the overall total calibration, raw totals corresponding to a Module 4 *ADOS* diagnostic classification of "ASD" were mapped on to CSS ranging from 4 to 10. Totals corresponding to a "Nonspectrum" classification were distributed across CSS of 1-3. Ranges of raw totals assigned to each point on the 10-point severity scale were determined by the percentiles of available data within that classification range. Because there are not separate SA and RRB cut-offs for *ADOS* classifications, the percentiles used for mapping the overall totals (i.e., SA+RRB) were used to inform the mapping of the raw SA and RRB totals to each respective domain CSS. As with Modules 1-3, raw RRB domain totals were mapped on to CSS values of 0 and 5-10, due to the limited range of RRB raw totals (i.e., 0-10; Hus et al. 2013). Confidence intervals were computed for each scale as 95% [CSS +/- 1.96*(*SE*)], where $SE = SD * \sqrt{1-\alpha}$ (Brown 1999).

Linear regression analyses were conducted using the ASD participants who had contemporaneous demographic data to examine the influences of participant characteristics on raw totals and on the calibrated domain scores. Participants' verbal and nonverbal IQs were entered into the first block, followed by participant chronological age, gender, maternal education and race in the second block. Only model R^2 are reported due to multicollinearity which limits interpretability of individual coefficients. Where there was more than one significant predictor, Forward Stepwise models were used to assess the relative contributions of those predictors in predicting raw totals and calibrated domain scores.

Four cases for whom longitudinal data were available were selected to demonstrate the utility of the calibrated total and domain scores for examining trajectories of ASD symptoms across the lifespan.

Results

Analysis of Original Module 4 Algorithm

Domain Total Distributions.

First, domain total distributions were examined for each domain separately to assess the utility of the original Module 4 in discriminating between diagnostic groups and in describing severity of core symptoms within ASD participants. As expected, original Module 4 Communication totals differed significantly by diagnostic group; $F(2,434)=87.04$, $p<.001$. Participants with Autism had significantly higher scores than participants with Other-ASD diagnoses and both ASD groups had higher scores than the Non-ASD cases. (See Table 4.1.) As shown in Figure 4.1, Communication domain scores in the ASD sample were roughly normally distributed, with a slight right-skew; totals of 2-4 were the most frequent (22.5-23.1% each). Maximum Communication totals of 7 or 8 were rarely observed (a total of 2.9% of Module 4 ASD cases received either score). For 3 of the 4 items comprising the Communication domain (*Stereotyped/Idiosyncratic Use of Words or Phrases*, *Conversation* and *Emphatic or Emotional Gestures*), scores of '1' were given in 45-56% of ASD cases. Over half of the participants scored a '0' on the fourth item, *Descriptive, Conventional, Instrumental, or Informational Gestures*.

Social totals also differed across diagnostic groups, such that the Autism group scored significantly higher than the Other-ASD group which scored significantly higher than the Non-ASD group, $F(2,434)=128.65$, $p<.001$ (see Table 4.1). As shown in Figure

4.1, Social domain scores were approximately normally distributed in the ASD participants; Social totals of 6-8 were most frequent (12.1-16.1% each). Fewer than 9% of ASD cases received maximum Social domain scores of 12-14. This is partially explained by a high frequency of '1' scores (i.e., 69-71%) for several items (*Facial Expressions Directed to Others*, *Quality of Social Overtures*, *Quality of Social Response*). The remaining 3 items (with the exception of *Unusual Eye Contact*, which does not have the option of a '1') had 39-44% of cases with '1'.

Correlation with Participant Characteristics.

Next, correlations between domain totals and participant characteristics were examined to inform the potential need for creation of algorithms based on ability level or age (i.e., as with the "Younger than 5 years" and "Greater or Equal to 5 years" algorithms for Module 2). Dividing Module 4 recipients by language level was not helpful because of limited variability in the *Overall Level of Language* item (88% scored a '0,' indicating that the participant "Uses sentences in a largely correct fashion"). Divisions of groups according to other items were similarly unhelpful in Module 3 (Gotham et al. 2007). Among ASD participants, correlations between Social-Communication totals and verbal (r=-.28; n=324; p<.001) and nonverbal IQ (r=-.21; n=314; p<.001) were significant, but weak (Cohen 1988). Correlations were further reduced when only individuals without intellectual disability were included (i.e., IQ $\geq$ 70; n=303; VIQ: r=-.17, p=.003; NVIQ: r=-.09, p=.25). When the sample was limited to individuals with verbal IQ $\geq$ 85 (n=259), there was not a significant relation with IQ, even though this still represented a substantial range (85-148). Social-Communication totals were not significantly associated with age (r=-.07, p=.186).

Analysis to Develop New Module 4 Algorithm

Exploratory Factor Analysis.

Exploratory Factor Analysis was performed with all preferred items included. As shown in Table 4.2, a 2-factor solution fit well, consistent with other modules (Gotham et al. 2007). Ten items loaded on to the Social Affect (SA) factor and 5 items loaded on the Restricted and Repetitive Behaviors (RRB) factor. Factors were significantly correlated (r=.46). As on Modules 1-3, some items from the Communication domain emerged on the SA domain (*Conversation and Emphatic Gestures*) and others on the RRB domain. *Stereotyped/Idiosyncratic Use of Words or Phrases* loaded on to the RRB factor, as had been the case for Modules 1-3. In addition, *Speech Abnormalities Associated with Autism*, previously a Communication domain item, also loaded on to the RRB factor. *Unusual Eye Contact* loaded solidly on both factors. When the sample was reduced to only participants with ASD, *Unusual Eye Contact* again loaded on both, but more strongly on the Social Affect factor (.43) compared to the RRB factor (.26). Because of the theoretical significance of this item and its inclusion as part of the SA domain for Modules 1-3 (Gotham et al. 2007) and the *ADOS*-T (Luyster et al. 2009), it was maintained in the same domain for Module 4.

Confirmatory factor analysis with each item assigned to one of two factors indicated good fit (CFI=.93); the 2-factor solution was a better fit than the 1-factor model (CFI=.91).

Logistic Regression Check on Weighting Domains.

Logistic regression for ASD (i.e., Autism, PDD-NOS and Aspergers) versus Non-ASD cases indicated that both SA (B=.37, SE=.06, z=6.20, Exp(B)=1.45) and RRB

(*B*=1.35, *SE*=.21, *z*=6.52, Exp(*B*)=3.85) totals were predictive of diagnosis. These results suggest a larger effect of RRB domain totals on predicting diagnosis compared to SA totals.

Correlations between Domain Totals, Items and Chronological Age and IQ.

Correlations between each algorithm item and the domain scores minus that item were significant. Correlations ranged from *r*=.40 to *r*=.71 for the SA domain and from *r*=.25 to *r*=.57 for the RRB domain. Domain totals were also significantly correlated (*r*=.48). Internal consistency, measured by Cronbach's alpha (Cronbach 1951), was comparable to other modules for both the SA (α=.84) and RRB (α=.61) domains.

In contrast to the original Module 4 algorithm totals, the new Module 4 algorithm total (SA+RRB) demonstrated a significant, albeit weak, correlation with age (*r*=-.21, *p*<.001) but not with verbal or nonverbal IQ. New SA and RRB domain totals were also weakly correlated with age (*r*=-.20 and *r*=-.15, respectively, *p*<.001), but not IQ.

Correlations between each of the items comprising the new algorithm and chronological age and IQ were also examined. Ten items were significantly correlated with age, ranging from *r*=-.10 (*Amount of Reciprocal Social Communication)* to *r*=-.19 (*Quality of Social Overtures).* Only *Communication of Own Affect and Insight* were significantly, though weakly, correlated with verbal IQ (*r*=-.13, *r*=-.26, respectively) and nonverbal IQ (*r*=-.13, *r*=-.19).

Sensitivity and Specificity Comparison

Receiver Operating Characteristic (ROC) curves were computed to provide information regarding where to set cut-offs in order to maximize sensitivity and specificity of the old and new algorithms. For the new algorithm, ROC curves were run

separately for the SA total and the combined SA + RRB total. Based on analyses indicating correlations between previous algorithm totals and IQ, sensitivity and specificity for the combined SA+RRB total was also examined for three verbal IQ groups (below average (<85), average (85-115), above average (>115). As in the past, scores of 3 were recoded to 2 for this procedure.

As shown in Table 4.3, the new algorithm performed better than the old algorithm. The combined SA+RRB total yielded somewhat higher sensitivity and considerably higher specificity than the SA total alone, both in the overall sample, as well as in each of the three IQ groups. Specificity was also generally higher in the new algorithm compared to the old algorithm, with the exception of the average IQ group. The difference here (i.e., between specificity of 77 vs. 82%) was accounted for by 2/44 fewer Non-ASD participants accurately classified by the revised algorithm compared to the old algorithm.

Development of Calibrated Severity Score

Examining overall total and domain score distributions across age groups.

Although correlations between new algorithm totals and age were weak, score distributions were examined across age groups to confirm whether there was need for age-based calibration cells. As shown in Figure 4.2a, c, and e, distributions of total and domain scores were relatively similar across age groups, with the exception of the oldest two age groups (40-49 and 50-59 years), which included a total of only 9 ASD participants. Nonetheless, given the larger number of algorithm items in Module 4 (15) than Modules 1-3 (14), a calibrated severity metric was warranted in order to allow for comparison across modules. As such, it was decided to create a single calibrated severity

metric for all Module 4 participants aged 9-39 years, including a total of 338 participants with ASD for calibration. Table 4.4 shows the mappings of raw overall, SA and RRB totals to the 10-point severity scale and confidence intervals for each scale.

Comparison of New Module 4 Algorithm Totals and Calibrated Domain Scores Across Age Groups.

As shown in Figure 4.2b, d, and f, distributions of overall calibrated scores remained relatively comparable across age groups. The 19-20 year old group had a somewhat narrower distribution of overall CSS compared to the other ages. Examination of SA-CSS and RRB-CSS distributions suggest this may be due to the higher proportion of 19-20 year olds that exhibited few repetitive behaviors during the *ADOS* (4% with raw RRB-CSS of 1, reflecting no RRB during the *ADOS*; 17% with RRB-CSS of 5, reflecting that RRBs during the *ADOS* were rare and unclear).

As noted above, the *ADOS* classification, based on the raw overall total (SA+RRB) was used to anchor the raw total-to-overall severity score mapping (i.e., "Autism Spectrum" classification mapped to CSS of 4-10). Using this approach, mean overall CSS distinguished between individuals grouped by clinicians' best estimate clinical diagnoses (i.e., Autism vs. Other-ASD vs. Non-ASD diagnoses); $F(2,409)=191.45$, $p<.001$.

Next, percentiles from the raw total-to-overall CSS mapping were used to map raw domain totals-to-domain severity score mappings. Ninety-nine percent of participants with an *ADOS* classification of "Autism Spectrum" had an SA-CSS of 4 or higher and 76% of individuals with an *ADOS* classification of "Nonspectrum" had an SA-CSS less than or equal to 3. With regard to RRB-CSS, 84% of participants with an *ADOS*

classification of "Autism Spectrum" had an RRB-CSS of 6 or higher and 79% with a "Nonspectrum" classification had an RRB-CSS of 5 or lower. Mean domain CSS also distinguished between best estimate clinical diagnostic groups (SA-CSS: $F(2,409)= 132.68$, $p<.001$; RRB-CSS: $F(2,409)= 150.63$, $p<.001$). Nonetheless, there was marked overlap in the distribution of scores (see Figure 4.3).

Correlations Between Domain Calibrations and Overall Calibrated Score.

In the ASD sample, associations between SA-CSS and RRB-CSS were significant, but weak ($r=.26$). Although correlations between each of the domain calibrated scores and the overall CSS were strong, the association between SA-CSS and CSS ($r=.90$) was greater than that observed for RRB-CSS and CSS (r=.60).

Comparison of Raw Overall and Domain Totals to Calibrated Scores

The final model including all predictors explained a total of 9% of variance in the overall Raw Total. Verbal IQ emerged as the only significant predictor of the Raw Total. The same model accounted for 7% of the variance in the overall CSS, with verbal IQ again making the only significant contribution to CSS. As there was only one significant predictor of the overall raw total and CSS, a Forward Stepwise model was not run.

Participant characteristics explained 10% of the variance in the raw Social Affect total, with Verbal IQ and chronological age the only significant predictors. This model accounted for approximately 8% of variance in the SA-CSS, with Verbal IQ and chronological age again emerging as significant predictors. Verbal IQ and age were entered into a Forward Stepwise model to assess the relative contributions of each of these variables in predicting SA-CSS. In the forward model, Verbal IQ explained 4.9% of

the variance; $F(1,270)=13.89$, $p<.001$. Age was excluded by the model, indicating that it was not significant.

In the model predicting RRB-Raw, participant characteristics accounted for 5% of the variance. Race (Caucasian vs. all others) was the only significant predictor of RRB-Raw. The overall model predicting RRB-CSS also accounted for approximately 5% of the variance, with race remaining the only significant predictor reflecting somewhat higher RRB-CSS for Caucasian participants compared to participants of other races.

Case Summaries

Four children were chosen to demonstrate the utility of the newly calibrated Module 4 scores for examining severity of ASD symptoms over time (see Table 4.5 for child characteristics at first and last assessments). Each child's overall CSS, SA-CSS, and RRB-CSS are plotted by age in Figure 4.4.

Case 1. "John," a Caucasian male, was seen at 2 years of age as part of a clinical research study. He received a diagnosis of Autism at that time. During his first *ADOS*, John exhibited severe social-communication symptoms (i.e., an SA-CSS=9) and somewhat milder repetitive behaviors (RRB-CSS=7). A period of stability in toddlerhood was marked by limited use of nonverbal communication behaviors (i.e., gestures, facial expressions and eye contact) and poor quality social overtures. John showed an apparent decrease in severity of social-communication symptoms at age 10, reflecting an increased range of gestures and facial expressions and improved quality of overtures. At age 18, however, John's SA-CSS increased by one point. At this age, he maintained gains in nonverbal communication, but exhibited more unusual overtures and responses. In contrast, John's severity of RRB symptoms increased from ages 2 to 3, and then

remained relatively stable across childhood and adolescence. Scores reflect persistent sensory interests, preoccupations with objects and references to highly specific topics during the *ADOS*. In John's case, the overall CSS showed an apparent increase in symptoms during toddler years, followed by apparent stability across childhood and adolescence. It did not capture the possible divergence in social-communication and repetitive behavior symptom severity in middle childhood.

Case 2. "Parker," a Caucasian male, was diagnosed with Autism at 2 years of age when seen as part of a clinical research study. At first assessment, Parker showed severe repetitive behaviors (RRB-CSS=10) and moderately severe social-communication symptoms (SA-CSS=7). However, his overall CSS demonstrated an apparently steady decrease in symptoms across childhood and adolescence. Separate examination by domain indicates that Parker demonstrated a decrease in repetitive behaviors (i.e., sensory interests and unusual preoccupations) during his *ADOS* at age 3, but persistent challenges in social-communication behaviors (e.g., limited eye contact, facial expressions and mildly inappropriate social overtures). Through middle-childhood, the severity of his repetitive symptoms appeared stable. However, he showed significant improvements in social-communication, particularly in nonverbal communication and quality of social overtures. By his final assessment at age 18, Parker demonstrated only subtle difficulties with social reciprocity and no repetitive behaviors.

Case 3. "Emily," an African American female, was seen as part of a clinical research study and diagnosed with Autism just before her 3rd birthday. Emily's overall and domain CSS scores followed very similar trajectories, demonstrating relative stability of symptom severity in toddlerhood, followed by an apparent decrease in symptoms at

age 5. At age 11, she showed a possible worsening of both social-communication and repetitive behaviors; however, at her last assessment just before her 20th birthday, severity scores again decreased, returning to the milder range that was observed at age 5. The apparent increase in symptoms at age 11 may reflect a shift in her skill set. During her first three assessments, the Module 1, No Words algorithm was employed and Emily demonstrated improvements in initiation of social-communication behaviors, such as pointing, showing and initiation of joint attention, and a decrease in repetitive interests. At age 11, Emily gained sufficient language to be assessed with a Module 3, during which she exhibited limited social reciprocity, inappropriate overtures, repetitive speech and frequent references to unusual or highly specific topics. These symptoms were notably improved at her last assessment.

Case 4. "Robert," an African American male was assessed at age 3 for a clinical research study and diagnosed with PDD-NOS. Robert's overall CSS suggests a steady worsening of symptoms across childhood and adolescence. Separate examination of symptoms by domain indicates that Robert exhibited an increase in social-communication severity (SA-CSS=4 at age 3 and 10 at age 19). Across time, Robert exhibited limited eye contact and mildly unusual social overtures. As he grew older, his social overtures and social reciprocity decreased and he displayed a flatter affect than previously observed. In contrast, Robert's repetitive behaviors remained relatively stable (RRB-CSS=5 at ages 3 and 19 reflecting mild speech abnormalities and brief repetitive interests), with the exception of his age 11 *ADOS*, during which he did not exhibit any repetitive behaviors.

Discussion

In the current study, the original Module 4 algorithm domain totals discriminated between diagnostic groups (i.e., Autism vs. Other-ASD vs. Non-ASD) and provided good sensitivity (89.6%) and adequate specificity (72.2%). However, the somewhat restricted range of the Communication and Social domain totals suggested that the original Module 4 algorithm was not the best combination of items to describe severity of core symptoms within the ASD group. Although items were originally designed to capture a higher proportion of '1' scores (i.e., approximately 50%), the finding that several items received scores of '1' more than two-thirds of the time supported the need to consider items with more variability in the creation of new algorithms.

This larger, more diverse sample provided the opportunity to revise the Module 4 algorithm using items corresponding or equivalent to the revised algorithms for Modules 1-3 (Gotham et al., 2007). The two-domain Module 4 algorithm is consistent with DSM-5 criteria for ASD. Moreover, the addition of the RRB domains improves the diagnostic utility of the *ADOS* to discriminate between individuals with ASD and Non-ASD diagnoses. The new algorithm yields improved sensitivity and specificity, both above 80% in the overall sample.

In light of changes in DSM-5 ASD diagnostic criteria, a single cut-off score that yields a good combination of sensitivity and specificity is provided to differentiate between ASD and Non-ASD classifications. This contrasts to other modules, which have separate cut-offs available for Autism and ASD. For researchers who may be interested in achieving a higher level of specificity, at the cost of somewhat lower sensitivity (i.e., equivalent to an *ADOS-2* classification of "Autism" only on other modules), a cut-off of

10 may be useful. This cut-off yields an overall specificity of 91.1%, but there is also a sizeable reduction in sensitivity (79.3% overall; 71.3% for individuals with above average IQs; data not shown, additional information available from authors upon request). It is also noteworthy that, although DSM-5 criteria now require that an individual exhibit deficits in both social-communication and restricted and repetitive behaviors, separate domain cut-offs are not provided on the *ADOS* algorithm. As noted above, the time-limited, standardized nature of the *ADOS* may influence the extent to which some individuals exhibit RRBs during the assessment. Thus, it is likely that implementing a separate RRB-domain cut-off would reduce sensitivity of the instrument (i.e., some individuals with ASD will exhibit few or no RRBs during the 40 minute observation period). For example, in this sample, the SA cut-off of 6 shown in Table 4.3 and an RRB cut-off of 1, specificity improves to 96.7%; however, sensitivity is reduced to 75.5% [data not shown, additional information available from authors upon request]. Nonetheless, inclusion of both domains (SA+RRB) in the overall total results in considerably better specificity (and somewhat higher sensitivity; see Table 4.3) than relying on the SA domain alone. While the *ADOS* is not designed for use in isolation as a DSM "checklist" to determine a clinical diagnosis of ASD, it provides highly valid instrument classifications and a useful context in which to observe behaviors relevant to clinical diagnosis.

The new Module 4 algorithm totals were weakly, but significantly correlated with age. In the cross-sectional design of this study, it is not clear if these differences are due to recruitment effects or if they reflect true developmental variation. The new algorithm totals were not significantly correlated with IQ. However, correlations between previous

algorithm totals and IQ suggested a need to examine performance across different levels of cognitive ability. Although an estimate of IQ or developmental level is recommended as a key part of a diagnostic evaluation (Hus & Lord 2011), we acknowledge that such information is not always available. As such, it did not seem practical to make separate algorithms for individuals of different cognitive levels, all of whom had fluent, complex language. Overall, IQ appears to be considerably less influential for Module 4 scores than was observed for Modules 1-3 (Gotham et al. 2007). Nonetheless, researchers are encouraged to be attentive to the fact that specificity may be poorer in individuals with average IQs. Because the Non-ASD samples within IQ ranges are relatively small, it is difficult to know exactly what this means. It is notable that, though this group demonstrated lower specificity than other IQ groups for Module 4, specificity was actually comparable to that observed for most other algorithms (i.e., Modules 1-3). This contrast reflects that, in fact, Module 4 specificity is generally quite good. This finding also highlights the fact that, as with all of the ASD diagnostic instruments, the *ADOS* is best used as one measure of behavior in combination with other sources of information.

Associations between the new Module 4 raw totals and participant characteristics were small (i.e., 5-10% variance explained by participant characteristics not specific to ASD, such as verbal IQ, age and race). Calibrated severity scores were derived in order to facilitate comparison to Modules 1-3 used with younger children and individuals with limited language skills, for which greater influences of participant characteristics on algorithm totals are observed (Gotham et al. 2009; Hus et al. 2013). It was also important to take into account differences in number of items across modules on the RRB domain, given the decision to include both *Stereotyped/Idiosyncratic Use of Words or Phrases*

and *Speech Abnormalities* on the new Module 4 algorithm, while also keeping theoretically meaningful items capturing sensory interests, hand mannerisms and excessive references to highly specific topics. In the end, this yielded a total of 5 items on the Module 4 RRB domain, compared to only 4 items for Modules 1-3.

As observed for Modules 1-3, there was overlap in the distributions of domain calibrated scores across diagnostic groups. Overlap of the Non-ASD group with the Autism and Other-ASD groups may reflect recruitment bias (i.e., some of our Non-ASD sample had been referred for assessment of ASD, but received a clinical Non-ASD diagnosis). However, this also reflects the intention of the calibrated score to provide a continuous, quantitative dimensional measures of social-communication and repetitive behaviors extending beyond diagnostic categorization, consistent with the collapsing of diagnostic categories in DSM-5.

It is hoped that the newly revised Module 4 algorithm and CSS will help to expand research efforts to better understand the specific strengths and difficulties in social-communication and repetitive behaviors experienced by adults with ASD. The Module 4 CSS offers the opportunity for comparisons to scores obtained from Modules 1-3, used with younger and less verbally fluent children and adolescents. Thus, the CSS allows examination of longitudinal trajectories of ASD symptoms across childhood and into young adulthood.

Module 4 scores can also be used to further our understanding of how ASD symptom severity interacts with other factors, such as verbal, cognitive and adaptive ability to predict functional outcomes for adults with ASD. Moreover, the Module 4 CSS

may be a useful phenotyping measure for neurobiological studies seeking to draw associations between dimensions of ASD and differences in brain structure or function.

Clinically, the Module 4 revisions yield scores that provide a more accurate summary of ASD symptoms, with an algorithm that is more closely aligned with DSM-5 criteria than the original algorithm. It also affords good sensitivity and improved specificity compared to the original Module 4 algorithm. Although it is always recommended that the *ADOS* be used as one source of information in a diagnostic battery, good specificity is particularly important in the assessment of adults, for whom parents are not always available to provide the comprehensive developmental history that is often helpful in making differential diagnoses. Finally, clinicians may use the Module 4 CSS to monitor symptom severity during the course of treatment. However, it is important to remember that the *ADOS'* primary use is as a *diagnostic instrument* and the CSS is intended to capture severity of *core symptoms* that may not be expected to remit in the same way symptoms of depression or anxiety are reduced in response to treatment. Moreover, because the CSS is *not* intended as a measure of functional impairment, it may not be as sensitive to more subtle changes as measures of adaptive social functioning. Thus, while a significant reduction in scores over time may be viewed as evidence of improvement, stability of scores should not be viewed as discouraging. Notably, confidence intervals (shown in Table 4.4) should be taken into account when assessing the clinical significance of a change in score.

Limitations

Sensitivity and specificity of the algorithm may vary in different clinical and research settings as a consequence of differences in examiner skill, sequence of

administration and other factors (Gotham et al. 2007). While the Non-ASD group is the largest to-date used in the validation of the *ADOS*, it is a diagnostically diverse group. Future studies examining the diagnostic utility of the *ADOS* in more specific comparison samples (e.g., individuals with mood disorders) would be useful to inform understanding of the behavioral patterns observed in other groups and the Module 4's ability to differentiate between ASD and Non-ASD diagnoses.

Results of a recent study examining the validity of the *ADOS* in a sample of adults suggest that our revisions to the Module 4 algorithm will increase discriminative validity in difficult to differentiate psychiatric groups (Bastiaansen et al., 2011). This study demonstrated good overall specificity of the original Module 4 algorithm (.82) in adults with Psychopathy, Schizophrenia or typical development. Domain totals discriminated between the ASD vs. the Psychopathy and typically developing groups, but did not discriminate between the ASD and Schizophrenia group. This was thought to be due to the overlap in negative symptoms observed in both ASD and Schizophrenia (e.g., limited range of directed facial expressions and lack of asking the examiner for information). Application of the revised Module 3 algorithm (Gotham et al., 2007) differentiated between the ASD and each of the three groups, including the Schizophrenia group. Examination of individual items suggested that only three of the 22 Module 4 items distinguished the ASD from the Schizophrenia group: *Stereotyped Language*, *Quality of Social Response* and *Overall Quality of Rapport.* All three of these items are included in the revised Module 4 algorithm, in addition to seven items found to differentiate ASD from the psychopathy and typically developing groups. Given that our changes to the Module 4 algorithm have increased comparability to the revised Module 3 algorithm and

that the new Module 4 algorithm comprises many items shown to differentiate groups in Bastiaansen's study, we would expect that the revised Module 4 algorithm will better differentiate between ASD and these Non-ASD groups than the original algorithm. Examination of Module 4 performance with these and other diagnostic groups will be an important future direction for validating the revised Module 4 algorithm.

New Module 4 totals were weakly associated with age and there was some variability in score distributions across ages, such that 19-20 year olds in this sample had a somewhat narrower distribution of scores compared to other age groups, which showed more similar distributions. It is likely that this difference reflects recruitment bias in this clinical sample. As such, it will be critical that the validity of the Module 4 revised algorithm and calibrated severity metric be replicated in other samples. In addition, older adults (>40 years) in this sample tended to have considerably lower *ADOS* raw totals (Mean=5.4) compared to other age groups, which tended to vary around means of 10-14. This is likely to also reflect sample recruitment biases (i.e., older individuals in our sample were generally self-referred to one of the clinics for a first-time diagnosis and may have had more subtle symptom patterns than those referred at younger ages). Because there were only 9 participants in this older age group, it was decided to exclude them from the calibrated severity score derivation. A larger sample of participants over 40 years of age is needed to explore these differences in symptom presentation for older individuals.

Effects of race on both raw and calibrated RRB totals are also likely to be an artifact of recruitment bias in this predominantly Caucasian sample. As noted for Module

1-3, studies of the *ADOS* in population-based or clinical samples recruited outside of North America will be crucial (Hus et al. 2013).

Conclusion

The revised Module 4 algorithm provides improved sensitivity, while maintaining or increasing specificity across individuals of different cognitive levels. The revised algorithm is consistent with the revised DSM-5 two-domain criteria for ASD and offers increased comparability to recently published algorithms for *ADOS* Modules T and 1-3 (Lord, Rutter, et al. 2012; Lord, Luyster, et al. 2012). Module 4 calibrated severity scores provide quantitative estimates of severity of social-communication and repetitive behaviors that are relatively independent of participant characteristics. The new severity scores also extend the ability to compare domains and overall totals across modules. These changes will facilitate future research efforts to increase understanding of the strengths and challenges experienced by adults with ASD.

Table 4.1 Sample characteristics

	Autism					Other-ASD					Non-ASD				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
age (years)	177	20.12	6.30	10.33	55.58	170	21.14	7.79	9.92	54.00	90	25.17	12.35	13.33	62.25
VIQ	145	96.79	23.52	34	148	136	107.43	20.27	38	150	82	91.94	27.37	23	144
NVIQ	144	94.42	21.99	32	140	135	103.44	19.60	49	153	81	94.48	26.61	33	147
ADI Social	139	19.30	6.38	0	30	122	16.41	6.31	2	28	66	9.67	6.55	0	24
ADI Comm-V	139	15.73	4.99	4	24	122	13.12	5.13	0	24	66	8.18	5.82	0	24
ADI Comm-NV	139	8.58	3.80	0	14	122	7.43	3.85	0	14	66	4.18	3.80	0	14
ADI-RRB	139	6.32	2.62	0	12	122	4.87	2.49	0	12	66	3.52	2.71	0	11
ADOS Comm	177	3.85	1.58	0	8	170	2.85	1.45	0	6	90	1.32	1.34	0	6
ADOS Social	177	8.38	2.63	3	14	170	6.50	2.49	0	14	90	3.09	2.50	0	11
ADOS RRB	177	1.75	1.55	0	8	170	.91	1.00	0	4	90	.33	0.65	0	3

VIQ = verbal IQ; NVIQ = nonverbal IQ; ADI=Autism Diagnostic Interview-Revised; Comm-V=Verbal Communication; Comm-NV=Nonverbal Communication; RRB=Restricted, Repetitive Behavior; ADOS=Autism Diagnostic Observation Schedule

Table 4.2 Revised Module 4 algorithm mapping

Domains	Module 4 N=437	Factor Loadings	Module 3 N=398*	Factor Loadings
Social Affect	Unusual Eye Contact**	0.37	Unusual Eye Contact	0.51
	Amount of Social Communication	0.86	Amount of Social Communication	0.81
	Facial Expressions	0.54	Facial Expressions	0.67
	Quality of Rapport	0.67	Quality of Rapport	0.72
	Comm. Own Affect	0.65	Shared Enjoyment	0.84
	Quality Social Overtures	0.57	Quality Social Overtures	0.65
	Conversation	0.74	Conversation	0.73
	Emphatic Gestures	0.57	Descriptive Gestures	0.71
	Quality of Social Response	0.49	Quality of Social Response	0.6
	Insight	0.59	Reporting of Events	0.65
Eigen Value	6.4		6.2	
Restricted Repetitive Behaviors	Speech Abnormalities	0.59		
	Stereotyped Language	0.81	Stereotyped Language	0.60
	Unusual Sensory Interest	0.60	Unusual Sensory Interest	0.44
	Highly Specific Topics	0.46	Highly Specific Topics	0.63
	Hand Mannerisms	0.54	Hand Mannerisms	0.48
Eigen Value	1.7		1.7	
RMSEA	0.08		0.06	
Rho	0.46		0.38	

*reproduced from Gotham et al., 2007 for comparison

**loads .45 on RRB when all sample (ASD & nonASD) included; loading is higher for SA when only ASD included in EFA

RMSEA = Root Mean Square Error (values 0.08 or less indicate a good fit).

Rho = correlation between Social Affect & Restricted Repetitive Behaviors factors.

Items from the 2000 algorithm not included in new Module 4 algorithm: Descriptive Gestures, Responsibility

Loadings from FA including all participants (ASD & Non-ASD; N=437)

Table 4.3 Sensitivity and specificity of previously used and revised algorithms

Sensitivity and specificity of previously used and revised algorithms

		Overall		VIQ <85		VIQ 85-115		VIQ >115	
		Sens	Spec	Sens	Spec	Sens	Spec	Sens	Spec
		ASD=347	NS=90	ASD=67	NS=26	ASD=156	NS=44	ASD=101	NS=14
2000 algorithm	Met 3 domains*	89.6	72.2	89.5	73.1	84.0	81.8	79.2	85.7
New algorithm	SA only (cut=6)	89.0	72.2	91.0	65.4	90.4	68.2	86.1	85.7
	SA+RRB (cut=8)	90.5	82.2	94.0	80.8	91.7	77.3	87.1	92.9

*Met or exceeded cut-offs on Social, Communication and Social+Communication domains; VIQ=Verbal IQ; Sens=Sensitivity; Spec=Specificity; SA=Social Affect; RRB=Restricted, Repetitive Behavior

Table 4.4 Mapping of ADOS Module 4 raw overall and domain totals to calibrated severity scores for ages 9-39 years

	Raw totals		
CSS	Overall Total	SA Domain	RRB Domain
1	0-2	0-1	0
2	3-5	2-3	-
3	6-7	4	-
4	8	5	-
5	9	6	1
6	10-11	7-8	2
7	12-13	9-10	3
8	14-15	11-12	4
9	16-19	13-15	5
10	20-29	16-20	6-10

CSS=Calibrated Severity Score; SA=Social Affect; RRB=Restricted and Repetitive Behaviors;
95% CI: CSS +/- 2.16, SA-CSS +/- 1.94; RRB-CSS +/- 2.99

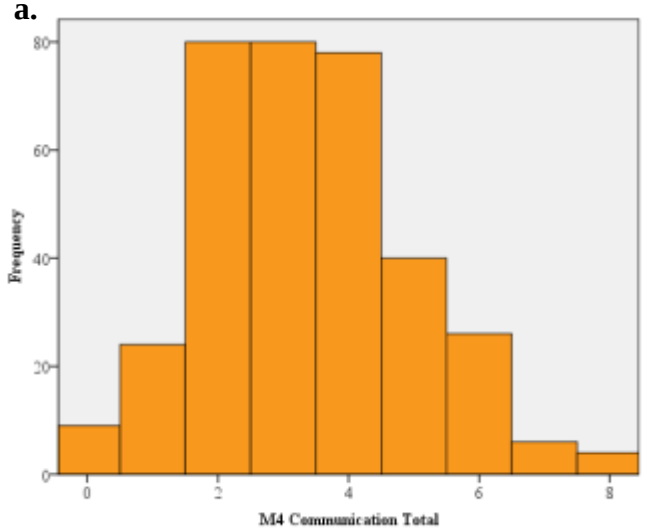
Table 4.5 Case summary characteristics

	First Assessment					Last Assessment				
	Age	VIQ	NVIQ	Mod	Dx	Age	VIQ	NVIQ	Mod	Dx
John	2.1	26	68	1	Autism	18.8	188	85	4	Autism
Parker	2.1	95	92	1	Autism	17.8	108	99	4	No Dx
Emily	2.9	48	85	1	Autism	19.9	72	81	4	Autism
Robert	3.1	53	80	1	PDD-NOS	19.8	72	78	4	PDD-NOS

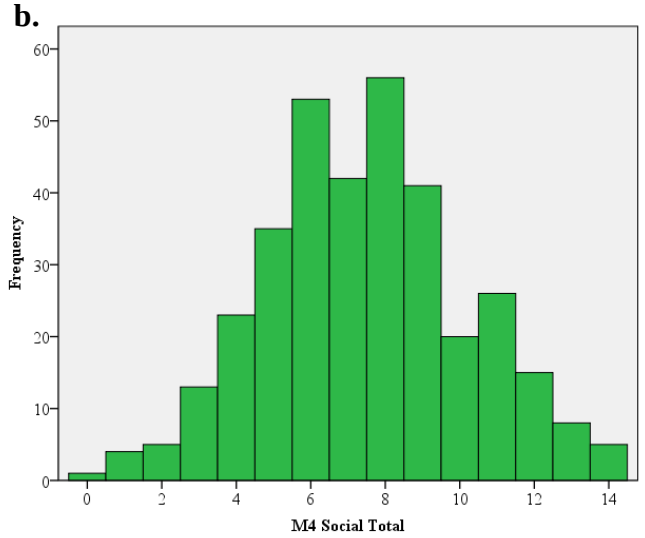
All ages in years; VIQ = verbal IQ; NVIQ = nonverbal IQ; Diagnosis at Last Assessment was based on presentation of symptoms at the time of assessment (examiners were blind to previous diagnosis)

Figure 4.1 Distribution of original Module 4 algorithm domain totals

a.

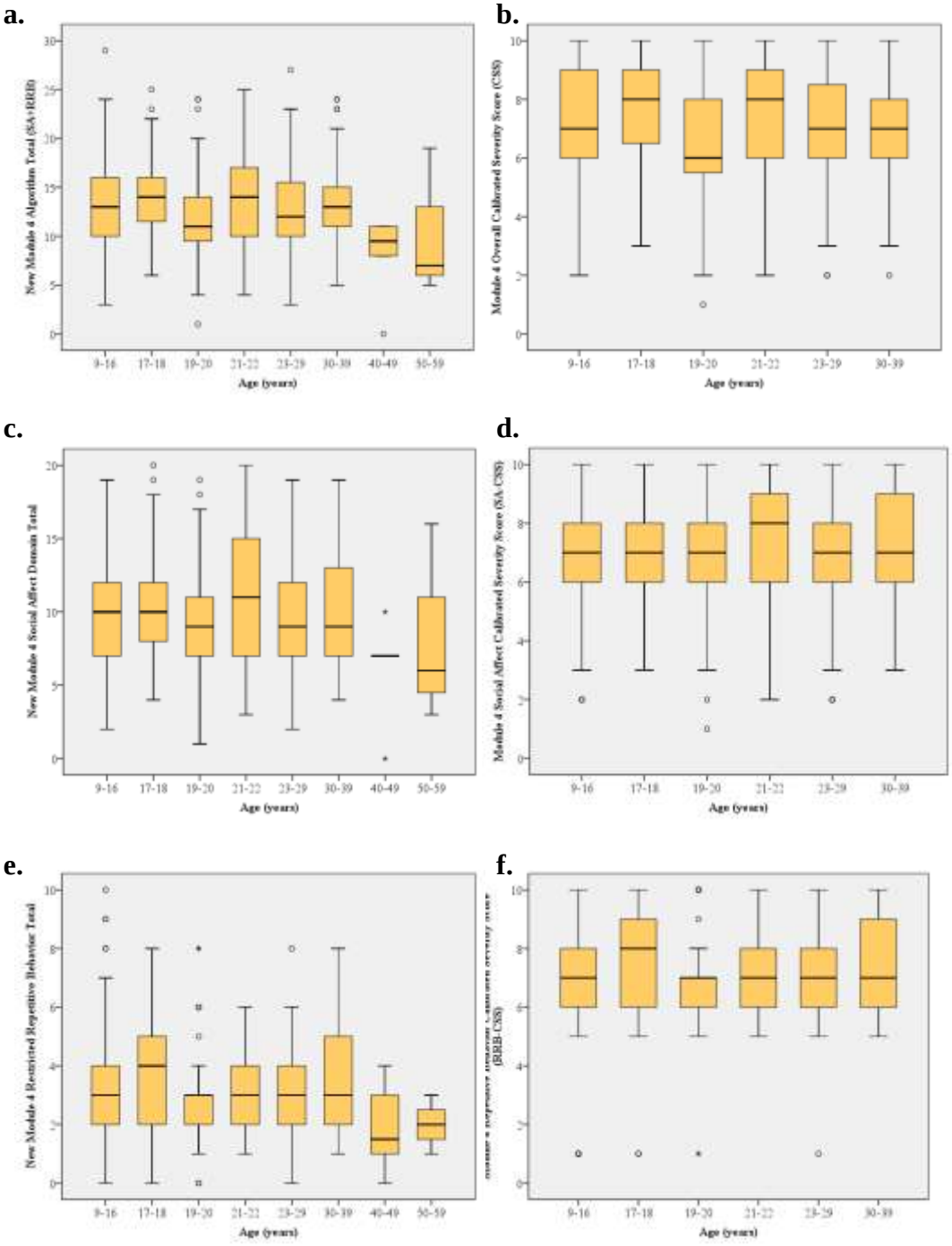


b.



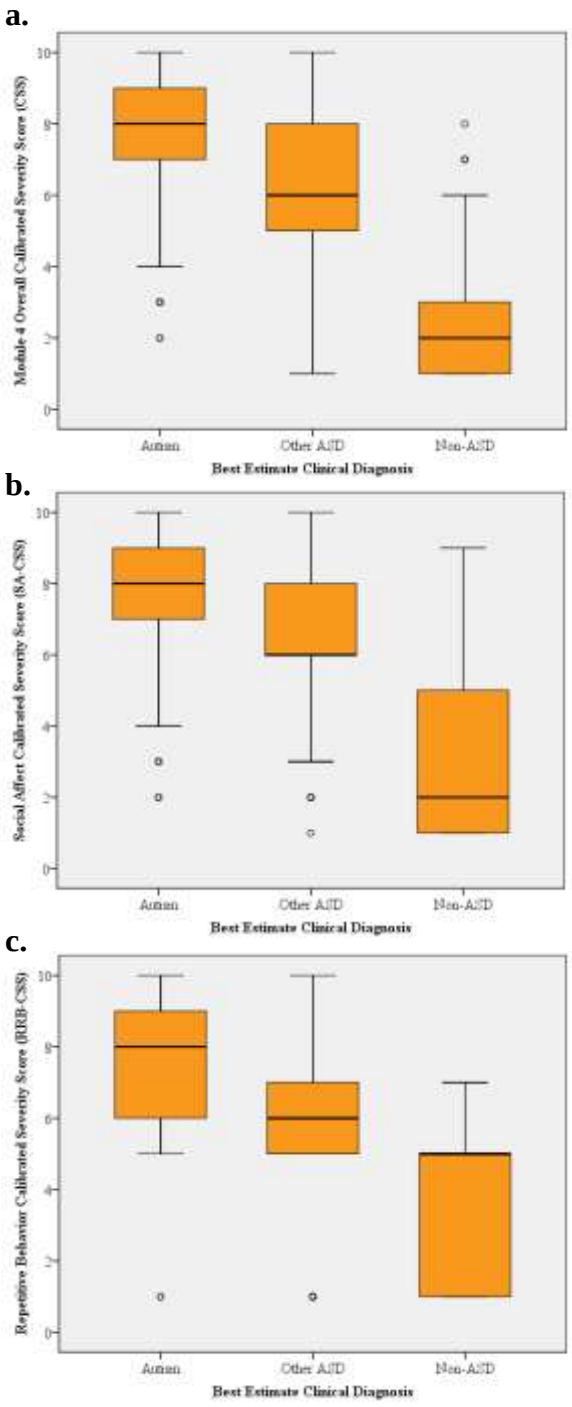
a (top) Module 4 original algorithm raw Communication Total; **b** (bottom) Module 4 original algorithm raw Social Total

Figure 4.2 Distributions of Module 4 algorithm Raw Totals and Calibrated Scores



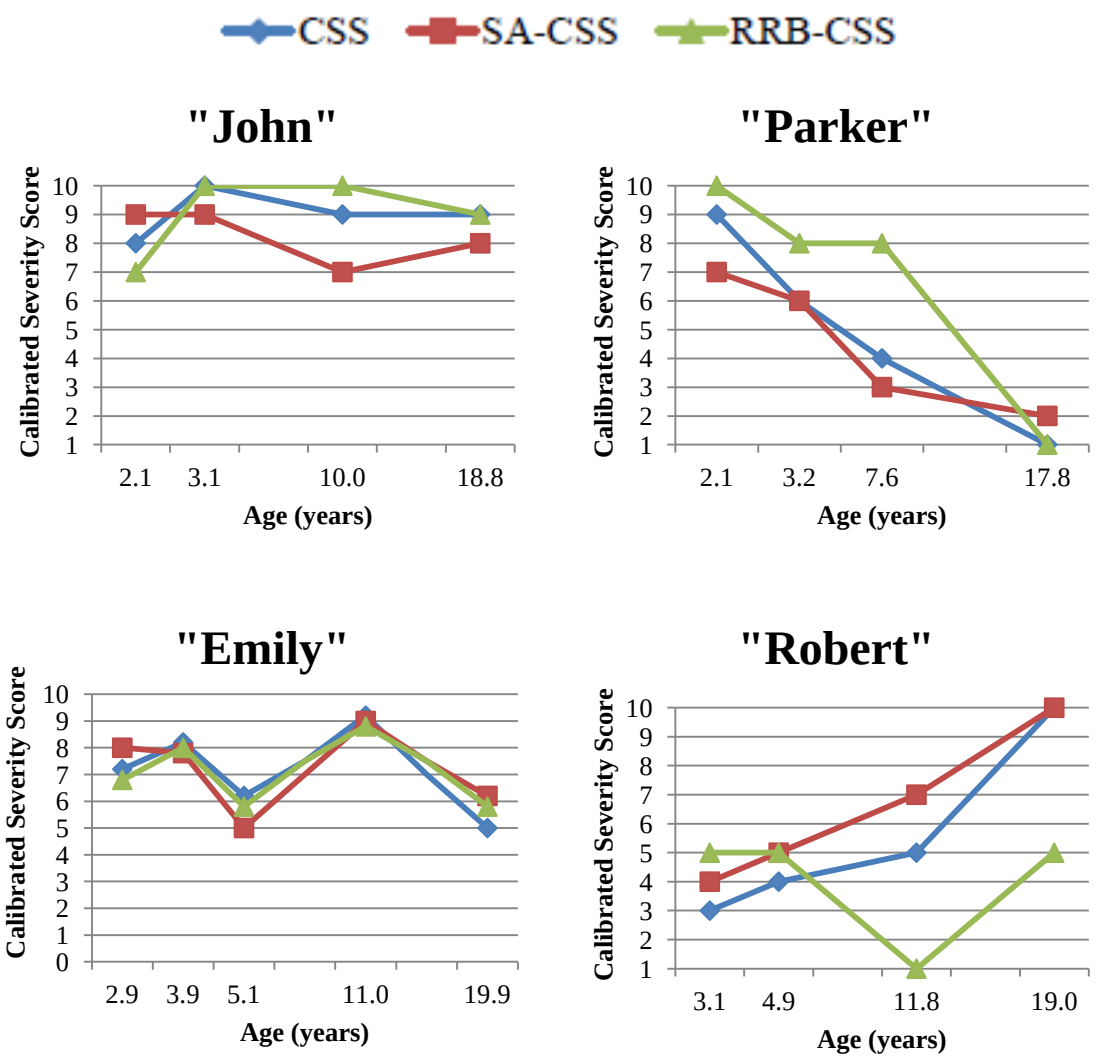
a (top, left) Module 4 Algorithm raw overall totals by age cells; **b** (top, right) Module 4 calibrated overall scores by age cells; **c** (middle, left) Raw Social Affect domain totals by age cells; **d** (middle, right) Calibrated Social Affect domain scores by age cells; **e** (bottom, left) Raw Repetitive Behavior domain totals by age cells; **f** (bottom, right) Calibrated Repetitive Behavior domain scores by age cells.

Figure 4.3 Distributions of Module 4 Calibrated Scores by best estimate clinical diagnosis



a (top) Module 4 Overall Calibrated Severity Score by best estimate clinical diagnosis; **b** (middle) Calibrated Social Affect domain scores by best estimate clinical diagnosis; **c** (bottom) Calibrated Repetitive Behavior domain scores by best estimate clinical diagnosis.

Figure 4.4 Case summaries of longitudinal Calibrated Severity Scores



References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: American Psychiatric Association.

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

Anderson, D. K., Liang, J. W., & Lord, C. (2013). Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. *Journal of Child Psychology and Psychiatry.* Advanced Online Publication. doi:10.1111/jcpp.12178

Brown, J. D. (1999). Statistics Corner questions and answers about language testing statistics: Standard error vs. standard error of measurement. *Shiken: JALT Testing & Evaluation SIG Newsletter,* 3(1), 20–25.

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.

Brugha, T. S., McManus, S., Smith, J., Scott, F. J., Meltzer, H., Purdon, S., Berney, T., Tantam, D., Robinson, J., Radley, J., & Bankart, J. (2012). Validating two survey methods for identifying cases of autism spectrum disorder among adults in the community. *Psychological medicine*, *42*(3), 647–656.

Dunn, L. M., & Dunn, L. M. (2007). *Peabody Picture Vocabulary Test, 4th ed*. Pearson, Bloomington.

Eaves, L. C., & Ho, H. H. (2008). Young adult outcome of autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *38*(4), 739–747.

Elliott, C. D. (2006). *Differential Ability Scales—Second Edition (DAS-II).* San Antonio, TX: The Psychological Corporation.

Elliott, Colin D. (1990). *Differential Ability Scales*. San Antonio, TX: Psychological Corporation.

Farley, M. A., McMahon, W. M., Fombonne, E., Jenson, W. R., Miller, J., Gardner, M., Block, H.., Pingree, C.B., Ritvo, E.R., Ritvo, R.A. & Coon, H. (2009). Twenty-year outcome for individuals with autism and average or near-average cognitive abilities. *Autism Research*, *2*(2), 109–118.

Fein, D., Barton, M., Eigsti, I.-M., Kelley, E., Naigles, L., Schultz, R. T., Stevens, M., Helt, M., Orinstein, A., Rosenthal, M., Troyb, E., & Tyson, K. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry*, *54*(2), 195–205.

Fischbach, G. D., & Lord, C. (2010). The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron*, *68*(2) 192–195.

Ganz, M. L. (2007). The lifetime distribution of the incremental societal costs of autism. *Archives of Pediatrics & Adolescent Medicine*, *161*(4), 343–349.

Gillespie-Lynch, K., Sepeta, L., Wang, Y., Marshall, S., Gomez, L., Sigman, M., & Hutman, T. (2012). Early childhood predictors of the social competence of adults with autism. *Journal of Autism and Developmental Disorders*, *42*(2), 161–174.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(5), 693–705.

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

Helt, M., Kelley, E., Kinsbourne, M., Pandey, J., Boorstein, H., Herbert, M., & Fein, D. (2008). Can children with autism recover? If so, how? *Neuropsychology Review*, *18*(4), 339–366.

Henninger, N. A., & Taylor, J. L. (2013). Outcomes in adults with autism spectrum disorders: a historical perspective. *Autism*, *17*(1), 103–116.

Howlin, P., Moss, P., Savage, S., & Rutter, M. (2013). Social outcomes in mid- to later adulthood among individuals diagnosed with autism and average nonverbal IQ as children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(6), 572–581.e1.

Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry*, *54*(2), 216–224.

Hus, V., & Lord, C. (2011). Evaluation & Testing. In Hollander, Eric, Kolevzon, Alexander, & Coyle, Joseph T. (Eds.), *Textbook of autism spectrum disorders* (pp. 49–65). Arlington, VA: American Psychiatric Publishing, Inc.

Hus, V., & Lord, C. (2013). Effects of child characteristics on the Autism Diagnostic Interview-Revised: Implications for use of scores as a measure of ASD severity. *Journal of Autism and Developmental Disorders*, *43*(2), 371–381.

Hus, V., Taylor, A., & Lord, C. (2011). Telescoping of caregiver report on the Autism Diagnostic Interview – Revised. *Journal of Child Psychology and Psychiatry*, *52*(7), 753–760.

Interagency Autism Coordinating Committee. (2011, January 18). The 2011 Interagency Autism Coordinating Committee Strategic Plan for Autism Spectrum Disorder Research. Retrieved from http://iacc.hhs.gov/strategic-plan/2011/index.shtml

Jones, R.M., Risi, S., Wexler, D., Anderson, D., Corsello, C., Pickles, A., & Lord, C. (2013). How interview questions are placed in time influences caregiver description of social communication symptoms on the ADI-R. *under review*.

Levy, A., & Perry, A. (2011). Outcomes in adolescents and adults with autism: A review of the literature. *Research in Autism Spectrum Disorders*, *5*(4), 1271–1282.

Lord, C., Luyster, R., Gotham, K., & Guthrie, W. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part II): Toddler Module*. Torrance, CA: Western Psychological Services.

Lord, C., Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D. M., Ousley, O., ... Risi, S. (2012). A multisite study of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*, *69*(3), 306–313.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part I): Modules 1-4.* Torrance, CA: Western Psychological Services.

Luyster, R., Gotham, K., Guthrie, W., Coffing, M., Petrak, R., Pierce, K., Bishop, S., Esler, A., Hus, V., Oti, R., Richler, J., Risi, S. & Lord, C. (2009). The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(9), 1305–1320.

Perry, A., Condillac, R. A., & Freeman, N. L. (2002). Best practices and practical strategies for assessment and diagnosis of autism. *Journal on Developmental Disabilities*, *9*(2), 61–75.

Piven, J., Harper, J., Palmer, P., & Arndt, S. (1996). Course of behavioral change in autism: a retrospective study of high-IQ adolescents and adults. *Journal of the American Academy of Child and Adolescent Psychiatry*, *35*(4), 523–529.

Raven, J. C. (1960). *Guide to the standard progressive matrices: sets A, B, C, D and E*. HK Lewis.

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

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

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders, 10*(1), 91–103.

Shattuck, P. T., Seltzer, M. M., Greenberg, J. S., Orsmond, G. I., Bolt, D., Kring, S., Lounds, J. & Lord, C. (2006). Change in autism symptoms and maladaptive behaviors in adolescents and adults with an autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *37*(9), 1735–1747.

Sigman, M., & McGovern, C. W. (2005). Improvement in cognitive and language skills from preschool to adolescence in autism. *Journal of Autism and Developmental Disorders*, *35*(1), 15–23.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling*. Boca Raton, FL: Chapman & Hall/CRC Press.

U.S. Department of Education, Office of Special Education Programs, Data Analysis System (DANS. (2011, July 15). Table 1-3. Number of students ages 6 through 21 served under IDEA, Part B, by disability category and state: Fall 2010. Retrieved from https://www.ideadata.org/TABLES34TH/AR_1-3.pdf

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation.

CHAPTER V

Conclusion

In the last 35 years since being added to the DSM, definitions of Autism Spectrum Disorders (ASDs) have undergone considerable revision – from distinct categorical syndromes to a single disorder encompassing vast symptom heterogeneity. Development of standardized diagnostic measures has informed these changes by providing reliable ways of documenting the wide range of variability in presentation of symptoms across individuals with ASD. However, this research has also highlighted the considerable impact of non-ASD-specific child characteristics, such as developmental level, on measurement of core ASD symptoms (i.e., social-communication deficits and restricted, repetitive behaviors). Imprecise measurement can contribute to decreased diagnostic accuracy and misinterpretation of scores on diagnostic measures as indicators of ASD severity. In clinical settings, inaccurate diagnosis may delay or hinder development of appropriate treatment plans and make it difficult to measure treatment efficacy. In the research domain, misinterpreted scores may hinder efforts to elucidate etiological mechanisms and understand trajectories of development and predictors of outcome.

Now, as we take on the more dimensional approach to diagnosis delineated by the DSM-5, there is a need to further understanding of how core ASD symptoms interact with commonly co-occurring non-ASD-specific dimensions of behavior, such as externalizing behaviors and mood. We must carefully examine the ability of our

diagnostic instruments to quantify ASD-related behaviors and revise these measures to allow us to capture symptom severity in a manner that is less influenced by developmental characteristics and non-ASD-specific dimensions of behaviors. If we can do this, we will be able to make more meaningful comparisons over time and across individuals.

Each of the three studies comprising this dissertation aims to enhance the validity of currently used ASD screening and diagnostic measures to quantify ASD symptom severity. The first study demonstrates the influence of non-ASD-specific child characteristics on scores from the Social Responsiveness Scale (SRS) and the potential utility of statistically controlling for nonspecific factors to increase the probability that associations between SRS scores and genetic or neurobiological mechanisms are due to ASD-related behaviors. In clinical settings, findings that SRS scores are considerably influenced by child characteristics, particularly general behavior problems, highlights the need to exercise caution when using the SRS for ASD screening in clinical populations. More systematic study of how non-ASD-specific child characteristics influence the specificity of the SRS is needed to inform specific recommendations regarding its use in general clinical settings (e.g., hospital outpatient clinics, community mental health).

The second and third studies of this dissertation expand the use of the scores from the Autism Diagnostic Observation Schedule (ADOS) to quantify ASD symptom severity. Clinically, the newly revised Module 4 algorithm provides a more accurate assessment of ASD symptoms in young adults, which will contribute to better diagnostic accuracy and provide a profile of individual strengths and challenges. Moreover, the domain calibrated scores for each of the four developmentally determined modules offer

an indicator of severity of ASD symptoms that can inform treatment planning and track symptom response to intervention. The results of these studies will also further research that seeks to describe trajectories of ASD symptoms over time, predictors of outcome, and behavioral presentation of ASD in adults.

The three studies comprising this dissertation represent steps toward expanding the use of screening and diagnostic measures to quantify ASD symptom severity in a manner that allows for meaningful comparison across time and individuals. Further research on these, and other ASD measures, will enhance our ability to quantify the heterogeneity of symptoms characterizing ASD and expand our understanding of how symptom severity interacts with other child characteristics to produce functional impairments. This knowledge will be key to identifying biological mechanisms that play a role in the causation or manifestation of ASD and informing development of targeted interventions to promote positive outcomes for individuals with ASD.