Examining the Autism Phenotype: The Structure of Autism Spectrum Disorders as Measured by the Autism Diagnostic Observation Schedule

Dissertation

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Megan Norris, M.A.
Graduate Program in Psychology

****

The Ohio State University
2010

Dissertation Committee:
Professor Luc Lecavalier, Advisor
Professor Michael Aman
Professor Michael Edwards
Professor Susan Havercamp
ABSTRACT

Autism spectrum disorders (ASDs) are widely studied yet poorly understood. They are characterized by impairments in social and communication skills and the presence of repetitive and stereotyped behaviors and/or circumscribed interests (RRBs). Much recent attention has been directed towards elucidating the structure of autistic symptoms. A better understanding of the phenotype can lead to improved diagnoses and clarification of etiology and pathogenesis. Factor analytic studies are one way researchers have pursued this end. Most often, two- or three-factor solutions have been reported. The objective of the current study was to test several competing models of the autism phenotype using the Autism Diagnostic Observation Schedule (ADOS). Participants included individuals with ASDs aged 3-18 years (N = 1,409) from the Autism Genetic Resource Exchange database. ADOS data from 720 participants for Module 1 and 689 participants for Module 3 were used in analyses. Samples were divided into more homogenous subgroups to examine the impact of age and level of functioning on model fit. Confirmatory factor analyses (CFAs) were performed on total samples and subsamples. Four primary models were tested: (a) a one-factor model; (b) a two-factor model (one factor consisting of social/communication items and the other consisting of RRBs); (c) a three-factor model based on DSM-IV-TR criteria (social, communication, and RRB factors); and (d) a two-factor model based on proposed DSM-V criteria (one
factor consisting of social/communication symptoms and one factor consisting of restricted and repetitive behaviors and language). Additionally, ADI-R RRB items were added to analyses because these symptoms may not be well captured with the ADOS. Bi-factor models were also examined for the DSM-IV analyses in order to examine the possibility that ASD symptoms were best explained by one general ASD factor and three specific factors. Results of the CFAs with Module 1 indicated all models fit reasonably well, with RMSEAs ranging from .056 (DSM-IV model) to .062 (one-factor and two-factor models). RMSEA confidence intervals overlapped, suggesting no model fit significantly better than other models. Generally, fit improved in the analyses with more homogenous subgroups. Addition of ADI-R RRB items resulted in un-interpretable results. Results of the CFAs with Module 3 indicated unacceptable fit for most models, with RMSEAs ranging from .074 (DSM-V) to .083 (one-factor model). RMSEA confidence intervals again overlapped, indicating all models fit similarly. Unlike Module 1 analyses, indices of fit improved when ADI-R RRB items were included in analyses, but there was again little differentiation between models. Fit improved in the analyses with more homogenous subgroups by age, but not level of functioning. Finally, the bi-factor DSM-IV model did not aide interpretation in either module. The lack of differentiation between models in both modules suggests that the structure of ASD symptoms is complex and several research methods will be necessary to understand the symptom structure. It may also help explain why different solutions are found across
studies; that is, models are similar to each other and different fit indices are found with different subgroups.
ACKNOWLEDGMENTS

I wish to thank my advisor, Dr. Luc Lecavalier for his support, encouragement, and patience, which made this dissertation possible.

I thank my committee, Drs. Michael Aman, Michael Edwards, and Susan Havercamp for their thoughtful comments and suggestions regarding this project.

Finally, I thank my family and friends for their unwavering support and encouragement throughout my graduate career.
VITA

June 16, 1982…………………………………….. Born – Canton, OH
2004……………………………………………… B.A. Psychology, Kent State University
2004- 2005…………………………………….. University Fellow, Psychology
                          The Ohio State University
2005- 2006……………………………………. Graduate Teaching Associate
                          The Ohio State University
2006-present…………………………………. Graduate Associate
                          The Ohio State University

PUBLICATIONS

Research Publications


Research Publications In Preparation


Fields of Study

Major Field: Psychology
Appendix I: LISREL Syntax for Module 3, DSM-V Model

Appendix J: LISREL Syntax for Module 3, DSM-IV Bi-factor Model
List of Tables

Table 1. Summary of studies of ASD structure ................................................................. 19
Table 2. Module 1 sample characteristics ...................................................................... 30
Table 3. Module 3 sample characteristics ...................................................................... 33
Table 4. Summary of Module 1 items ............................................................................ 36
Table 5. Summary of Module 3 items ............................................................................ 37
Table 6. Summary of RRB items from the ADI-R included in some analyses .............. 38
Table 7. Recommended guidelines for fit indices ............................................................. 43
Table 8. Module 1 indices of fit across models ............................................................... 49
Table 9. Indices of fit for Module 1 with addition of ADI-R RRB items ....................... 52
Table 10. Module 3 indices of fit across models .............................................................. 55
Table 11. Indices of fit for Module 3 with addition of ADI-R RRB items ....................... 59
List of Figures

Figure 1. One-factor model.................................................................45
Figure 2. Two-factor model..............................................................45
Figure 3. DSM-IV model.................................................................46
Figure 4. DSM-V model.................................................................46
Figure 5. DSM-IV bi-factor model....................................................47
CHAPTER 1

INTRODUCTION

Autism Spectrum Disorders (ASDs) are a group of neurodevelopmental disorders characterized by three main symptom clusters: impairments in reciprocal social behavior; impairments in communication; and the presence of restricted, repetitive, or stereotyped patterns of behavior or interests (APA, DSM-IV-TR, 2000). ASDs include autistic disorder (AD), Asperger’s disorder (AS), and pervasive developmental disorder-not otherwise specified (PDD-NOS). Current prevalence estimates for ASDs are about 60 individuals per 10,000 (Fombonne, 2005).

ASD Symptoms and Diagnostic Criteria

Historical background. Autism was first described by Leo Kanner (1943) who focused on two main domains: impaired social interaction, and various behaviors related to resistance to change and/or insistence on sameness. Rutter (1978) worked from this description and incorporated subsequent research when he defined autism as an early-onset disorder characterized by distinctive impairments in social and communication development and the presence of various behaviors related to insistence on sameness. This description was influential on the DSM-III diagnostic criteria (APA, 1980), the first version of the DSM to include autism as a diagnostic category. However, the DSM-III
category of “infantile autism” seemed to apply best to younger, more delayed children and its shortcomings led to significant changes in DSM-III-R (Volkmar, 1998).

Among the changes made to criteria in DSM-III-R (APA, 1987) were the elimination of early onset as a required criterion, and the organization of symptoms into three now-familiar domains of social interaction impairment, communication and play impairment, and restricted, repetitive interests and activities. Studies seemed to indicate that DSM-III-R criteria were over-inclusive (i.e., a high false-positive rate), leading to more changes in criteria with the advent of DSM-IV (APA, 1994).

Field trials for the development of DSM-IV criteria included 125 raters from 21 sites around the world. This revision attempted to address a number of problematic issues, including improving sensitivity and specificity and applicability to a range of levels of functioning. These 125 raters evaluated approximately 1,000 cases of autism or developmental disability according to their best-judgment diagnosis as well as DSM-III, DSM-III-R, ICD-10 (WHO, 1992), and newly proposed diagnostic criteria. Results indicated that DSM-III-R criteria were over-inclusive, with specificity being especially poor among severely impaired individuals. Recommended modifications included increasing the total number of criteria required to obtain a diagnosis of autism and including an age of onset before 3 years. Based on field trial data, these changes improved both sensitivity and specificity of the category. Asperger’s Syndrome, Rett’s Disorder, and Childhood disintegrative disorder were also included in DSM-IV, increasing consistency with the ICD-10 classification system.

Current criteria. In 2000, APA published the DSM-IV-TR, and changes to ASD categories included more descriptive text expanding on the features of AS, as well as
revision of PDD-NOS criteria to specify that individuals must have impairments in social interaction along with impairment in communication or restricted, repetitive and stereotyped behaviors (RRBs) to receive this diagnosis; the DSM-IV had permitted for individuals with impairment in only one domain to receive a PDD-NOS diagnosis. AD criteria remained unchanged between versions. According to DSM-IV/DSM-IV-TR, six symptoms total are needed in order to qualify for a diagnosis of AD, but two are required from the social impairments domain and only one each from the communication impairments and RRBs domains. The social domain is given more weight, prompting some to conceptualize AD and other ASDs as primarily social disorders.

By DSM-IV-TR criteria, social impairment symptoms include (a) impairment in the use of nonverbal behaviors (e.g., eye-to-eye gaze), (b) failure to develop developmentally appropriate relationships, (c) lack of seeking to share enjoyment, and (d) lack of social/emotional reciprocity. Communication impairment symptoms include (a) delay in or lack of spoken language and nonverbal compensatory strategies, (b) problems initiating or sustaining conversations, (c) use of stereotyped/repetitive/idiosyncratic language, and (d) lack of developmentally appropriate pretend or social imitative play. Finally, RRB symptoms include (a) abnormal preoccupation with a minimum of one stereotyped and restricted interest, (b) inflexible adherence to nonfunctional routines or rituals, (c) stereotyped and repetitive motor behaviors, and (d) persistent preoccupation with parts of objects. Delays in functioning must be present before three years of age, and the impairment must not be better described by Rett’s Disorder or Childhood Disintegrative Disorder, two other PDDs which are distinct from ASDs.
For a diagnosis of AS, individuals must not have significant language, cognitive, or life skill (e.g., self-help skills) delays with the exception of social interaction. They must have at least two symptoms of impaired social interaction and at least one symptom of RRBs as described above. PDD-NOS is a diagnosis based on exclusion. Individuals assigned this diagnosis should not meet criteria for any other PDDs, but must still demonstrate pervasive impairment in reciprocal social interaction and verbal/nonverbal communication skills, or present with RRBs. To meet criteria for PDD-NOS, one also should not qualify for diagnoses of Schizophrenia, Schizotypal Personality Disorder, or Avoidant Personality Disorder.

Recent studies have demonstrated that while AD is one of the most reliably diagnosed childhood disorders (Volkmar & Klin, 2005) the diagnosis of AS and PDD-NOS is more variable (Volkmar, Lord, Bailey, Schultz, & Klin, 2004). AD can be identified in children as young as two years of age (Cox, Klein, Charman, Baird, Baron-Cohen, Swettenham, et al., 1999) although sometimes ASD diagnoses are delayed. For instance, Wiggins, Baio, and Rice (2006) found a mean age of diagnosis of 5 years in a group of 108 youngsters with ASDs. Higher functioning individuals, or those with average or higher IQs, language skills, and/or adaptive behaviors, are often not recognized until even later (e.g., Sivberg, 2003). It is possible that the variability in PDD-NOS and AS diagnosis is due in large part to inconsistent interpretation and application of diagnostic criteria. In fact, Towbin (2005) outlined four different situations in which PDD-NOS is likely to be diagnosed: when detailed background is unavailable; when impairment in one of the core domains is mild or absent; when age of onset is after 3 years; and when other disorders are present but significant social impairments exist.
Likewise, Volkmar and colleagues (2004) highlighted five definitions of AS aside from the DSM and ICD criteria which have been put forth in the literature. Such diversity in application of criteria has complicated the study of ASDs and their structure.

*The future of criteria.* Conceptualization of ASDs continues to evolve. A new version of the DSM is scheduled to be published in May 2013. Significant changes have been proposed to PDD criteria. The new criteria have been published to allow for feedback from interested parties (APA, 2010). Under the proposed DSM-V criteria, AD, AS, PDD-NOS, and CDD would all be called Autism Spectrum Disorder, and the criteria would consist of three main areas. The first is impairment in social communication and interactions as evidenced by: (a) deficits in verbal and nonverbal communication used for social interaction, (b) lack of social reciprocity, and (c) failure to develop and maintain peer relationships appropriate to developmental level. All three of these symptoms must be present and clinically significant. The second area is restricted, repetitive patterns of behavior, interests, and activities evidenced by at least two of the following: (a) stereotyped motor or verbal behaviors or unusual sensory behaviors, (b) excessive adherence to routines and ritualized patterns of behavior, and (c) restricted, fixated interests. The third and final mandatory area is that symptoms must manifest in early childhood, though no specific age range is given, and it is noted that difficulties may not be fully evident until “social demands exceed limited capacities,” (APA, 2010).

The proposal of such significant changes is a reaction to the wealth of recent research examining ASDs and their structure. For instance, some studies have indicated that symptoms manifest differently with development and change with age (e.g., Eaves & Ho, 2004; Lord, Risi, DiLavore, Shulman, Thurm, & Pickles, 2006). Thus, current DSM
(and ICD) symptoms may not apply equally well across development; very young children (i.e., under age 3 years) may not have had the opportunity to develop peer relationships, so evaluating whether or not they have formed such relationships appropriate to their developmental level is difficult if not impossible. Lord and Corsello (2005) have suggested that current diagnostic criteria may be most applicable to school-aged children with autism and mild to moderate intellectual disability (ID).

Another area of interest is the validity of ASD subtypes. Witwer and Lecavalier (2008) reviewed the literature to examine the validity of autism, AS, and PDD-NOS as distinct categories. The included studies used either DSM-IV or ICD-10 criteria. Twenty-two studies were identified which examined differences between subtypes in clinical and demographic variables, neuropsychological profiles, comorbidity, and prognosis. Overall, results indicated that differences between groups were best explained by differences in IQ.

Numerous other studies also have pointed to the important role of IQ in ASD symptom expression. For instance, Munson, Dawson, Sterling, Beauchaine, Zhou, Koehler, and colleagues (2008) used latent class analysis and taxometric methods in a sample of 24 to 66 month-olds (n = 456). Both methods yielded evidence of discrete subgroups within ASDs based on IQ, suggesting that IQ may be a more reliable means of subtyping ASDs than the current system (Munson et al., 2008). Still other research has shown that symptom severity and expression vary with level of functioning (e.g., Dawson, Estes, Munson, Schellenberg, Bernier, & Abbott, 2007; Joseph, Tager-Flusberg, and Lord, 2002; Szatmari, Georgiades, Bryson, Zwaigenbaum, Roberts, Mahoney, et al., 2006). In addition to providing further reason to revisit DSM criteria, these studies also
point to the importance of homogenous subgroups for studying ASDs. Specifically, they highlight the need to examine the impact of level of functioning (e.g., low IQ may indicate different etiologies).

A significant challenge facing any diagnostic system (including the DSM-V) is the heterogeneity of autism presentation. Severity and range of symptoms vary considerably from one individual to the next (Szatmari, Mérette, Bryson, Thivierge, Roy, Cayer, et al., 2002) making it difficult to capture ASDs succinctly with one set of criteria. Individuals with ASDs also present with a wide range of language and cognitive abilities (Lecavalier, Snow, and Norris, in press). Furthermore, many other disorders have symptoms similar to those of ASDs. For instance, ID is defined as cognitive functioning approximately two standard deviations below the population mean with concurrent deficits in adaptive behavior, and can be difficult to differentiate from ASDs, particularly at lower levels of functioning (Lecavalier et al., in press). Language disorders also may present similarly to ASDs, making accurate identification difficult (e.g., Lord, 1995; Ventola, Kleinman, Pandey, Wilson, Esser, Boorstein, et al., 2007).

The “gold standard” for diagnosis of ASDs includes clinician consensus following a multidisciplinary assessment; this assessment often includes the Autism Diagnostic Interview- Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994) and the Autism Diagnostic Observation Schedule (ADOS; Lord, Rutter, DiLavore, & Risi, 1999). These are distinct instruments, each with advantages and disadvantages. Research has indicated that optimal sensitivity and specificity are achieved when they are used together (Risi, Lord, Gotham, Corsello, Chrysler, Szatmari, et al., 2006).
The ADI-R is a semi-structured, clinician-administered interview and includes a range of behaviors/symptoms from each of the three DSM domains of impairment. It covers past and current behavior. However, it is lengthy (lasting 2 or more hours), requires training to administer, and consequently is costly. It also relies to an extent on parental memory, which can be biased as more time passes.

By contrast, the ADOS is a semi-structured observation which allows clinicians to elicit (or not) different behaviors. It consists of four different but similar modules appropriate for different verbal levels and ages. Module 1 is designed for those without consistent use of phrased speech; Module 2 is for those with some phrase speech; Module 3 is for children who are both verbally fluent and for whom toy play is considered age-appropriate; Module 4 is for verbally fluent adolescents and adults. Most ADOS items are coded from zero (0, no evidence of abnormality) to three (3, markedly abnormal behavior). A code of seven (7) is used when the behavior is abnormal but in a different way than the protocol specifies and eight (8) is used when the rating is not applicable.

Ratings are made during/immediately after administration. It is the most well-studied observation instrument available, but it has relatively few items related to repetitive and stereotyped behaviors (as the relatively brief administration period may not permit observation of such behaviors), and does not address early history. It takes substantially less time than the ADI-R to administer, but also requires specialized training.

Structure of the ASD Phenotype

In recent years, many different types of studies (including genetic/family, longitudinal, and factor analytic endeavors) have aimed to elucidate the ASD phenotype. One reason the structure of ASDs has been examined is that the three-domain system of
the DSM/ICD is based largely on clinical judgment, and empirical examination of this model has yielded conflicting results, as detailed below (e.g., Constantino, Gruber, Davis, Hayes, Passanante, & Przybeck, 2004; Georgiades, Szatmari, Zwaigenbaum, Duku, Bryson, Roberts, et al., 2007). As already discussed, validity of the DSM subtypes has also been questioned (e.g., Witwer & Lecavalier, 2008) leading researchers to seek better ways to classify individuals with ASDs into homogenous subgroups (e.g., on the basis of IQ or specific ASD symptoms).

Clarifying the structure of ASDs is important for many reasons. First, classification systems (like the DSM) are periodically revised to reflect advances in research. Having a better understanding of the ASD symptom structure aides in the revision of these systems, which then hopefully leads to more valid diagnoses. Refined diagnosis is important at the individual level to ensure those in need of services and treatment have access to them, and to facilitate communication between professionals. Furthermore, diagnostic instruments (like the ADI-R and ADOS) are often based on classification symptoms, and as diagnostic categories are more reliably described, instruments also may be refined. This may be one of the reasons so few instruments are available to assess PDD-NOS; its definition per current classification systems is imprecise.

Broader implications of improved diagnosis relate to etiology. As classification systems are refined and diagnoses become more accurate, studies can more precisely examine the genetic underpinnings of ASDs. Indeed, more homogenous phenotypes (like those based on language or insistence on sameness) have proven helpful in linkage studies (Abrahams & Geschwind, 2008). In fact, Hus, Pickles, Cook, Jr, Risi, and Lord
(2007) summarized several genetic studies which have used various symptom profiles (e.g., age of first words of phrases, insistence on sameness) to examine ASD samples with some success. For example, Shao, Cuccaro, Hauser, Raiford, Menold, Wolpert, et al. (2003) found that individuals with high insistence on sameness scores had increased logarithm of the odds (LOD) scores for the 15q11-q13 chromosome region. Such research also has implications for understanding the pathogenesis of ASDs. Research indicates that quantitative phenotypes may be more reliable and therefore more informative than categories (Szatmari, Maziade, Zwaigenbaum, Mérette, Roy, Joober, et al., 2007; Szatmari, White, & Marikangas, 2007). Several researchers have pointed to the lack of well-characterized, empirical ASD phenotypes as an obstacle to pathogenic research (Skuse, 2007; Szatmari, White, Marikangas, 2007). This line of research is especially timely, given the impending DSM-V revision.

**Relationship between domains.** One way in which scientists have attempted to elucidate the structure of autism and its symptoms is by examining the relationship (or lack thereof) between the current triad of behavioral domains found in the DSM. Overall, studies seem to indicate that there is a weak to moderate relationship between domains (e.g., Happé, Ronald, & Plomin, 2006; Mandy & Skuse, 2008; Snow, Lecavalier, & Houts, 2008). This is true even of the social and communication domains, which are often thought to be closely related, if not inseparable (APA, 2010). For instance, Mandy and Skuse conducted a review of available literature to determine the relationship between the social and communication domains of ASDs and the restricted, repetitive behavior/interests domain. They found only three studies which directly addressed this question, the results of which were inconclusive. Two of the three studies (Ronald,
Happé, Bolton, Butcher, Price, Wheelwright, et al., 2006; Ronald, Happé, & Plomin, 2005) found at most only modest correlations between domains. Ronald et al. (2005) examined questionnaire data completed by parents and teachers on a large community sample (i.e., not ASD-specific) of twins and found unimpressive associations between social-communication impairment and RRBs ($r = .15$ from teacher report, $n = 3,090$; $r = .29$ from parent report, $n = 3,996$). However, methodological limitations of this study (including the instrument used) tempered impact of the findings. Using the same sample but a more valid and reliable instrument, Ronald et al. (2006) found similar results, with correlations between domains as follows: communication and social $r = .34$, communication and RRBs $r = .38$, social and RRBs $r = .23$.

The third study Mandy and Skuse (2008) reviewed was one of the first efforts to examine association between ASD domains. Wing and Gould (1979) found notable association in their sample of children with either severe cognitive impairment but no social impairment ($n = 58$) or severe cognitive and social impairment ($n = 74$). Results indicated all individuals in the socially impaired group also exhibited RRBs, and most also exhibited language and symbolic play abnormalities (e.g., 74% of the socially impaired group did not engage in symbolic play). Overall, this study supported the hypothesis that the theoretical domains of autism were highly correlated, but generalizability was limited due to a possible sampling bias. It is also difficult to compare these results with more recent studies due to the unclear definitions of “social impairment” as this study was conducted prior to DSM-III and the development of standardized definitions of autism.
Additionally, some researchers have pointed to the apparent different developmental trajectories of the three core symptom domains as evidence of their relative independence (Happé, Ronald, and Plomin, 2006). For example, Charman, Taylor, Drew, Cockerill, Brown and Baird (2005) followed 26 children diagnosed with autism at age 2 years through assessments at ages 3 and 7 years. An additional data point was obtained by asking questions about ages 4 to 5 at the 7-year assessment (resulting in data for age 3, age 4-5, and age 7 years). They found that impairments in social interaction tended to become less severe between aged 4 to 5 and 7 years, communication symptoms decreased at each time point, and report of RRBs increased from the age 3 assessment to the 4 to 5 year report, then decreased at age 7 assessment. That is, the three DSM domains as measured by the ADI-R demonstrated different patterns over time.

Piven, Harper, Palmer, and Arndt (1996) examined autistic symptoms of thirty-eight 13-to 28-year-olds at their current age, and retrospectively at age 5 years. Generally, there was a trend for symptoms to improve with age, though there was a difference between social/communication symptoms and RRBs. The number of participants exhibiting social and communication symptom improvement was significantly higher than the number exhibiting improvement on RRBs. While providing evidence that the domains of ASD symptoms may be distinct, these studies also have implications for sample composition of studies examining ASD structure. Since ASD symptom patterns change with age, differences in ASD structure may be missed if preschool-aged children, school-aged children, adolescents, and adults are all included in the same analyses. This underlines the importance of examining homogenous subgroups in terms of age (e.g., ASD may present different in a 5-year-old than it does in a 15-year-old).
Factor analytic and principal component studies. A different but related vein of research has relied on factor analysis (FA) in an attempt to uncover underlying constructs and compare possible symptom models. Though some recent work has supported the current three-domain DSM system compared to other models (e.g., Lecavalier, Gadow, DeVincent, Houts, & Edwards, 2009) other research has indicated that the current classification system may not adequately describe the structure of ASD symptoms. Below, the evidence supporting various models (particularly two- and three-factor models) is examined.

Two-factor/component models. Two factor models have been proposed in three studies. Snow, Lecavalier, and Houts (2009) submitted the ADI-R items to exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) in a sample of 1,861 individuals from multiplex families aged 4 to 18 years. Analyses were performed separately for verbal ($n = 1,329$) and nonverbal ($n = 532$) samples. In addition, separate analyses were conducted with all items or only algorithm items. Results supported a two-factor structure of autism with a social/communication factor and a restricted/repetitive behaviors factor. Ranges for fit indices for the two-factor solution in different analyses were as follows: Root Mean Square Error of Approximations (RMSEAs) from .026-.068, Comparative Fit Indexes (CFIs) from .95-.99, Standardized Root Mean Square Residuals (SRMRs) from .077-.089, and Goodness of Fit Indexes (GFIs) from .94-.97. Interestingly, the authors chose the two-factor model, but the three-factor model had virtually identical indices of fit.

Using data from the same database, Frazier, Youngstrom, Kubu, Sinclair, and Rezai (2008) found similar results in a sample of 1,170 verbal individuals aged 2 to 46
years. ADI-R *domain scores* were submitted to analyses. EFA results supported a two-factor model, with a social/communication factor and a stereotyped behavior factor (which also included stereotyped language). CFAs supported both this two-factor model as well as a three-factor model, which separated difficulties in peer relationships and imaginative play from the social/communication factor. However, in this three-factor model, the peer relationships/imaginative play factor only contained two items. Fit indices for both the three-factor and two-factor models were as follows: RMSEA = .07, Non-Normed Fit Index (NNFI) = .92, and CFI = .94. Analyses examining factorial invariance across age (subgroups of 2 to 6 years, and greater than 7 years) for these two models indicated no significant differences in factor loadings or in model fit.

Mick (2005) performed principal components analysis (PCA) on the 12 ADOS Module 1 algorithm items (*n* = 220). Communalities generally were high (.43-.71). Results indicated that two components best fit the data: Social Interaction (6 items) and Impaired Communication (5 items). These components exactly matched the scoring algorithm domains, except for one item (“*Gestures*”), which did not load above the meaningful cut-off (.40) on either component. One item (“*Unusual eye contact*”) had a loading of 1.0 on the first component.

*Three-factor/component models.* Seven studies have indicated that three-factor models may best describe the ASD phenotype. In contrast to findings on Module 1 data, Mick (2005) examined the twelve ADOS Module 2 algorithm items (*n* = 100), and analyses supported a 3-component solution: Social Communication (6 items), Social Language (4 items), and Joint Attention (2 items). The final component included the
items “Pointing” and “Spontaneous initiation of joint attention.” Examination of Module 2 was performed with a relatively small sample size.

Lecavalier, Aman, Scahill, McDougle, McCracken, Vitiello, et al. (2006) performed an EFA on ADI-R items from a sample of 226 individuals with ASDs aged 5 to 17 years. Results supported a three-factor solution similar to the ADI-R algorithm, except all of the nonverbal communication items in the analysis loaded onto the social factor; the other two factors were communication and RRB. Georgiades, Szatmari, Zwaigenbaum, Duku, Bryson, Roberts, et al. (2007) submitted ADI-R domains to PCA in a sample of 209 individuals aged two to forty years. They found a three-component solution was preferable over two-, four-, and five component solutions. This three-component solution was only somewhat similar to the DSM model. Components were named: social-communication (6 items), inflexible language and behavior (4 items), and repetitive sensory and motor behavior (2 items). This model was compared to the DSM model via CFA with a separate sample (n not specified) and results supported the proposed model. For this model, RMSEA was .067, NNFI = .9, CFI = .92, and SRMR = .08, generally indicating acceptable fit. For the DSM-IV model, RMSEA = .12, NNFI = .59, CFI = .69, and SRMR = .012, indicating poor fit.

Similar to Georgiades et al. (2007), van Lang, Boomsma, Stema, de Bildt, Kraiger, Ketelaars, et al. (2006) reported a different model which fit their data better than the DSM-IV three-domain model. Van Lang et al. (2006) submitted ADI-R domains obtained from 255 individuals aged 4 to 20 years with either an ASD diagnosis or social/communication problems to CFA. Results supported a three-factor solution with factors named: impaired social communication (5 items), impaired make-believe and play (2
items), and stereotyped language and behavior (5 items). This model was preferred over
the DSM-IV model (based on either current scores or scores from age 4 to 5), a one-
factor model, and a two-factor model which consisted of a social/communication factor
and a RRB factor. For the proposed model, fit indices were as follows: RMSEA = .03,
NNFI = .98, SRMR = .05. Analysis performed with only lower-functioning individuals (n
= 234) found no substantial differences in model fit compared to the total sample, though
it should be noted that the samples differed by only 21 participants.

Boomsma, van Lang, Jonge, de Bilt, van Engeland, and Minderaa (2008)
replicated these findings in an independent sample of children and adolescents with ASD
aged 4 to 24 years (n = 263). Again, the proposed model was compared to both a one-
factor and two-factor model, and the proposed model was preferable. Fit indices for the
proposed model based on ADI-R current and 4 to 5 year scores were as follows:
RMSEAs = .05-.07, NNFIs = .91-.96, CFIs = .93-.97, and SRMRs = .05-.06. Multiple
group analyses of this model were performed to examine the impact of age, IQ, and ASD
subtype on model fit, and the investigators found that there were no large differences.

Lecavalier et al. (2009) subjected data from the PDD sections of the Early
Childhood Inventory, Fourth Edition (ECI-4; Gadow & Sprafkin, 2000; designed for ages
3 to 5 years) and the Child Symptom Inventory, Fourth Edition (CSI-4; Gadow &
Sprafkin, 2002; designed for ages 6 to 12 years) to CFA. Analyses were conducted on the
entire sample (n = 730) and on individual subgroups in order to examine the impact of
subject characteristics (i.e., age and ASD subtype). One-, two-, and three-factor models
were tested, and the three-factor model (representing the DSM domains) was supported
for both parent and teacher ratings. DSM model fits for the parent data were as follows:
RMSEA = .088, CFI = .949, GFI = .98, and SRMR = .066. Fits for the teacher data were as follows: RMSEA = .073, CFI = .98, GFI = .99, and SRMR = .051. Including only individuals with IQ > 70 improved fit with school-aged children with parent data but not with teacher data. The best fit (for both parent and teacher data) was found for the subgroup with Asperger’s disorder. Overall, these results supported the DSM-IV model and indicated rater, level of functioning, and ASD subtype impacted fit.

Robertson, Tanguay, L’Ecuyer, Sims, and Waltrip (1999) performed an EFA of the 13 social communication scoring algorithm items of the PL-ADOS with a sample of infants and children aged 36 months to 6 years (n = 51). They reported that a 3-factor solution best fit the data: Factor 1 (8 items) was named Joint Attention, Factor 2 (4 items) was named Affective Reciprocity, and Factor 3 (1 item) was named Theory of Mind. All items loaded strongly (i.e., >.55) onto their respective factors. It is noteworthy that this study used a relatively young sample, a small sample size, and an earlier version of the ADOS, so results should be interpreted with caution.

Other Models. There is also limited support in the literature for other models. Constantino et al. (2004) examined the structure of autistic traits in a sample of individuals aged 4 to 18 years with and without ASDs. SRS items from parent report (n = 91) and teacher-report (n = 91) and then a larger sample of 168 subjects (with 245 total test administrations) were submitted to PCA. Results from all analyses supported the idea of one component underlying ASD symptoms. Still other studies propose models with more than three domains. For instance, Tadevosyan-Leyfer, Dowd, Mankoski, Winklosky, Putnam, McGrath, et al. (2003) performed a PCA on ADI/ADI-R data in a sample of 292 individuals with autism aged 2 to 47 years. Six components emerged from
analyses: spoken language, social intent, compulsions, developmental milestones, savant skills, and sensory aversions.

A summary of studies examining the structure of ASD symptoms can be found in Table 1. Overall, these studies largely support models composed of either two or three domains for the structure of ASD symptoms. There was variability in not only the number of factors, but also the composition of factors. Results may differ for several reasons, including the instrument used, type of FA procedures employed, and composition of the sample.
<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Instrument(s)</th>
<th>No. factors</th>
<th>Names</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lecavalier et al. (2009)</td>
<td>730</td>
<td>ECI-4, CSI-4 (all items)</td>
<td>3</td>
<td>Social Communication RRB</td>
<td>CFA; included 3-12 year-olds with PDD</td>
</tr>
<tr>
<td>Snow et al. (2009)</td>
<td>1,861</td>
<td>ADI-R</td>
<td>2</td>
<td>Social/Comm RRB</td>
<td>EFA, CFA; 4-18 year-olds with PDD; also supported 3-factor DSM-IV model</td>
</tr>
<tr>
<td>Boomsma et al. (2008)</td>
<td>263</td>
<td>ADI-R (subdomains)</td>
<td>3</td>
<td>Social/Comm Stereotyped Speech/Beh Impaired Play Skills</td>
<td>CFA; 4 to 24 year-olds with PDD</td>
</tr>
<tr>
<td>Frazier et al. (2008)</td>
<td>1,170</td>
<td>ADI-R (subdomains)</td>
<td>2</td>
<td>Social/Comm RRB</td>
<td>EFA, CFA; verbal 2 to 46 year-olds; also supported 3-factor model</td>
</tr>
<tr>
<td>Georgiades et al. (2007)</td>
<td>209</td>
<td>ADI-R (subdomains)</td>
<td>3</td>
<td>Social/Comm Inflexible Lang/Beh RRB</td>
<td>PCA; 2 to 40 year-olds with PDD; CFA in separate sample supported this over DSM-IV model</td>
</tr>
<tr>
<td>Lecavalier et al. (2006)</td>
<td>226</td>
<td>ADI-R (algorithm items)</td>
<td>3</td>
<td>Social Communication RRB</td>
<td>EFA; 5 to 17 year olds with PDD</td>
</tr>
<tr>
<td>Van Lang et al. (2006)</td>
<td>255</td>
<td>ADI-R (subdomains)</td>
<td>3</td>
<td>Social/Comm Stereotyped Speech/Beh Impaired Play Skills</td>
<td>CFA; 4 to 20 year-olds; PDD and non-PDD</td>
</tr>
<tr>
<td>Mick (2005)</td>
<td>220</td>
<td>ADOS, Module 1 (algorithm items)</td>
<td>2</td>
<td>Social Communication</td>
<td>PCA; 17 to 71 month-olds; PDD and non-PDD</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>ADOS, Module 2 (algorithm items)</td>
<td>3</td>
<td>Social/Comm</td>
<td></td>
</tr>
<tr>
<td>Constantino et al. (2004)</td>
<td>168</td>
<td>SRS (all items)</td>
<td>1</td>
<td>ASD</td>
<td>PCA; 245 parent, 91 teacher ratings of 168 total subjects aged 4-18 years</td>
</tr>
<tr>
<td>Tadevosyan-Leyfer et al. (2003)</td>
<td>292</td>
<td>ADI, ADI-R (98 items common between both versions)</td>
<td>6</td>
<td>Spoken Language Social Intent Compulsions Dev Milestones Savant Skills Sensory Aversions</td>
<td>PCA; 2 to 47 year-olds with PDD</td>
</tr>
<tr>
<td>Robertson et al. (1999)</td>
<td>51</td>
<td>PL-ADOS (social-communication algorithm items)</td>
<td>3</td>
<td>Joint Attention Affective Reciprocity Theory of Mind</td>
<td>EFA; 3 to 6 year-olds with PDD</td>
</tr>
</tbody>
</table>

*Comm = Communication; Lang = Language, Beh = Behavior, Dev = Developmental

Table 1: Summary of studies of ASD structure.
A majority of studies (7/11 studies) used data from the ADI or ADI-R. This is important because the instrument used impacts results. For example, different reporters can impact model fit (e.g., Lecavalier et al., 2009) so data from a parent-report measure may suggest different results than data from a clinician-rated instrument. Nine of the reported studies used data from either a clinician-rated interview (based on parent report) or a parent-rated checklist. Only two studies used an instrument based on clinician observation (Mick, 2005; Robertson et al., 1999) and these supported different models.

Procedural variations can also impact results (Norris & Lecavalier, 2010a). Four of eleven studies (36%) used PCA, two (18%) used EFA only, three (27%) used CFA only, and two (36%) used both EFA and CFA. CFA allows for the testing of models, while EFA is more often used to generate hypotheses, and PCA is a data reduction technique. Method of estimation (e.g., maximum likelihood versus principal axis factoring) and rotation (i.e., orthogonal or oblique) also can impact results. Using either all items or item domains can also affect results.

It is also important to consider sample characteristics when reviewing studies. Several studies included a wide age range of participants, but only two of these examined the impact of different age groups on model fit (i.e., Boomsma et al., 2008; Frazier et al., 2008). One study included only verbal individuals (Frazier et al., 2008) while another analyzed data separately for verbal and nonverbal participants (Snow et al., 2009). Three studies examined fits of models within homogenous subgroups based on level of functioning (Boomsma et al., 2008; Lecavalier et al., 2009; van Lang et al., 2006) with mixed results.
Generally, studies using data (either item or domain level) from the ADI-R seemed to support either two- and three-domain structures. Both studies using ADI-R data which supported two-factor models (Frazier et al., 2008; Snow et al., 2009) also somewhat supported three-factor models. Both the DSM three-factor model and a three-factor model with factors related to general social/communication impairments, repetitive/stereotyped language and behavior, and impairments in joint attention/imaginative play seemed to have some support in the literature.

The majority of factor analytic studies examining the structure of ASDs have employed ADI-R data. This can be viewed as a positive given the ADI-R is a well-studied, standardized instrument. However, since the nature of input data obviously impacts results, it is essential to use other instruments to address the structure of ASDs as well. Although numerous instruments have been developed to assess ASD symptoms, few have been rigorously studied (Norris & Lecavalier, 2010b). In addition to the ADI-R, one well-established measure with strong psychometric support is the ADOS. As the ADOS is the primary focus of this study, it will be further described.

**ADOS: Development and Initial Validation.** The current form of the ADOS resulted from merging the previous version of the ADOS (Lord, Rutter, Goode, Heemsbergen, Jordan, Mawhood, et al., 1989) and a version designed for children with very limited to no language skills, the Pre-Linguistic ADOS (or PL-ADOS; DiLavore, Lord, and Rutter, 1995). The current version was developed with 381 individuals referred for evaluation due to suspicion of an ASD. Approximately 20-30 individuals were selected from this larger pool for reliability of each module, and included those with autism, PDD-NOS, and non-spectrum diagnoses. For validity analyses, matched groups
on verbal mental age (module 1-3) or VIQ (module 4) were created. Sample sizes for validity analyses were as follows: module 1 \((n = 54)\); module 2 \((n = 55)\); module 3 \((n = 59)\); and module 4 \((n = 45)\). Interrater reliability (IRR) of the final set of items exceeded 80% agreement for all modules. Items with poor IRR were eliminated or rewritten. Intraclass correlation coefficients (ICCs) were computed for 27 individuals administered the same module twice within an average of 9 months in order to assess test-retest reliability. ICCs ranged from .59 (Restricted, Repetitive domain) to .82 (Social and Communication domains combined). Internal consistency assessed by alpha for modules was highest for the Social domain (range of 86-.91), followed by the Communication domain (.74-.84), with the Stereotyped Behaviors and Repetitive Interests domain being lowest (.63–.65 for Modules 2 and 1; .47–.56 for Modules 4 and 3, respectively).

Exploratory factor analyses (EFAs) were conducted for each module, and results indicated that one factor explained a majority of the variance. It consisted of almost all social and communication items which were included in analyses. Items dealing with gestures and various forms of speech loaded onto a second factor, but had high cross-loadings (i.e., above .30) on the first factor. Stereotyped behaviors and repetitive interest items loaded onto separate factors which varied across modules. Ultimately, the authors decided to separate social, communication, and RRBs into individual domains within the scoring algorithm.

ANOVA\s were used to examine items’ ability to differentiate individuals with autism from those not on the spectrum; some items were retained despite not meeting this criterion if they were diagnostically important for particular subgroups (e.g., low functioning children) or if they provided other valuable information (e.g., anxiety items
were included as anxiety-related behaviors could impact testing performance). Receiver operating characteristic (ROC) curves were calculated to identify optimal cut-off values for identifying ASD compared to non-ASD, autism compared to not autism (PDD-NOS and non-spectrum), PDD-NOS compared to non-spectrum, and autism versus non-spectrum. Sensitivity and specificity values varied by module and comparison groups, but overall were high, supporting the instrument’s diagnostic validity. Module 1 sensitivity and specificity ranged from .94-1.00 and .79-1.00, respectively; module 2 ranged from .89-.95 for sensitivity and .73-.94 for specificity; module 3 sensitivities ranged from .80-1.00 and specificities ranged from .68-1.00; module 4 sensitivity and specificity ranged from .86-.93 and .76-.93, respectively. Values tended to be lowest for differentiating autism from non-autism.

**ADOS: Additional Research.** Numerous additional studies have examined the ADOS, particularly its diagnostic and convergent validity. Overall, studies have found it has sound psychometric properties. In 2004, de Bilt, Sytema, Ketelaars, Kraijer, Mulder, Volkmar, et al. compared the ADOS to the ADI-R and consensus diagnosis in a sample of 184 individuals aged 5 to 20 years with ID. Based on DSM-IV-TR criteria with team consensus, 48 individuals were diagnosed with autism, 47 with PDD-NOS, and 89 were non-spectrum. For those with autism, the ADOS was more sensitive (.87) than the ADI-R (.72) but the ADI-R was much more specific (.79 compared to .47). Among those with PDD-NOS, the ADOS out-performed the ADI-R and had high sensitivity (.92) and fair specificity (.65). Logistic regression indicated the ADOS was more accurate with increasing age and decreasing level of functioning. Overall agreement between the
instruments was good (64%) but when controlling for chance, agreement values plummeted \( (k = .27) \) for total sample, .29 for autism and PDD-NOS.

Risi et al. (2006) found similar results in a sample of 1,039 individuals aged 14 months to 18 years. Correlation between ADOS and ADI-R total scores was .57, but both instruments performed best when used together. Poorer agreement was found between the ADI-R and ADOS in a sample of 6 to 11 year olds (Bishop & Norbury, 2002), which the authors attributed at least partly to the different formats of the instruments. Mazefsky and Oswald (2006) compared the ADOS, ADIR, and Gilliam Autism Rating Scale (GARS; Gilliam, 1995) in a sample of children aged 22 months to 8 years \( (n = 78) \). Both the ADOS and ADI-R performed well, with the ADOS correctly classifying 77% of children and the ADI-R correctly classifying 73%, while the GARS was unable to differentiate between those with an ASD \( (n = 59) \) and those without an ASD \( (n = 19) \).

Other researchers have focused on examining the ADOS in younger samples. Gray, Tonge, and Sweeney (2008) compared the ADOS to the ADI-R in a sample of 209 children aged 20 to 55 months suspected of having developmental delays. After diagnostic evaluation, 120 children were diagnosed with autism, 23 with PDD-NOS, and 66 with developmental delay and/or language impairment. Children with autism scored significantly higher than those without autism on all module domains, even after controlling for developmental age. ADOS classification showed good agreement with consensus clinical diagnosis of autism \( (k = .73, p < .001) \) and ASD \( (k = .62, p < .001) \), out-performing the ADI-R in this young sample. Sensitivity and specificity of the ADOS diagnostic algorithms were .85 and .89, respectively. Similarly, Le Couteur, Haden, Hammal, and McConachie (2008) compared these two instruments in 101 children aged
24 to 49 months and found good agreement between the instruments ($k = .62$ for autism cut-off, .54 for ASD cut-off).

Noterdaeme, Sitter, Mildenberger, and Amorosa (2000) assessed diagnostic validity of the ADOS in a sample of 49 children with ICD-10 diagnoses of infantile autism ($n = 11$), receptive language disorder ($n = 20$), or expressive language disorder ($n = 18$). Results indicated significant differences on subscale scores between the autism group and the language disorder groups, but few differences were found between the two language disorder groups. Eight of 11 children with autism were correctly classified by the social impairment and communication impairment algorithms, while none of the language disordered children met ASD cut-offs on both these domains, indicating very high specificity and good sensitivity. Noterdaeme, Mildenberger, Sitter, and Amorosa (2002) found similar results in a sample of 11 children with childhood autism and 16 with severe receptive language disorder.

Gotham, Risi, Pickles, and Lord (2007) performed both EFA and CFA on Modules 1 to 3 with children aged 14 months to 16 years in order to improve the diagnostic validity of the scoring algorithms. The sample included 1,630 ratings based on 1,139 individuals. Most participants were diagnosed with autism (56% of sample), though non-spectrum cases (17%) were also included. Participants were divided into groups based on language level in Module 1 (no words or some words) and age (< 5 years or 5 to 15 years) in Module 2, with Module 3 not being subdivided. Subsample sizes ranged from 137 to 495 participants. Social and communication items were included in analyses if no more than 20% of individuals with autism scored a “0” on the item, and no more than 20% of non-autistic individuals scored a “2” or “3” on the item.
Two additional items considered theoretically important were also included (“gestures” in Module 3 and “shared enjoyment” in Modules 2 and 3). RRB items did not need to meet these criteria. Consequently, 14 items were included in analyses. Results of the EFA indicated that a two factor solution was a good fit for all groups; factors were named Social Affect and Restricted, Repetitive Behaviors. Confirmatory analyses yielded RMSEAs ranging from .08 (Module 3) to .09 (Module 1), which indicate marginally acceptable to poor fit according to guidelines (Browne & Cudeck, 1993). Factors again were named Social Affect (10 items), and Restricted, Repetitive Behaviors (4 items). A three-factor solution was examined, with the third factor named “Joint Attention,” which overlapped with the first factor. This solution yielded acceptable fit indices; however, it was abandoned in favor of the more parsimonious model which performed best among all cells.

In a companion study, Gotham, Risi, Dawson, Tager-Flusberg, Joseph, Carter, et al. (2008) replicated the revised algorithms in a sample of children with Autism, PDD-NOS, or non-spectrum disorders aged 19 months to 16 years (total $N = 1,282$). Participants were divided into subsamples according to the same cells as in Gotham et al. (2007); subsample sizes ranged from 53 to 339 participants. The Module 2, older children ($n = 100$) were excluded from the factor analysis due to limited distribution across diagnostic groups. Results of EFA with 14 items supported the two-factor solution, with RMSEAs ranging from .05 (Module 1, no words) to .08 (Module 2, <5 years). The Module 2, <5 years solution contained a different pattern of loadings from what Gotham et al. (2007) found, but this could be explained by the low communalities of items and
low subject-to-item ratio. Analyses on only ASD cases ($n = 1,068$) yielded similar results.

Overall, studies of the ADOS have found good sensitivity, fair specificity, and good agreement with other instruments designed to identify ASDs. Little independent research has examined its reliability, although internal consistency, interrater reliability, and test-retest values were good in the development sample. To date, most of the work using the ADOS has focused on its diagnostic validity. Given its demonstrated validity it can be a valuable tool for studying the autism phenotype. Yet only two studies have used it for this purpose (Mick, 2005; Robertson et al., 1999) and one of these used an older version of the ADOS and a very small sample.

**Study Objective**

The current study examined the structure of ASD symptoms as measured primarily by the ADOS. This study expanded on previous research by examining the structure of ASDs according to an observation-based measure of current behaviors, and by examining several models of ASD symptoms within the same sample. It was also the first to incorporate data from multiple sources to this end. Specifically, a separate set of analyses was conducted using both the ADOS items as well as items from the ADI-R RRB domain (current scores). This is important because these symptoms are a core ASD symptom domain, and therefore important to assess, but the ADOS has relatively few items tapping these behaviors due to the limited amount of time those administering the ADOS spend with the individual. In order to ensure this symptom domain was adequately sampled for analyses, it was necessary to gather additional data on these behaviors from another source (in this case, the ADI-R).
Only two studies have attempted to examine the structure of ASDs using the ADOS, but sample sizes were small, homogeneous subgroups were not examined, and subscales were submitted to analyses (as opposed to all items). This study is unique in that it included all ADOS items related to DSM domains. This is important as including only algorithm items presupposes a structure on the data, limiting interpretability of results. Additionally, this study used data from a standardized instrument and a large sample. Hence, this study is capable of providing valuable insight into how the field conceptualizes and diagnoses what is currently a heterogeneous group of disorders.

Due to the possible impact of sample characteristics, special attention was paid to how different subgroups impacted model fit. While the ADOS controls for level of functioning and age to some extent through appropriate module selection, previous studies using the ADOS used heterogeneous groups in terms of language and IQ. Previous studies have demonstrated that ASD symptoms may manifest differently as a function of IQ and age; this provided the rationale for selecting the variables in this study.

Based on previous research, it was expected that data would support a three-factor model (i.e., the DSM-IV). It also was expected that models fit would improve within the different subgroups based on age and level of functioning.
CHAPTER 2

METHOD

Participants

Participant data were collected from a large database available for research purposes, the Autism Genetic Resource Exchange (AGRE; Geschwind, Sowinski, Lord, Iversen, Shestack, et al., 2001). The majority of children in the AGRE database are from families in which at least two members have an ASD diagnosis (i.e., multiplex families). Participants with an ASD diagnosis are referred to AGRE; diagnoses are then confirmed via team consensus, based in part on both ADI-R and ADOS scores. Individuals are assigned to one of four diagnostic categories: Autism, Not Quite Autism (NQA, those within one point of domain autism cut-offs on the ADI-R), Broad Spectrum (more than one point away from domain cut-offs on the ADI-R, but still determined to fall on the spectrum), or Not Spectrum. Additional data are collected on cognitive and adaptive functioning, vocabulary, and family history. Genetic material also is collected via blood draws. It is important to note that due to the nature of the database, not all data are available for all participants. Individuals were excluded if referred with an ASD but categorized as Not Spectrum by AGRE researchers; all other individuals were included if they met spectrum cut-offs on the ADOS and had complete ADOS and ADI-R data.
Children and adolescents aged 3 to 18 years were included in analyses. Due to the relatively small number of individuals with Module 2 scores and the size of subgroups which would have resulted from further dividing this module into smaller cells, it was decided to examine only Modules 1 and 3. There was a total of 1,409 participants, with 720 individuals for Module 1 and 689 individuals for Module 3. Module 1 sample characteristics are reported in Table 2.

<table>
<thead>
<tr>
<th>Module 1</th>
<th>Total (N=720)</th>
<th>Under 6 (n=290)</th>
<th>AB &lt; 55 (n=339)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age years (M, SD)</td>
<td>7.2 (3.1)</td>
<td>4.5 (.84)</td>
<td>7.9 (3.1)</td>
</tr>
<tr>
<td>Gender (males) n (%)</td>
<td>561 (78%)</td>
<td>224 (77%)</td>
<td>266 (79%)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>n = 605</td>
<td>n = 237</td>
<td>n = 288</td>
</tr>
<tr>
<td>Non-Hispanic/Latino</td>
<td>480 (79%)</td>
<td>184 (78%)</td>
<td>234 (81%)</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>125 (21%)</td>
<td>53 (22%)</td>
<td>54 (19%)</td>
</tr>
<tr>
<td>Diagnostic category</td>
<td>n = 636</td>
<td>n = 252</td>
<td>n = 303</td>
</tr>
<tr>
<td>Autism n (%)</td>
<td>611 (96%)</td>
<td>238 (94%)</td>
<td>293 (97%)</td>
</tr>
<tr>
<td>NQA</td>
<td>11 (2%)</td>
<td>8 (3%)</td>
<td>6 (2%)</td>
</tr>
<tr>
<td>Broad Spectrum</td>
<td>14 (2%)</td>
<td>6 (2%)</td>
<td>4 (1%)</td>
</tr>
<tr>
<td>ADOS Classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autism</td>
<td>665 (92%)</td>
<td>265 (91%)</td>
<td>314 (93%)</td>
</tr>
<tr>
<td>Spectrum</td>
<td>55 (8%)</td>
<td>25 (9%)</td>
<td>25 (7%)</td>
</tr>
<tr>
<td>ADOS Domain Scores (M, SD)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social</td>
<td>10.8 (2.2)</td>
<td>10.9 (2.4)</td>
<td>11.0 (2.1)</td>
</tr>
<tr>
<td>Range</td>
<td>4-14</td>
<td>4-14</td>
<td>4-14</td>
</tr>
<tr>
<td>Communication</td>
<td>6.2 (1.6)</td>
<td>6.2 (1.6)</td>
<td>6.3 (1.6)</td>
</tr>
<tr>
<td>Range</td>
<td>2-10</td>
<td>2-10</td>
<td>2-10</td>
</tr>
<tr>
<td>RRB</td>
<td>3.9 (1.5)</td>
<td>3.9 (1.6)</td>
<td>3.9 (1.5)</td>
</tr>
<tr>
<td>Range</td>
<td>0-6</td>
<td>0-6</td>
<td>0-6</td>
</tr>
<tr>
<td>VABS (M, SD)</td>
<td>n = 542</td>
<td>n = 211</td>
<td>n = 339</td>
</tr>
<tr>
<td>AB composite</td>
<td>46 (15.8)</td>
<td>54.1 (12.9)</td>
<td>36.6 (11)</td>
</tr>
<tr>
<td>DLS SS</td>
<td>44.6 (18.5)</td>
<td>52.4 (14.3)</td>
<td>33.5 (13)</td>
</tr>
<tr>
<td>Raven (range)</td>
<td>n = 198</td>
<td>n = 52</td>
<td>n = 83</td>
</tr>
<tr>
<td>NVIQ</td>
<td>28-143</td>
<td>73-132</td>
<td>28-143</td>
</tr>
</tbody>
</table>

*Under 6 = under 6 years of age, AB = Adaptive Behavior

Table 2: Module 1 sample characteristics.
Module 1. The Module 1 sample was predominately male (78%, n = 561) and the mean age was 7.2 years (SD = 3.1). Ethnicity was available for 605 participants (84%); the majority of these were identified as non-Hispanic/Latino (79%, n = 480). For those assigned to a diagnostic category, 611 (96%) were classified as having autism, 11 (2%) were classified as NQA, and 14 (2%) were classified as Broad Spectrum. The remaining 12% (n = 84) that did not have an assigned category met either Autism (n = 77) or Spectrum (n = 7) cut-offs on the ADOS. Mean and standard deviations for ADOS domains are also presented in Table 1. Ranges for domain scores in the total sample were as follows: Social = 4-14, Communication = 2-10, and RRB = 0-6.

Module 1 Subgroups. Formation of subgroups is detailed below, within the Statistical Analyses section. Two subsamples were created in Module 1: a youngest subgroup (under age 6 years, n = 290) and a lowest-functioning subgroup (based on both Vineland Adaptive Behavior Scales (VABS) Adaptive Behavior composite and Daily Living Skills (DLS) standard score < 55, n = 339).

Within the Module 1 youngest subgroup (n = 290), the majority of participants were male (77%; n = 224) and the mean age was 4.5 years (SD = .84). The lowest-functioning subgroup (n = 339) was 79% male (n = 266) with a mean age of 7.9 years (SD = 3.1). The VABS Adaptive Behavior composite scores ranged from 20 to 54 (mean = 36.6, SD = 11) while the DLS standard scores ranged from 20 to 54 (mean = 33.5, SD = 13). The ADOS domain score ranges for both subgroups were identical to those of the total sample. There also was no difference between subgroups and the total sample on demographic variables (gender, ethnicity, race).
Module 3. Module 3 sample characteristics are presented in Table 3. The Module 3 sample was predominately male (80%, \( n = 549 \)) and the mean age was 9.5 years (SD=2.7). Ethnicity and race were available for 573 participants (83%); the majority of these were identified as non-Hispanic/Latino (67%, \( n = 460 \)) and Caucasian (74%, \( n = 512 \)). For those assigned to a category, 542 (79%) were classified as having autism, 21 (3%) were classified as NQA, and 36 (5%) were classified as Broad Spectrum. Of the remaining 13% (\( n = 90 \)) that did not have an assigned category, 65 met Autism and 25 met Spectrum cut-offs on the ADOS. ADOS domain score ranges for the total sample were as follows: Social = 0-14, Communication = 0-8, and RRB = 0-8.
<table>
<thead>
<tr>
<th>Module 3</th>
<th>Total (N=689)</th>
<th>Over 10 (n=277)</th>
<th>AB ≤ 70 (n=229)</th>
<th>AB &gt; 70 (n=217)</th>
<th>IQ ≥ 100 (n=399)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age years (M, SD)</strong></td>
<td>9.5 (2.7)</td>
<td>12.2 (1.7)</td>
<td>10.3 (2.8)</td>
<td>8.7 (2.4)</td>
<td>8.9 (2.3)</td>
</tr>
<tr>
<td><strong>Gender (males) n (%)</strong></td>
<td>549 (80%)</td>
<td>227 (82%)</td>
<td>186 (81%)</td>
<td>164 (76%)</td>
<td>323 (81%)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td>n = 573</td>
<td>n = 224</td>
<td>n = 193</td>
<td>n = 185</td>
<td>n = 338</td>
</tr>
<tr>
<td>Non-Hispanic/Latino</td>
<td>460 (80%)</td>
<td>181 (81%)</td>
<td>153 (79%)</td>
<td>151 (82%)</td>
<td>277 (82%)</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>113 (20%)</td>
<td>43 (19%)</td>
<td>40 (21%)</td>
<td>34 (18%)</td>
<td>61 (18%)</td>
</tr>
<tr>
<td><strong>Diagnostic category</strong></td>
<td>n = 599</td>
<td>n = 234</td>
<td>n = 206</td>
<td>n = 188</td>
<td>n = 345</td>
</tr>
<tr>
<td>Autism</td>
<td>542 (90%)</td>
<td>221 (94%)</td>
<td>188 (91%)</td>
<td>158 (84%)</td>
<td>302 (88%)</td>
</tr>
<tr>
<td>NQA</td>
<td>21 (4%)</td>
<td>6 (3%)</td>
<td>7 (3%)</td>
<td>11 (6%)</td>
<td>13 (4%)</td>
</tr>
<tr>
<td>Broad Spectrum</td>
<td>36 (6%)</td>
<td>7 (3%)</td>
<td>11 (5%)</td>
<td>19 (10%)</td>
<td>30 (9%)</td>
</tr>
<tr>
<td><strong>ADOS Classification</strong></td>
<td>n = 599</td>
<td>n = 234</td>
<td>n = 206</td>
<td>n = 188</td>
<td>n = 345</td>
</tr>
<tr>
<td>Autism</td>
<td>479 (70%)</td>
<td>201 (73%)</td>
<td>181 (79%)</td>
<td>121 (56%)</td>
<td>258 (65%)</td>
</tr>
<tr>
<td>Spectrum</td>
<td>177 (26%)</td>
<td>63 (23%)</td>
<td>37 (16%)</td>
<td>83 (38%)</td>
<td>123 (31%)</td>
</tr>
<tr>
<td>Not Spectrum</td>
<td>33 (5%)</td>
<td>13 (5%)</td>
<td>11 (5%)</td>
<td>13 (6%)</td>
<td>18 (5%)</td>
</tr>
<tr>
<td><strong>ADOS Domain Scores (M, SD)</strong></td>
<td>n = 511</td>
<td>n = 213</td>
<td>n = 229</td>
<td>n = 217</td>
<td>n = 295</td>
</tr>
<tr>
<td>Social</td>
<td>8.2 (2.5)</td>
<td>8.3 (2.6)</td>
<td>8.6 (2.5)</td>
<td>7.7 (2.5)</td>
<td>8.0 (2.5)</td>
</tr>
<tr>
<td>Range</td>
<td>0-14</td>
<td>0-14</td>
<td>2-14</td>
<td>1-14</td>
<td>1-14</td>
</tr>
<tr>
<td>Communication</td>
<td>3.8 (1.6)</td>
<td>3.8 (1.6)</td>
<td>4.0 (1.6)</td>
<td>3.4 (1.5)</td>
<td>3.7 (1.5)</td>
</tr>
<tr>
<td>Range</td>
<td>0-8</td>
<td>0-8</td>
<td>0-7</td>
<td>0-8</td>
<td>0-8</td>
</tr>
<tr>
<td>RRB</td>
<td>1.8 (1.6)</td>
<td>1.8 (1.6)</td>
<td>2.1 (1.8)</td>
<td>1.5 (1.5)</td>
<td>1.7 (1.6)</td>
</tr>
<tr>
<td>Range</td>
<td>0-8</td>
<td>0-6</td>
<td>0-8</td>
<td>0-6</td>
<td>0-8</td>
</tr>
<tr>
<td><strong>VABS (M, SD)</strong></td>
<td>n = 617</td>
<td>n = 260</td>
<td>n = 202</td>
<td>n = 203</td>
<td>n = 399</td>
</tr>
<tr>
<td>AB composite</td>
<td>69.1 (16.4)</td>
<td>63.6 (15.5)</td>
<td>55.2 (10.8)</td>
<td>83.2 (9.7)</td>
<td>72.1 (15.7)</td>
</tr>
<tr>
<td>DLS SS</td>
<td>67.9 (20.6)</td>
<td>64.9 (20.6)</td>
<td>50.7 (15.6)</td>
<td>85.4 (9.8)</td>
<td>70.5 (20.5)</td>
</tr>
<tr>
<td>Raven (range)</td>
<td>n = 617</td>
<td>n = 260</td>
<td>n = 202</td>
<td>n = 203</td>
<td>n = 399</td>
</tr>
<tr>
<td>NVIQ</td>
<td>43-140</td>
<td>43-131</td>
<td>43-140</td>
<td>58-140</td>
<td>100-140</td>
</tr>
</tbody>
</table>

*Over 10 = Over 10 years of age, AB = Adaptive Behavior

Table 3: Module 3 sample characteristics.
Module 3 Subgroups. Four subgroups were created in Module 3: an oldest subsample (over age 10 years, \( n = 277 \)); a lowest-functioning subgroup (based on VABS Adaptive Behavior composite and DLS standard score \( \leq 70 \), \( n = 229 \)); a highest-functioning group (based on VABS Adaptive Behavior composite and DLS standard score \( > 70 \), \( n = 217 \)); and a second highest-functioning subgroup (based on Raven NVIQ \( \geq 100 \), \( n = 399 \)).

Within the Module 3 oldest subgroup (\( n = 277 \)), 82% of participants were male (\( n = 227 \)) and the mean age was 12.2 years (SD = 1.7). ADOS domain score ranges were as follows: Social = 0-14, Communication = 0-8, and RRB = 0-6. The low-functioning subgroup based on VABS scores (\( n = 229 \)) was predominately male (81%, \( n = 186 \)) with a mean age of 10.3 years (SD = 2.8). ADOS domain score ranges for this subgroup were as follows: Social = 2-14, Communication = 0-7, and RRB = 0-8. The VABS Adaptive Behavior composite scores ranged from 20 to 70 (mean = 55.2, SD = 10.8) while the DLS standard scores ranged from 20 to 69 (mean = 50.7, SD = 15.6). The high-functioning subgroup based on VABS scores (\( n = 217 \)) consisted of primarily males (76%, \( n = 164 \)); the mean age was 8.7 years (SD = 2.4). ADOS domain score ranges were as follows: Social = 1-14, Communication = 0-8, and RRB = 0-6. The VABS Adaptive Behavior composite scores ranged from 71 to 114 (mean = 83.2, SD = 9.7) while the DLS standard scores ranged from 71 to 120 (mean = 85.4, SD = 9.8). The high-functioning subgroup based on Raven estimated non-verbal IQ (\( n = 399 \)) was primarily male (81%, \( n = 323 \)) and the mean age was 8.9 years (SD = 2.3). ADOS domain score
ranges were as follows: Social = 1-14, Communication = 0-8, and RRB = 0-8. Raven NVIQs ranged from 100 to 140. There was no difference between subgroups and the total sample on demographic characteristics (gender, ethnicity, race).

**Instruments**

The *Autism Diagnostic Observation Schedule – Generic* (ADOS-G; Lord, Rutter, DiLavore, & Risi, 1999) is a semi-structured, standardized observation for the assessment of ASD. It consists of social “presses” which require a trained examiner to engage an individual in various social interactions. It consists of four modules; the appropriate module is selected based on the individual’s expressive language level and chronological age. For example, within Modules 1 and 2, examiners try to engage children in a mock birthday party for a baby doll, while in Modules 3 and 4 examiners converse with the individual about what it means to be a friend, or why someone would want to get married. Administration of each module takes approximately 30-45 minutes.

As modules 1 and 3 will be examined in this study, they are detailed below. Most ADOS items are rated from zero to three, though ratings of seven are given when behavior is abnormal but in a different way than is specified on the protocol, and eight is used when the rating is not applicable. For analyses, ratings of seven and eight were transformed to zeros. Module 1 consists of 10 activities and 29 ratings; 12 items are in the scoring algorithm. Module 3 consists of 14 activities and 28 ratings; 11 items are in the scoring algorithm. A summary of Module 1 items are presented in Table 4. A summary of Module 3 items are found in Table 5.
Module 1
A1 Overall level of non-echoed language
A2 Frequency of vocalization directed to others
A3 Intonation of vocalizations of verbalizations
A4 Immediate Echolalia
A5 Stereotyped/ idiosyncratic use of words of phrases
A6 Use of other’s body to communicate
A7 Pointing
A8 Gestures

B1 Unusual eye contact
B2 Responsive social smile
B3 Facial expressions directed to others
B4 Integration of gaze and other behaviors during social overtures
B5 Shared enjoyment in interaction
B6 Response to name
B7 Requesting
B8 Giving
B9 Showing
B10 Spontaneous initiation of joint attention
B11 Response to joint attention
B12 Quality of social overtures

C1 Functional play with objects
C2 Imagination/ creativity

D1 Unusual sensory interest in play material/ person
D2 Hand and finger and other complex mannerisms
D3 Self-injurious behavior*
D4 Unusually repetitive interests or stereotyped behaviors

E1 Overactivity*
E2 Tantrums, aggression, negative, or disruptive behavior*
E3 Anxiety*

* Indicates item not included in analyses.

Table 4: Summary of Module 1 items.
Module 3
A1 Overall level of non-echoed language
A2 Speech abnormalities associated with autism (Intonation, volume, rhythm, rate)
A3 Immediate Echolalia
A4 Stereotyped/ idiosyncratic use of words of phrases
A5 Offers information
A6 Asks for information
A7 Reporting of events
A8 Conversation
A9 Descriptive, conventional, instrumental, or informational gestures

B1 Unusual eye contact
B2 Facial expressions directed to others
B3 Language production and linked nonverbal communication
B4 Shared enjoyment in interaction
B5 Empathy/ comments on others’ emotions
B6 Insight
B7 Quality of social overtures
B8 Quality of social response
B9 Amount of reciprocal social communication
B10 Overall quality of rapport

C1 Imagination

D1 Unusual sensory interest in play material/ person
D2 Hand and finger and other complex mannerisms
D3 Self-injurious behavior*
D4 Excessive interest in or references to unusual or highly specific topics or objects or repetitive behaviors
D5 Compulsions or rituals

E1 Overactivity/ agitation*
E2 Tantrums, aggression, negative, or disruptive behavior*
E3 Anxiety*

* Indicates item not included in analyses.

Table 5: Summary of Module 3 items.

The diagnostic validity of the ADOS has been supported by multiple studies (de Bildt et al., 2004; Gotham et al., 2007; Gotham et al., 2008; Gray et al., 2008; Le Couteur
et al, 2008; Mick, 2005) as has the convergent validity of the ADOS (de Bildt et al., 2004; Gray et al., 2008; Le Couteur et al, 2008; Mick, 2005; Robertson et al., 1999). The *Autism Diagnostic Interview-Revised* (ADI-R; Lord, Rutter, & Le Couteur, 1994) is a semi-structured parent interview consisting of 93 items which are scored from zero (0, no definite behavior of the type identified) to two (2, definite abnormal behavior of the type described in the definition and coding). Scores of three (3) may also be assigned, indicating severe abnormality, and scores of seven (7) and (8) may be used as in the ADOS (and were transformed to zeros, as with the ADOS data). Thirty-four of the items contribute to the scoring algorithm; scores of three are treated as scores of two in the algorithm. Thirteen RRB items were used in this study and are presented in Table 6. Because the ADOS is thin in this area, these items were added to some of the analyses.

<table>
<thead>
<tr>
<th>Unusual preoccupations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Circumscribed interests</td>
</tr>
<tr>
<td>Repetitive use of objects or interest in parts of objects</td>
</tr>
<tr>
<td>Compulsions/rituals</td>
</tr>
<tr>
<td>Unusual sensory interests</td>
</tr>
<tr>
<td>Undue general sensitivity to noise</td>
</tr>
<tr>
<td>Abnormal, idiosyncratic, negative response to specific sensory stimuli</td>
</tr>
<tr>
<td>Difficulties with minor changes in subject’s own routines or personal environment</td>
</tr>
<tr>
<td>Resistance to trivial changes in the environment (not directly affecting the subject)</td>
</tr>
<tr>
<td>Unusual attachment to objects</td>
</tr>
<tr>
<td>Hand and finger mannerisms</td>
</tr>
<tr>
<td>Other complex mannerisms or stereotyped body movements (do not include isolated rocking)</td>
</tr>
<tr>
<td>Midline hand movements</td>
</tr>
</tbody>
</table>

Table 6: Summary of RRB items from the ADI-R included in some analyses.
The *Raven’s Progressive Matrices* (Raven, Raven, & Court, 2003) is a nonverbal measure of cognitive functioning. It consists of five sets of 12 items, with subsequent items increasing in difficulty. Estimated non-verbal age was used to select participants for the high-functioning subgroup in Module 3.

The *Vineland Adaptive Behavior Scale* (VABS; Sparrow, Balla, and Cicchetti, 1984) is a semi-structured interview of adaptive behavior. Items are organized into four domains: Communication, Daily Living Skills, Socialization, and Motor Skills. One overall score also is obtained, the Adaptive Behavior Composite. Both the overall composite and the Daily Living Skills standard score were used to select participants for the low-functioning subgroups; this was because the overall score incorporates functioning in areas essential to the diagnosis of autism (i.e., socialization and communication) so a less-related domain was also used as criterion to ensure low-functioning status was not the result of severe autistic symptomology. The VABS was used to create subgroups based on level of functioning because: (a) most of the participants had available data, and (b) researchers have suggested that adaptive behavior may be a better estimator of level of functioning in those with ASDs than IQ measures (Kraijer, 2000).

**Procedure**

Participant families for AGRE are found through various sources, including the Cure Autism Now mailing list and web site, at conferences, through posters in schools and physicians’ offices, and by word-of-mouth. At initial contact, a packet is given to families containing information about the program and a preliminary enrollment form. On the form, families identify the diagnosis they have received from their physician or
specialist; Autism, PDD-NOS and Asperger’s Disorder are accepted. Then an ADI-R is conducted by a trained interviewer. Raters need to be reliable, and are assessed periodically to ensure this (AGRE, 2009).

Interviews are conducted in families’ homes. The interview is audio-taped (with consent). Only individuals whose ASD diagnosis is confirmed are considered for blood collection. Researchers conduct full in-home exams, administer the ADOS, and collect data on other measures. Training on the ADOS is provided by certified ADOS trainers who are supervised by Catherine Lord, Ph.D., at the University of Michigan. Then data are entered into the phenotypic database.

Statistical Analyses

Confirmatory Factor Analysis. Modules were examined by CFA individually. Analyses were conducted on the entire module sample, then also on subgroups based on particular participant characteristics (i.e., age and level of functioning). CFA was performed using LISREL (Jöreskog & Sörbom, 2004). Polychoric correlation matrices were used due to the ordinal nature of the data. Diagonally weighted least squares (DWLS) was used for estimation because it is preferred over other methods when data are categorical, and it requires smaller sample sizes than other methods (Wirth & Edwards, 2007). Independent clustering solutions (in which each item loads onto only one factor) were examined with ADOS items and with ADOS and ADI-R items together. Both independent clustering solutions and bi-factor structures (in which items may load onto both specific and general factors) were examined in DSM-IV model analyses (those with only ADOS items, and those with ADOS and ADI-R items). Bi-factor structures were included in order to test the hypothesis that ASD symptoms were best explained by
both a general ASD factor and specific domain factors of social impairment, communication impairment, and RRBs. In other words, it was an additional way to conceptualize ASDs which may have helped explain the various results between previous studies. Fit indices (described below) were examined to evaluate models.

Internal consistency of factors, or how well items on factors hang together, was examined with Cronbach’s alpha. A desirable alpha level is a minimum of .80. Values above .90 are considered excellent, while values between .7-.79 are considered fair (Cicchetti, 1994).

*Measures of fit.* There are many different indices of fit, but little consensus regarding which measure (or combination of measures) is best to use for evaluating models (Bollen, 1989). Adding to the uncertainty, sometimes these different indices contradict each other. Some measures are especially sensitive to sample size (e.g., Normed Fit Index, or NFI), while others seem less impacted by $N$ (e.g., RMSEA). Model complexity is another consideration, as some indices penalize for more complex models, or models with more parameters (e.g., Parsimonious Comparative Fit Index, or PCFI) while others do not. The chosen indices were selected for several reasons, including recommendations which suggest selecting indices from different categories to reflect various criteria (Brown, 2006; Garson, 2009) and to maximize the ability to compare this study to previous work.

RMSEA is defined as the discrepancy (or the amount of error of approximation) per degree of freedom (Brown, 2006). It was selected as the primary measure of model fit because it is one of the measures least impacted by sample size and does not involve a null model (M. W. Browne, personal communication, June 1, 2010). Also, it has a
confidence interval, and graduated guidelines exist for its interpretation (rather than a single cut-off). Finally, it is commonly reported in the literature, allowing for direct comparison between previous studies and results found here.

Additional measures considered were the NNFI, CFI, SRMR, and GFI. NNFI and CFI both reflect the proportion by which the hypothesized (i.e., the specified) model improves fit compared to the null model. Also, NNFI controls for model complexity. SRMR is defined as the mean difference between the predicted and observed variances and covariances in the specified model, based on standardized residuals. GFI is defined as the percent of observed covariances explained by the covariances implied by the model. Although recently GFI may have fallen out of favor as a preferred measure of fit (Garson, 2009) it is frequently reported in previous studies, so was included here.

Guidelines have been proposed to help researchers interpret goodness of fit indices, though there is some variability in recommendations. Generally accepted cut-offs were used and are presented in Table 7. For the RMSEA, values below .08 are acceptable, with values below .05 indicating “good” fit (Browne & Cudeck, 1993). For the NNFI, values greater than or equal to .95 are considered acceptable and values below .9 indicate poor fit and a need to re-specify the model (Hu & Bentler, 1999). For the CFI, values above .95 indicate good fit (Hu & Bentler, 1999). For SRMR, values below .10 indicate acceptable fit (Kline, 2005) while values below .05 indicate good fit. For GFI, values above .95 indicate acceptable fits.
**Model Specification.** Four models were tested within each module total sample and subgroup. First, a one-factor model was specified (Figure 1). Second, a two-factor model based on previous studies (e.g., Frazier et al., 2008; Snow et al., 2008) and the scoring algorithm was tested; it consisted of a social/communication factor and a restricted, repetitive behavior factor (Figure 2). The third model was the 3-domain DSM-IV model (Figure 3). Finally, a fourth model based on recently proposed changes to the DSM was specified, called hereafter DSM-V (Figure 4); it consists of one factor composed of social and communication items and one factor composed of restricted and repetitive behavior and language, hereafter referred to as the RRB/L factor. Thus, one one-factor model was tested; one three-factor model was tested; and two two-factor models were tested, with repetitive language items grouped with social and communication items in one of these models, and with repetitive behaviors items in the other two-factor model. Figure 5 represents the bi-factor DSM-IV models, with the specific factors of social impairment, communication, and RRB and the general factor of ASD. Bi-factor models were used to assess whether three specific factors accounted for variance unique from variance accounted for by a general ASD factor. Since a one-factor model has been proposed in the literature and the concept of a broader autism phenotype (BAP) is gaining in popularity (Piven, Palmer, Jacobi, Childress, & Arndt, 1997), a bi-
factor structure could be relevant in explaining discrepant results found in the literature by taking into account both general and specific latent variables.

Subgroups. In order to examine the impact of subject characteristics on the structure of the ASD phenotype, modules were divided into more homogenous subgroups by age and level of functioning (as determined by IQ or adaptive behavior scores). This was done in case heterogeneity in the sample may obscure differences in ASD structure related to age and level of functioning. Decisions related to dividing modules 1 and 3 were made with two goals in mind: to create homogenous subgroups based on variables of interest, and to retain enough participants for stable factor solutions. As such, analyses were conducted on modules as a whole; modules by age (a “youngest subgroup” in Module 1 consisting of those 6 years and younger, and an “oldest subgroup” in Module 3 consisting of those 10 years and older); and modules by level of functioning (“lowest-functioning subgroup” in Module 1 based on VABS composite and DLS <55; a “lowest-functioning subgroup” in Module 3 based on VABS composite and DLS ≤70; a “highest-functioning subgroup” based on VABS composite and DLS >70; and a highest-functioning subgroup” in Module 3 based on Raven estimated NVIQ ≥100). This resulted in subgroups ranging from 217 participants (Module 3, VABS Composite and DLS ≥70) to 399 participants (Module 3, Raven estimated NVIQ ≥100). It was necessary to divide groups in this manner due to missing data for many participants (e.g., it was not possible to use Raven scores for all participants, as not all participants had valid Raven scores).

Indices of fit were examined to evaluate model fit.
Figure 1: One-factor model.

Figure 2: Two-factor model.
Modules 1 and 3 not included:
D3
E1-E3

Figure 3: DSM-IV model.

Modules 1 and 3 not included:
D3
E1-E3

Figure 4: DSM-V model.
Figure 5: DSM-IV bi-factor model.
CHAPTER 3

RESULTS

Results are presented for ADOS items only first, then analyses which included both ADOS and ADI-R items. Results are presented by module, with Module 1 first. Unless specified otherwise, results are presented for total sample analyses.

CFA Results

Module 1

Model Fit. Selected indices of fit for Module 1, all models, for the total sample and both subsamples are presented in Table 8.
<table>
<thead>
<tr>
<th>Model</th>
<th>N=720, total</th>
<th>n=290, youngest</th>
<th>n=339, lowest functioning</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One-factor model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSEA (90% CI)</td>
<td>.062 (.058-.066)</td>
<td>.059 (.052-.066)</td>
<td>.055 (.049-.062)</td>
</tr>
<tr>
<td>RMSEA CI</td>
<td>.058-.066</td>
<td>.052-.066</td>
<td>.049-.062</td>
</tr>
<tr>
<td>NNFI</td>
<td>.96</td>
<td>.97</td>
<td>.97</td>
</tr>
<tr>
<td>CFI</td>
<td>.96</td>
<td>.98</td>
<td>.97</td>
</tr>
<tr>
<td>SRMR</td>
<td>.088</td>
<td>.105</td>
<td>.096</td>
</tr>
<tr>
<td>GFI</td>
<td>.96</td>
<td>.97</td>
<td>.95</td>
</tr>
<tr>
<td><strong>Two-factor model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSEA (90% CI)</td>
<td>.062 (.058-.066)</td>
<td>.059 (.052-.066)</td>
<td>.055 (.049-.062)</td>
</tr>
<tr>
<td>RMSEA CI</td>
<td>.058-.066</td>
<td>.052-.066</td>
<td>.049-.062</td>
</tr>
<tr>
<td>NNFI</td>
<td>.96</td>
<td>.97</td>
<td>.97</td>
</tr>
<tr>
<td>CFI</td>
<td>.96</td>
<td>.98</td>
<td>.97</td>
</tr>
<tr>
<td>SRMR</td>
<td>.088</td>
<td>.105</td>
<td>.096</td>
</tr>
<tr>
<td>GFI</td>
<td>.96</td>
<td>.97</td>
<td>.95</td>
</tr>
<tr>
<td><strong>DSM-IV model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSEA (90% CI)</td>
<td>.056 (.052-.060)</td>
<td>.057 (.050-.065)</td>
<td>.050 (.043-.056)</td>
</tr>
<tr>
<td>RMSEA CI</td>
<td>.052-.060</td>
<td>.050-.065</td>
<td>.043-.056</td>
</tr>
<tr>
<td>NNFI</td>
<td>.97</td>
<td>.96</td>
<td>.97</td>
</tr>
<tr>
<td>CFI</td>
<td>.97</td>
<td>.98</td>
<td>.97</td>
</tr>
<tr>
<td>SRMR</td>
<td>.084</td>
<td>.102</td>
<td>.093</td>
</tr>
<tr>
<td>GFI</td>
<td>.96</td>
<td>.97</td>
<td>.96</td>
</tr>
<tr>
<td><strong>DSM-IV Bi-factor model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSEA (90% CI)</td>
<td>.047 (.043-.052)</td>
<td>.057 (.050-.065)</td>
<td>.040 (.032-.047)</td>
</tr>
<tr>
<td>RMSEA CI</td>
<td>.043-.052</td>
<td>.050-.065</td>
<td>.032-.047</td>
</tr>
<tr>
<td>NNFI</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
</tr>
<tr>
<td>CFI</td>
<td>.98</td>
<td>.98</td>
<td>.99</td>
</tr>
<tr>
<td>SRMR</td>
<td>.070</td>
<td>.092</td>
<td>.079</td>
</tr>
<tr>
<td>GFI</td>
<td>.98</td>
<td>.98</td>
<td>.97</td>
</tr>
<tr>
<td><strong>DSM-V model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RMSEA (90% CI)</td>
<td>.061 (.057-.065)</td>
<td>.059 (.051-.066)</td>
<td>.054 (.048-.061)</td>
</tr>
<tr>
<td>RMSEA CI</td>
<td>.057-.065</td>
<td>.051-.066</td>
<td>.048-.061</td>
</tr>
<tr>
<td>NNFI</td>
<td>.96</td>
<td>.97</td>
<td>.97</td>
</tr>
<tr>
<td>CFI</td>
<td>.96</td>
<td>.98</td>
<td>.98</td>
</tr>
<tr>
<td>SRMR</td>
<td>.087</td>
<td>.105</td>
<td>.095</td>
</tr>
<tr>
<td>GFI</td>
<td>.96</td>
<td>.97</td>
<td>.95</td>
</tr>
</tbody>
</table>

Table 8: Module 1 indices of fit across models.

Within the total Module 1 sample (N = 720), RMSEAs ranged from .056 (DSM-IV model) to .062 (the one-factor and 2-factor models). RMSEA confidence intervals overlapped, indicating that based on the RMSEA, no single model fit the data significantly better than other models. However, the DSM-IV model, which is a three-factor model, performed somewhat better than either of the two-factor models or the one-factor model. When other measures of model fit (including NNFI, CFI, SRMR, and GFI) were considered, all four models had acceptable to good fits. In general, subsample fits were an improvement over the total sample fits. However, in the youngest subsample, the SRMR fell within the unacceptable range for all four models. Fits tended to be best for each model within the lowest-functioning children.
When the DSM-IV bi-factor model was specified, indices of fit improved over the independent clustering DSM solution in the total sample (RMSEA = .047 versus .056) and the lowest functioning subgroup (RMSEA = .04 versus .05) while in the youngest subgroup indices were very similar (RMSEA = .057 for both). However, nine items in the total sample analysis had very weak loadings (i.e., < .1) on their specific factors, and three items had similarly weak loadings on the general ASD factor, compared to only two items in the independent clustering solution.

**Factor loadings.** Overall, the mean factor loadings for the total sample analyses were on the order of .50-.60 for social and communication factors and .35 for RRB factors the RRB/L factor of the DSM-V model. For the one-factor model, the mean item loading was .50. For the two-factor model, mean loadings were as follows: social-communication = .53, and RRB = .36. For the DSM-IV model, the average loadings were: social = .59, communication = .51, and RRB = .37. For the DSM-V model, mean loadings were: social-communication = .57, and RRB/L = .35. Two items loaded below .10 in all four total sample analyses: “overall level of non-echoed language” and “use of other’s body to communicate.” Within the youngest subsample, “overall level of non-echoed language,” also loaded below .10 in all four models, while in the lowest-functioning subsample, “use of other’s body to communicate,” “hand and finger and other complex mannerisms,” and “unusually repetitive interests or stereotyped behaviors,” had loadings below .10 in all four model analyses. In the total sample, items “integration of gaze and other behaviors during social overtures,” and “imagination/creativity,” had the highest loadings (> .80) in all four models. The item, “functional play with objects,” loaded above .80 in two of four models (two-factor and
DSM-IV) and the item “frequency of vocalization directed to others,” loaded above .80 in the DSM-IV model only.

*Internal consistency.* Alpha coefficients varied from acceptable to very poor. Alpha for the one-factor model was .73. For the two-factor model, alpha coefficients were as follows: .70 for social-communication and .35 for RRB. For the DSM-IV model, alpha coefficients were as follows: social = .79, communication = .18, and RRB = .35 (same items as two-factor RRB factor). For the DSM-V model, alphas were: social-communication = .81, RRB/L = .09.

*Association between factors.* The inter-factor correlations for each model were quite high. For the two-factor model the correlation between factors was .86. Inter-factor correlations for the DSM-IV model were: social and communication factors = .86, social and RRBs = .75, and communication and RRBs = .92. For the DSM-V model, correlation between factors was .89.

*Module 1 Analyses with Addition of ADI-R RRB Items*

*Model fit.* When the 13 RRB domain ADI-R items were added to Module 1 analyses, fits generally worsened or analyses yielded un-interpretable solutions due to estimation problems. Estimation problems can occur for several reasons, including small sample size, outliers in the data, a poorly specified model, or lack of variance. In this case, a majority of participants obtained a score of zero in seven out of the 13 ADI-R items. The lack of variability seemed to be even more present in the youngest subsample. Indeed, there was a greater proportion of scores of 0 in this subsample compared to the total sample for 9 of 13 ADI-R RRB items. This may be a scenario where a lack of variability in scores is the cause of estimation difficulties.
For the total sample, RMSEAs and 90% confidence intervals indicated acceptable fit in all models (as they did with only ADOS items), but the other four indices fell within the unacceptable range for all models (unlike analyses with ADOS only items). RMSEAs ranged from .066 (one factor model) to .07 (two-factor model). Indices of fit are presented in Table 9. While it is not uncommon for fit indices to disagree because they are based on different sources of information, this does not explain why adding ADI-R items drastically altered fit indices. A possible explanation is that the ADI-R RRB items are poorly associated with ADOS items. Examination of item intercorrelations supported this idea. The highest correlation between items was .375 between ADOS item “hand and finger and other complex mannerisms,” and ADI-R item “hand and finger mannerisms”.

<table>
<thead>
<tr>
<th></th>
<th>RMSEA (90% CI)</th>
<th>NNFI</th>
<th>CFI</th>
<th>SRMR</th>
<th>GFI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One factor model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=720, total</td>
<td>.066 (.063-.068)</td>
<td>.91</td>
<td>.92</td>
<td>.127</td>
<td>.90</td>
</tr>
<tr>
<td>n=290, youngest</td>
<td>.080 (.076-.084)</td>
<td>.91</td>
<td>.92</td>
<td>.174</td>
<td>.81</td>
</tr>
<tr>
<td>n=339, lowest functioning</td>
<td>.062 (.058-.066)</td>
<td>.92</td>
<td>.93</td>
<td>.138</td>
<td>.88</td>
</tr>
<tr>
<td><strong>Two-factor model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=720, total</td>
<td>.070 (.067-.072)</td>
<td>.90</td>
<td>.91</td>
<td>.127</td>
<td>.91</td>
</tr>
<tr>
<td>n=290, youngest</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>n=339, lowest functioning</td>
<td>.063 (.059-.067)</td>
<td>.92</td>
<td>.92</td>
<td>.138</td>
<td>.88</td>
</tr>
<tr>
<td><strong>DSM-IV model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=720, total</td>
<td>.067 (.065-.070)</td>
<td>.91</td>
<td>.91</td>
<td>.126</td>
<td>.91</td>
</tr>
<tr>
<td>n=290, youngest</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>n=339, lowest functioning</td>
<td>.061 (.057-.065)</td>
<td>.93</td>
<td>.93</td>
<td>.136</td>
<td>.88</td>
</tr>
<tr>
<td><strong>DSM-V model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=720, total</td>
<td>.068 (.065-.070)</td>
<td>.91</td>
<td>.91</td>
<td>.127</td>
<td>.90</td>
</tr>
<tr>
<td>n=290, youngest</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>n=339, lowest functioning</td>
<td>.062 (.058-.066)</td>
<td>.92</td>
<td>.93</td>
<td>.138</td>
<td>.88</td>
</tr>
</tbody>
</table>

Table 9: Indices of fit for Module 1 with addition of ADI-R RRB items.
In the subgroup analyses with the 13 ADI-R items added, three of four models (all but the one-factor model) encountered estimation problems with the youngest subgroup (i.e., the theta delta matrix, which is a matrix of measurement error terms, was not positive definite and the solution did not converge). As explained above, this could be due to a lack of variability in scores in the subsample. Analyses with the lowest-functioning subgroup followed a similar pattern to those with the total sample; that is, RMSEAs indicated acceptable fit (range from .061-.063) but all other indices indicated poor fit.

DSM-IV bi-factor analyses encountered estimation problems in the total sample and both subsamples and subsequently is not included in Table 9. In the youngest subsample, theta delta was not positive definite and the solution did not converge, while in the other two analyses the fitted covariance matrix was not positive definite. Results were not interpretable.

Factor loadings. Generally speaking, within the total sample factor loadings were quite variable, and several ADI-R items had very low loadings (<.1). The mean factor loading for the one-factor model was .41. For the two-factor model, the mean loadings were as follows: .56 for social-communication and .26 for RRB. For the DSM-IV model, mean loadings were: .63 for social, .51 for communication, and .26 for RRB. For the DSM-V model, mean factor loadings were as follows: .60 for social/communication and .26 for RRB/L. The lowest loading item in all models was unusual attachment to objects (≤ .01) and the highest loading item for all models was unusual eye contact (> .90).

Internal consistency. Overall, coefficient alphas for the total sample ranged from very poor to acceptable. Alpha coefficients for the RRB and RRB/L factors of models
improved with the addition of ADI-R items. Alpha for the one-factor model was .71. For
the two-factor model, social-communication remained unchanged with alpha = .70, but
RRB improved to .61. For the DSM-IV model, alpha coefficients were: social = .79,
communication = .18, while RRB = .61. For the DSM-V model, alpha coefficients were
.81 for social-communication and .53 for RRB/L.

Association between factors. Association between factors was high in all models.
The inter-factor correlation for the two-factor model was .82. For the DSM-IV model,
correlations were: .86 between social and communication, .78 between social and RRB,
and .81 between communication and RRB. For the DSM-V model, the correlation
between factors was .87.

Overall, adding ADI-R items did not help improve models in Module 1.

Module 3

Model fit. Selected indices of fit for Module 3, all models, for the total sample and
all subsamples are presented in Table 10.
Within the total Module 3 sample (N = 689) RMSEAs ranged from .074 (DSM-V model) to .083 (the one-factor model). As within Module 1, confidence intervals for RMSEAs overlapped between models, indicating each model fit approximately as well as the others. Based on RMSEA alone, the DSM-V model had slightly better fit. However, when other indices of fit were considered, no model had more than two indices indicating
acceptable fit. For the one-factor model, only the SRMR fell within the acceptable range. For the two-factor model, the SRMR was acceptable, but in addition to the RMSEA confidence interval extending above .08, the NNFI, CFI, and GFI were all unacceptable. The DSM-IV model followed a similar pattern, with only the SRMR falling within the acceptable range, and barely so (.099). For the DSM-V model, the NNFI, CFI, and GFI indicated poor fit, but RMSEA and SRMR indicated acceptable fit. Thus, the DSM-V model was slightly preferable as two indices fell within acceptable ranges.

Within subgroup analyses, only with children over 10 years ($n = 277$) did indices of fit improve or remain comparable to the total sample. Analyses with the two highest-functioning subgroups yielded poor fit, and the poorest fits for each model were found with the lowest-functioning subgroup ($n = 299$). As with the total sample, within subgroups there was little difference between models.

The bi-factor DSM-IV solution encountered estimation problems and did not yield meaningful results in the total sample and two of the four subsamples. Within the oldest subsample, all indices indicated acceptable fit (RMSEA = .063, 90% CI = .055-.071, NNFI = .95, CFI = .96, SRMR = .085, GFI = .96). Analyses with the highest-functioning subgroup (per VABS scores) yielded two acceptable indices (RMSEA = .069, CI = .06-.078, CFI = .95) and three unacceptable indices (NNFI = .94, SRMR = .10, GFI = .946), though these three were on the cusp of acceptable.

*Factor loadings.* As in Module 1, the average factor loadings were approximately .50-.60 for social and communication factors and .40 for RRB factors and the RRB/L factor of the DSM-V model. For the one-factor model, the mean loading was .47. For the two-factor model, the average loadings were as follows: social-communication = .53, and
RRB = .43. For the DSM-IV model, the mean loadings were: social = .57, communication = .51, and RRB = .43. For the DSM-V model, mean loadings were: social-communication = .59 and RRB/L = .40.

The item “compulsions or rituals,” loaded below .10 in two of four total sample analyses (one-factor and DSM-V). Within the oldest subsample, “language production and linked nonverbal communication,” loaded below .10 in all four models, while in the highest-functioning (by Raven) subsample “stereotyped/ idiosyncratic use of words or phrases,” loaded below .10 in three models (all but the DSM-V model) while “compulsions or rituals” was below .10 in only the DSM-V model. This item also loaded below .1 in the lowest-functioning subgroup in the DSM-V model. In the total sample analyses, the items, “offers information,” and “conversation,” had the highest loadings (> .70) in all four models. The item “amount of reciprocal social communication,” also had high loadings (> .70) in three of four models (not DSM-V). Within the DSM-IV model, the item “quality of social response” also exceeded .70. The item “unusual sensory interest in play material/person,” was the highest loading RRB item in the two-factor and DSM-IV models (.70 and .69, respectively), while in the DSM-V model the RRB/L factor highest loading item (at .65) was a language item, “immediate echolalia.”

**Internal consistency.** Coefficient alphas tended to be higher than those for Module 1. For the one-factor solution, alpha coefficient was .83. For the two-factor model, the social-communication coefficient was .84 and the RRB coefficient was .41. For the DSM-IV model, alpha coefficients were as follows: social = .73, communication = .71, and RRB = .41. For the DSM-V model, alpha coefficient was .85 for social-communication and .54 for RRB/L.
Association between factors. Inter-factor correlations for each model ranged from modest to strong. For the two-factor model the correlation was .35. Inter-factor correlations for the DSM-IV model were: between social and communication factors = .93, between social and RRBs = .36, and between communication and RRBs = .33; for the DSM-V model: .58.

Module 3 Analyses with Additional ADI-R RRB Items

Model fit. In contrast to Module 1, when the 13 RRB ADI-R items were added to Module 3 analyses, fit indices tended to improve compared to those obtained for ADOS only analyses. Selected indices of fit across models are presented in Table 11.
<table>
<thead>
<tr>
<th>Model</th>
<th>N=689, total</th>
<th>n=277, oldest</th>
<th>n=229, low functioning</th>
<th>n=217, highest AB</th>
<th>n=399, highest IQ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>One factor model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=689, total</td>
<td>.076 (.073-.079)</td>
<td>.86</td>
<td>.87</td>
<td>.112</td>
<td>.88</td>
</tr>
<tr>
<td>n=277, oldest</td>
<td>.070 (.066-.075)</td>
<td>.89</td>
<td>.90</td>
<td>.121</td>
<td>.87</td>
</tr>
<tr>
<td>n=229, low functioning</td>
<td>.104 (.099-.109)</td>
<td>.76</td>
<td>.77</td>
<td>.185</td>
<td>.82</td>
</tr>
<tr>
<td>n=217, highest AB</td>
<td>.093 (.088-.098)</td>
<td>.82</td>
<td>.83</td>
<td>.156</td>
<td>.82</td>
</tr>
<tr>
<td>n=399, highest IQ</td>
<td>.079 (.076-.083)</td>
<td>.86</td>
<td>.87</td>
<td>.128</td>
<td>.86</td>
</tr>
<tr>
<td><strong>Two-factor model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=689, total</td>
<td>.057 (.054-.060)</td>
<td>.92</td>
<td>.93</td>
<td>.088</td>
<td>.92</td>
</tr>
<tr>
<td>n=277, oldest</td>
<td>.057 (.052-.062)</td>
<td>.93</td>
<td>.93</td>
<td>.108</td>
<td>.90</td>
</tr>
<tr>
<td>n=229, low functioning</td>
<td>.107 (.102-.112)</td>
<td>.74</td>
<td>.76</td>
<td>.185</td>
<td>.82</td>
</tr>
<tr>
<td>n=217, highest AB</td>
<td>.126 (.122-.131)</td>
<td>.62</td>
<td>.68</td>
<td>.159</td>
<td>.83</td>
</tr>
<tr>
<td>n=399, highest IQ</td>
<td>.065 (.061-.068)</td>
<td>.91</td>
<td>.91</td>
<td>.108</td>
<td>.89</td>
</tr>
<tr>
<td><strong>DSM-IV model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=689, total</td>
<td>.056 (.054-.059)</td>
<td>.92</td>
<td>.93</td>
<td>.088</td>
<td>.93</td>
</tr>
<tr>
<td>n=277, oldest</td>
<td>.049 (.043-.055)</td>
<td>.93</td>
<td>.95</td>
<td>.108</td>
<td>.90</td>
</tr>
<tr>
<td>n=229, low functioning</td>
<td>-.402 (-.4-.407)</td>
<td>-.80</td>
<td>-.80</td>
<td>-.160</td>
<td>.84</td>
</tr>
<tr>
<td>n=217, highest AB</td>
<td>.102 (.972-.107)</td>
<td>.78</td>
<td>.80</td>
<td>.160</td>
<td>.84</td>
</tr>
<tr>
<td>n=399, highest IQ</td>
<td>.065 (.061-.069)</td>
<td>.90</td>
<td>.91</td>
<td>.107</td>
<td>.89</td>
</tr>
<tr>
<td><strong>DSM-IV Bi-factor model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=689, total</td>
<td>-.402 (-.4-.407)</td>
<td>-.80</td>
<td>-.80</td>
<td>-.160</td>
<td>.84</td>
</tr>
<tr>
<td>n=277, oldest</td>
<td>.049 (.043-.054)</td>
<td>.95</td>
<td>.95</td>
<td>.087</td>
<td>.94</td>
</tr>
<tr>
<td>n=229, low functioning</td>
<td>-.402 (-.4-.407)</td>
<td>-.80</td>
<td>-.80</td>
<td>-.160</td>
<td>.84</td>
</tr>
<tr>
<td>n=217, highest AB</td>
<td>-.402 (-.4-.407)</td>
<td>-.80</td>
<td>-.80</td>
<td>-.160</td>
<td>.84</td>
</tr>
<tr>
<td>n=399, highest IQ</td>
<td>-.402 (-.4-.407)</td>
<td>-.80</td>
<td>-.80</td>
<td>-.160</td>
<td>.84</td>
</tr>
<tr>
<td><strong>DSM-V model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N=689, total</td>
<td>.062 (.059-.065)</td>
<td>.91</td>
<td>.91</td>
<td>.100</td>
<td>.90</td>
</tr>
<tr>
<td>n=277, oldest</td>
<td>.066 (.061-.071)</td>
<td>.91</td>
<td>.91</td>
<td>.116</td>
<td>.88</td>
</tr>
<tr>
<td>n=229, low functioning</td>
<td>.103 (.099-.108)</td>
<td>.76</td>
<td>.77</td>
<td>.170</td>
<td>.85</td>
</tr>
<tr>
<td>n=217, highest AB</td>
<td>.080 (.074-.085)</td>
<td>.87</td>
<td>.87</td>
<td>.142</td>
<td>.86</td>
</tr>
<tr>
<td>n=399, highest IQ</td>
<td>.068 (.065-.072)</td>
<td>.89</td>
<td>.90</td>
<td>.117</td>
<td>.87</td>
</tr>
</tbody>
</table>

Table 11: Indices of fit for Module 3 with addition of ADI-R RRB items.

Within the total Module 3 sample, both the two-factor and DSM-IV models seemed to fit better than the alternatives, and were very similar to each other. Based on RMSEAs, the one-factor model performed worst, but confidence intervals for the other three models overlapped. Other fit indices besides the RMSEA again yielded mixed results, with no model having more than two indices (including RMSEA) indicating acceptable fit.
Fits obtained with the oldest subsample were comparable to those in the total sample. Analyses with the lowest-functioning group yielded the worst indices of fit in three of four models, and encountered estimation problems (i.e., the phi matrix was not positive definite) with the DSM-IV model. Within the subgroup analyses based on VABS scores Heywood cases occurred, a situation in which items have impossible loadings. In this case, the ADI-R item “midline hand movements” was loading > 10. When analyses were conducted without this item, RMSEAs improved substantially, indicating acceptable fit, but other indices of fit remained unacceptable. Although fit was poor for the subgroup with highest Raven scores when only ADOS items were included, when ADI-R items were added RMSEAs indicated acceptable fit in all but the one-factor model.

The DSM-IV bi-factor model encountered estimation problems with the total sample and three of four subsamples. For the oldest subsample, three of five indices indicated good fit (RMSEA = .049, CFI = .95, and SRMR = .087).

Factor loadings. Overall, within the total sample factor loadings were less variable than in Module 1, and very low loadings (<.10) occurred infrequently. The mean factor loading for the one-factor model was .36, with a range of .028 (compulsions or rituals) to .77 (amount of reciprocal social communication). ADI-R item loadings ranged from .052 (circumscribed interests) to .31 (repetitive use of objects or interest in parts of objects). For the two-factor model, the mean loadings were as follows: .53 for social-communication and .43 for RRB; the loadings ranged from .09 (stereotyped /idiosyncratic use of words or phrases) to .77 (offers information). ADI-R items ranged from .30 to .62 (midline hand movements and repetitive use of objects or interest in parts of objects).
of objects, respectively). For the DSM-IV model, mean item loadings were: .57 for social, .51 for communication, and .43 for RRB. Item loadings for this solution ranged from .093 (stereotyped/idiosyncratic use of words or phrases) to .80 (amount of reciprocal social communication). ADI-R item loadings ranged from .30 (midline hand movements) to .62 (repetitive use of objects or interest in parts of objects). For the DSM-V model, mean factor loadings were as follows: .59 for social/communication and .37 for RRB/L. Item loadings ranged from .1 (stereotyped/idiosyncratic use of words or phrases) to .79 (amount of reciprocal social communication). ADI-R item loadings ranged from .27 (midline hand movements) to .59 (repetitive use of objects or interest in parts of objects).

Internal Consistency. Within the total sample, coefficient alphas were good, and those for RRB and RRB/L factors were higher than those in analyses with only ADOS items. For the one-factor model, alpha coefficient was .81. For the two-factor model, alpha coefficients were .82 for social-communication and .72 for RRB. For the DSM-IV model, alpha coefficients were: social = .73, communication = .71, and RRB = .72. For the DSM-V, alpha coefficients were .85 for social-communication and .70 for RRB/L.

Association between factors. For the total sample, associations were strong between social and communication factors of the DSM, but modest between other factors. For the two-factor model, the correlation between factors was .24. For the DSM-IV model, correlations were as follows: .93 between social and communication, .23 between social and RRB, and .23 between communication and RRB. For the DSM-V model, the inter-factor correlation was .36.
CHAPTER 4

DISCUSSION

This study was the first to examine the structure of ASD symptoms with the all ADOS items in a large group of children and adolescents with ASDs. Despite having most items in common, results differed between Modules 1 and 3. Overall, analyses in Module 1 yielded better fits across all models than analyses in Module 3. Fits in Module 1 were acceptable but suggested room for improvement. Adding ADI-R items to analyses did not improve fit for Module 1 data. Within Module 3, most indices were unacceptable or borderline acceptable. Adding ADI-R items to analyses improved fits for Module 3, but there was still little differentiation between models. Within Module 1, fits generally improved within subgroup analyses. Within Module 3, fits generally worsened within subgroup analyses. Taken as a whole, these analyses did not support the hypothesis that a three-factor model would best fit the data. Results did indicate a one-factor solution is the least tenable, but otherwise, analyses did not support any single model over others. These results may help explain why so many plausible solutions are proposed in the literature: ultimately, there does not seem to be much difference between the models, and fits vary with sample characteristics.

Module 1

*Model fits.* Within the Module 1 total sample, all five indices of fit for all four models fell within the good to acceptable range. There was little differentiation between
models, and because all models were tested due to theoretical support, there is little reason to prefer one model over the others within this module. That being said, the DSM-IV model had slightly better indices of fit than other models. This echoed the findings of Lecavalier et al. (2006) and Lecavalier et al. (2009) who found the DSM-IV was superior to other models when using items from the ADI-R and ECI/CSI, respectively. Indices of fit in this study were generally comparable (and in some cases better) to Lecavalier et al.’s (2009) parent-derived data (e.g., DSM total sample RMSEA = .056 vs. .088; CFI = .97 vs. .95; GFI = .96 vs. .98; SRMR = .088 vs. .066, respectively) though less impressive than fits found with teacher-reported data.

Subgroup analyses did little to clarify the situation. That is, if one model had performed extremely well in a given subgroup, that may have been a justification to select it as being the preferable one overall. But this was not the case. Fit indices within subgroups generally did improve a little, but not to the extent that any model was clearly best. For instance, RMSEAs within the lowest-functioning subgroup ranged between .05-.055, compared to .056-.062 in the total sample, but other indices were comparable across groups. In fact, based on SRMR, models fit better in the total sample than in subgroups (e.g., total sample range = .084-.088, lowest-functioning subgroup range = .093-.096). Thus subgroup analysis did not indicate a preferable model.

**ADI-R items.** Adding ADI-R RRB items did not improve fits in Module 1. In fact, the addition of these items seemed to create estimation problems rather than clarify model evaluation.

**Bi-factor models.** The bi-factor DSM-IV model fits did improve over independent clustering DSM-IV models in Module 1, but the RMSEA confidence intervals overlapped...
with that of the independent clustering DSM-IV solution, suggesting this model was not necessarily preferable. Additionally, with the total sample, several items on each factor had very low (i.e., <.1) loadings while only two items had such weak associations in the independent clustering solution.

Loadings. Average factor loadings were weaker than seen in other studies (e.g., Georgiades et al., 2007, Snow et al., 2008) but still moderate, ranging between .35 (RRB/L domain of DSM-V model) to .59 (Social domain of DSM-IV). Generally, items on the RRB and RRB/L domains had the lowest loadings, and therefore the weakest relationship to their latent factors. One possible explanation for this is low endorsement given the relatively narrow time frame clinicians have to observe RRBs. However, the least frequently endorsed RRB item was still present in 70% of the sample (item D2, “hand and finger and other complex mannerisms”). Low endorsement is an unlikely cause of the lower factor loadings. In fact, the opposite may be true. That is, high endorsement rates coupled with a small number of items could have resulted in insufficient variability in scores. The mean RRB domain scores for Module 1 total sample and both subsamples was approximately 4 (the highest possible score being 6) while the standard deviation for all groups was approximately 1.5. The Visual inspection of score distribution indicated the RRB domain was skewed towards more severe scores. When there is little variance, there can be little covariance between items, which can result in weak loadings.

It is also possible that ADOS RRB items are not sensitive to the variety of RRB items found in younger children. There are only three RRB items (excluding self-injurious behavior) on the ADOS and it is possible that these do not fully capture the
range of RRBs present in younger children. Recent research has shown how variable
RRBs can be in young children (Ozonoff, Macari, Young, Goldring, Thompson, & Rogers, 2008) indicating it may be difficult to elicit specific RRBs during a 45-minute,
structured observation. Indeed, the ADOS authors considered this when deciding not to
include RRB items in the total score which classifies individuals as spectrum or non-

The two items with very low loadings (< .1) in all four models (“overall level of
non-echoed language” and “use of other’s body to communicate”) were both from the
communication domain of the DSM. It is noteworthy that the communication domain had
low internal consistency (further discussed below) perhaps indicating that in Module 1
(which is intended for younger, lower-functioning individuals), communication
symptoms are especially difficult to measure. Perhaps surprisingly, within the lowest-
functioning subgroup, two of the three RRB items loaded below .1 on their intended
factor (either RRB or RRB/L) in all four models: “hand and finger and other complex
mannerisms,” and “unusually repetitive interests or stereotyped behaviors.” It is unclear
why these items had such low loadings.

*Internal consistency.* Alpha coefficients for factors with the Module 1 total
sample ranged from very poor to good. The RRB factor of the two-factor and DSM-IV
models had low internal consistency ($\alpha = .35$), which is not unexpected since alpha
coefficient is related to the number of items on a scale, and the RRB factor only
contained four items. The DSM-V RRB/L factor’s internal consistency was much lower
($\alpha = .09$) despite including more items (7). This indicates items on the ADOS related to
stereotyped speech and repetitive behaviors are not closely related. Examination of inter-
item correlations indicated RRB items were minimally related to stereotyped/repetitive speech items (the largest relationship was -.14, between “idiosyncratic use of words or phrases” and “unusual sensory interest in play material/person). Given alpha coefficients were much higher for the RRB/L factor in Module 3 (for both ADOS analyses and analyses including ADI-R items) this further supports the idea that communication (including stereotyped language) is difficult to measure in participants likely to be administered Module 1 (by definition, younger, less verbal individuals). Internal consistency for the communication factor of the DSM-IV was similarly low (α = .18), adding further credence to this idea.

While adding ADI-R items did not improve fit within Module 1, it did improve alphas of the RRB and RRB/L factors. Again, as alpha tends to increase with the number of items, it is possible this improvement is an artifact of analyzing additional items. But it is also possible that the additional RRB items led to more cohesiveness of these factors by sampling a wider range, and therefore a more complete picture, of RRBs.

Module 3

Model fit. Within Module 3, only one model had two acceptable/good indices of fit (DSM-V), while two models had one indicating acceptable fit (two-factor and DSM-IV), and the one-factor model indices all suggested poor fit. Again, there was no clear-cut winner between models, but the DSM-V model did appear to perform marginally better than the others. This is encouraging as the proposed DSM-V criteria have only recently been published, and have yet to be studied. And although indices of fit in Module 3 were worse than in Module 1, and were not all within the acceptable range, they were comparable to several previous studies. For instance, Frazier et al. (2008) found the
following indices for the best models in their analyses: RMSEAs = .07, CFIs = .94, NNFIs = .92. Georgiades et al. (2007) found an RMSEA of .067, SRMR of .08, CFI of .92, and NNFI of .90 for their proposed model. With the exception of RMSEA, these values are not far off from indices found in the DSM-IV and DSM-V models here (RMSEAs = .074-.079, NNFIs = .92-.93, CFIs = .93, and SRMR = .094-.099).

As in Module 1, sample characteristics in Module 3 seemed to impact model fit, though not necessarily in the way expected: indices of fit seemed to improve for the oldest subgroup, were notably worse in the lowest-functioning subgroup, and were slightly worse in the highest-functioning subgroups. It is possible that worse indices of fits were obtained in the more homogenous subgroups due to a lack of variability within these samples. There is considerable evidence that IQ accounts for much of the variability within ASDs (e.g., Lecavalier et al., in press; Munson et al., 2008); when there is less variability in scores, there is less covariance to analyze, which could result in poorer indices of fit. The worst indices of fit in the lowest-functioning subgroup could be a reflection of Module 3 being less appropriate for lower-functioning individuals. The high-functioning group based on Raven scores produced worse indices than the high-functioning group based on VABS scores, possibly speaking to an over-estimation of level of functioning based on NVIQ in this sample.

**Bi-factor models.** Examination of bi-factor models did not aide in interpretation of results, and estimation errors occurred in several analyses.

**Item Loadings.** Average factor loadings were moderate in strength, ranging from .40 (RRB/L in DSM-V) to .59 (social-communication in DSM-V). In general, RRB items had stronger loadings in Module 3 compared to Module 1. This further supports the idea
that ADOS RRB items are better at measuring symptoms in older children and adolescents. Overall, there were also fewer items with very low loadings in Module 3 compared to Module 1. In the total sample, “compulsions or rituals” from the RRB domain had loadings below .10 in two models, while no other items had such weak loadings. This again may speak to the difficulty in measuring these behaviors within the confines of the ADOS (this symptom was observed in only 25% of the sample).

Internal consistency. Alpha coefficients for total sample analyses were acceptable to good for social and communication factors (.70 and above) and unsatisfactory for RRB and RRB/L (.41-.54), yet still higher than Module 1 values. When ADI-R items were added, all coefficients increased above .70, indicating factors were more cohesive with the addition of more RRB items.

ADI-R analyses. Unlike in Module 1, adding ADI-R items to Module 3 analyses improved fits and allowed for additional interpretation of results. With the additional 13 items, all but the one-factor model had two indices of fit within the acceptable range. Subgroup analyses followed the same pattern as in ADOS only analyses. That is, fit indices improved slightly for the oldest subgroup, but were generally comparable or worse in other subgroups, particularly the lowest-functioning subgroup. Again, poor fits in the lowest-functioning subgroup could be a reflection of Module 3 items not being designed for lower-functioning individuals. However, it is somewhat surprising that the highest-functioning subgroups did not have improved fits. The reason for this remains unclear, though it is possible that a lack of variability in scores accounts for this.

Comparisons Between Modules
While it is helpful to think of Modules 1 and 3 as separate instruments to some extent, it is also informative to consider similarities and differences of results between them. Overall, Module 1 indices of fit were substantially better than those obtained in Module 3. One explanation for this could be that Module 1 participants had more severe ASD symptoms than Module 3 participants, resulting in more clear-cut symptom relationships to latent factors. ADOS domain scores and autism diagnosis rates support this idea. However, when ADI-R items were included in the analyses, Module 3 fit indices improved (though not to the point that all indices were acceptable) whereas Module 1 indices of fit worsened. This could be further evidence that RRBs emerge later in development. Therefore, the structure of ASD symptoms with the younger Module 1 sample was not improved with more RRB items, as they are less apparent in younger children. Adding more support to this idea is the fact that RRB factor loadings were somewhat higher in Module 3 than in Module 1, both with and without the ADI-R items.

**Internal Consistency.** A notable difference between modules was that Module 3 communication, RRB, and RRB/L alpha coefficients were much higher than those obtained in Module 1. The difference was especially pronounced when ADI-R items were included in analyses. It is unclear why Module 3 items would “hang together” more than Module 1 items. It is possible that the lack of internal consistency in Module 1 is a function of ASD symptoms being more difficult to measure in younger, lower functioning individuals.

**Association between factors.** In Module 1, inter-factor correlations indicated a strong relationship between factors in all models (range from .75-.89). Even the RRB and RRB/L correlated highly with other factors. This is in contrast to previous research
indicating that the domains of ASD symptoms may be relatively distinct (e.g., Ronald et al. 2005, Ronald et al., 2006). However, these two studies included a community sample, not an ASD-specific sample, which likely resulted in greater variability in scores. In Module 3, inter-factor correlations were still quite strong between social and communication factors, but much lower between RRB and RRB/L factors and other domains when compared to Module 1. Together, these findings may suggest that core ASD symptom domains become more distinct with development. These findings also support previous findings (e.g., Charman et al., 2005) that the developmental trajectories of domains do indeed differ.

Considering both modules, one important point is that overall, the newly proposed DSM-V criteria performed as well or better than the other models. This is the first study to explicitly test this new model and to provide evidence supporting it. At a minimum, these results suggest the data are a better fit over the one-factor and the two-factor model. The original conceptualization of autism discussed two main symptom domains: social/ communication impairments and RRBs (Kanner, 1943). This study provides evidence that the social and communication domains may be so closely related that they represent one factor (as in Kanner’s work, and the recent DSM-V conceptualization). As discussed above, inter-factor correlations across modules for social and communication factors of the DSM-IV were very high (range from .86-.93). Additionally, in both modules, alpha coefficients for the combined social-communication factor of both the 2-factor and DSM-V models were higher than those for the separate social and communication factors of the DSM-IV. This was true both for analyses with ADOS only items as well as when ADI-R items were added. And while inter-factor
correlations were high in Module 1 between the RRB and RRB/L factors and other factors, in Module 3 RRB and RRB/L factors were less correlated with other factors (range from .24-.36) both with and without the addition of ADI-R items. In summary, this indicates the DSM-V model is promising, but the RRB/L domain may need to be further refined.

Model fit. Generally, indices of fit found in this study were comparable to those of previous, similar studies. For instance, Frazier et al. (2008) found an RMSEA of .07, CFI of .94, and NNFI of.92 in the best fitting ADI-R model. Snow et al. (2008) found RMSEAs between .053-.065, CFI of .95, SRMRs between .077-.084, and GFIIs between .94-.95 with different subsamples in their study of the ADI-R. These values are similar to those found in Module 1 analyses here (ranges for RMSEA=.054-.062, NNFI=.958-.969, CFI=.962-.972, SRMR between .084-.0875, and GFI=.96-.965) though Module 3 analyses, particularly with only ADOS items, tended to be poorer.

The primary method by which models were compared in this study was by examining different indices of fit, most notably the RMSEA. It is important to note that the guidelines for interpretation of these measures are arbitrary. Practically speaking, there is no difference between a CFI of .949 and .951, yet one indicates acceptable fit per guidelines and the other does not. Additionally, there is no statistical test to compare model fit; rather researchers need to consider several measures of fit and clinical and theoretical meaningfulness of solutions. It is important to note that even unacceptable fits can represent progress when compared to previous results (Bollen, 1989). It is also important to remember that not all indices of fit yield similar conclusions. For instance, there is less variability in NNFI values across modules and subsample compared to
RMSEA. Had NNFI been selected as the primary measure of fit in this study, models may have looked even more similar to each other. Furthermore, it is not clear how indices of fit translate into clinical significance. A model with an RMSEA of .03 will not necessarily be more clinically meaningful than a model with an RMSEA of .06. Research into this topic is lacking.

**Limitations**

The most salient limitation of this study relates to its sample. Almost all participants came from multiplex families, meaning more than one family member was diagnosed with an ASD. Because of this, results may not generalize to individuals from simplex families. Additionally, this was a clinical sample; all individuals had confirmed diagnoses of ASDs or met cut-off criteria for ASD diagnoses on the ADOS, indicating a significant level of ASD symptoms. Some researchers have pointed to the circularity of studying autistic symptoms among individuals with ASD diagnoses (Happé, Ronald, & Plomin, 2006), arguing the structure of such symptoms is best examined in the general population so that relationships between symptoms or domains are not artificially inflated. This also may relate to possible lack of variability discussed previously. Perhaps further examination of models within community samples or family members with ASD symptoms but without a full-blown ASD would increase variability in scores and prove valuable in studying the structure of ASD symptoms.

Another limitation deals with the nature of the AGRE database and the ADOS itself. While studying relatively homogenous subgroups can be helpful in order to examine the impact of subject characteristics on results, in this study, dividing modules into more homogenous subgroups was difficult due to missing data for participants. The
nature of the database dictates that not every child is rated on each instrument, so using some measures to divide modules was impossible. Additionally, because the ADOS controls somewhat for language level, age, and level of functioning, even total sample analyses were (theoretically) being performed on somewhat homogenous samples. This may have lead to less variability in scores, which resulted in less differentiation between models.

**Future Directions**

It will be important to examine different models of ASD structure using different data sources (e.g., the Children’s Social Behavior Questionnaire; Luteijn, Jackson, Volkmar, & Minderaa, 1998) as instruments impact results. Continued work with large samples able to be divided into homogenous subgroups will be important. If the trend for large research databases continues, such studies may become possible sooner rather than later. Based on results of this study, it seems further comparison of the DSM-IV and proposed DSM-V models is important.

One fruitful area of future endeavor would be further examining the proposed DSM-V model. Although it performed well based on CFA results, internal consistency was very poor for the RRB/L factor. The relationship between stereotyped, repetitive behaviors and stereotyped, repetitive speech needs to be further examined. Further research of the RRB items of the ADOS would be beneficial, although this may be difficult to accomplish given the few items and the limited observation period. Yet it is important, especially considering recent research with the ADOS has indicated the RRB items may contribute to diagnostic stability independent of other domains (Lord, Risi, DiLavore, Shulman, Thurm, & Pickles, 2006).
Given the variable results of factor analytic studies in the literature, it is important to remember that factor analysis is one of many methods which may need to be used to elucidate the ASD phenotype. Used in isolation, it is unlikely to yield completely satisfying results with regards to the question, “what is the structure of ASD symptoms?” Or perhaps more accurately, the question is, “what are the structures?” This is because the question of structure is complex. Overall, the three symptom domains which define ASDs are related, though research is inconclusive with regards to the extent of the association. Genetic and family studies suggest that a single underlying etiology is unlikely to explain the heterogeneity found in phenotypes. For example, some studies have demonstrated that over half the genes which contribute to variation in one symptom domain are independent from those responsible for variation in the other domains (Ronald et al., 2005; Ronald et al., 2006). Likewise, Piven et al. (1997) found family members of those with ASDs who were not on the spectrum but exhibited ASD symptoms often only exhibited one of the symptom domains. Ultimately, a growing body of literature is indicating that many different phenotypes make up ASDs, each associated with different symptom profiles, and each possibly with distinct etiologies (Abrahams & Geschwind, 2008). This too could help explain the lack of a clearly preferable model in this study. That is, categorical symptom models are important and their continued refinement is important, but they will be inadequate to best capture the “true” (and undoubtedly complicated) relationship between symptoms and symptom domains by themselves. Other methods will be key in elucidating the structure of ASD symptoms.

Future studies should compare external variables (e.g., treatment outcome, medication use, medical traits, motor skills, adaptive behavior) to ASD symptom
domains. Examining convergence and divergence of ASD symptoms with other variables may allow for a better understanding of diagnosis. It also could reveal subgroups of individuals with similar phenotypes who also share an underlying genetic link. This is important for etiology, prognosis, and treatment.

Conclusions

This study set out to examine the merit of multiple models of ASD symptomology. It was the first study to compare directly multiple competing models present in the literature using a measure based on direct observation. It also was the first study to use all items of the ADOS in a very large sample of children and adolescents, and to examine model fit amongst homogenous subgroups based on age and level of functioning. Results indicated that while there was no clear winner among the models, overall the DSM-IV and DSM-V models out-performed the one-factor and two-factor models. Furthermore, sample characteristics did impact model fit, with the more homogenous subgroups tending to yield better indices of fit on Module 1, though Module 3 subgroups fits were more variable. There were notable differences between Module 1 and Module 3 results, supporting previous research which indicates the structure of symptoms may change with development. Ultimately, improving the picture of ASD symptom structure is important in part because this information is used to develop and refine ASD diagnostic instruments and to inform decisions regarding classification systems. This study demonstrated not only the difficulty in studying the ASD symptom structure, but also that recent efforts to revise the symptom model may be moving in the right direction.
References


Outcome at 7 years of children diagnosed with autism at age 2: Predictive validity of assessments conducted at 2 and 3 years of age and pattern of symptom change over time. *Journal of Child Psychology and Psychiatry, 45*, 500-513.


Appendix A: LISREL Syntax for Module 1, One-factor Model

TI module1 1fact, nomiss
DA NI=25 NO=720 NG=1 MA=PM
PM = Mod1_1fac.poly
AC = Mod1_1fac.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 C1 C2 D1 D2 D4
MO NX=25 NK=1 LX=FU,FI TD=D1,FR PH=ST
FR LX 1 1 LX 2 1 LX 3 1 LX 4 1 LX 5 1 LX 6 1 LX 7 1 LX 8 1 LX 9 1 LX 10 1 LX 11 1 LX 12 1 LX 13 1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1 LX 19 1 LX 20 1 LX 21 1 LX 22 1 LX 23 1 LX 24 1 LX 25 1
PD
OU ALL RS MI ND=3 ME=DWLS
Appendix B: LISREL Syntax for Module 1, Two-Factor Model

TI liamodule1 2fact, nomiss
DA NI=25 NO=720 NG=1 MA=PM
PM = aMod1_2fac.poly
AC = aMod1_2fac.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 C1 C2 D1 D2 D4
MO NX=25 NK=2 LX=FU,FI TD=DLFR PH=ST
FR LX 1 1 LX 2 1 LX 3 1 LX 4 1 LX 5 1 LX 6 1 LX 7 1 LX 8 1 LX 9 1 LX 10 1 LX 11 1 LX 12 1 LX 13
1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1 LX 19 1 LX 20 1 LX 22 1
FR LX 21 2 LX 23 2 LX 24 2 LX 25 2
PD
OU ALL RS MI ND=3 ME=DWLS
Appendix C: LISREL Syntax for Module 1, DSM-IV Model

T1 amodule1 DSM, nomiss
DA NI=25 NO=720 NG=1 MA=PM
PM = Mod1_DSM.poly
AC = Mod1_DSM.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 C1 C2 D1 D2 D4
MO NX=25 NK=3 LX=FU,FI TD=DI,FR PH=ST
FR LX 9 1 LX 10 1 LX 11 1 LX 12 1 LX 13 1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1 LX 19 1 LX 20 1
FR LX 1 2 LX 2 2 LX 3 2 LX 4 2 LX 5 2 LX 6 2 LX 7 2 LX 8 2 LX 22 2
FR LX 21 3 LX 23 3 LX 24 3 LX 25 3
PD
OU ALL RS MI ND=3 ME=DWLS
Appendix D: LISREL Syntax for Module 1, DSM-V Model

TIMODULE1 vmodel, nomiss
DA NI=25 NO=720 NG=1 MA=PM
PM = Mod1_vmodel.poly
AC = Mod1_vmodel.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 C1 C2 D1 D2 D4
MO NX=25 NK=2 LX=FU,FI TD=DLFR PH=ST
FR LX 2 1 LX 3 1 LX 6 1 LX 7 1 LX 8 1 LX 9 1 LX 10 1 LX 11 1 LX 12 1 LX 13 1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1 LX 19 1 LX 20 1 LX 21 1 LX 22 1
FR LX 1 2 LX 4 2 LX 5 2 LX 21 2 LX 23 2 LX 24 2 LX 25 2
PD
OU ALL RS MI ND=3 ME=DWLS

87
Appendix E: LISREL Syntax for Module 1, Bi-factor DSM-IV Model

TIMODULE1 DSM bi, nomiss
DM NI=25 NO=720 NG=1 MA=PM
PM = Mod1_DSM.poly
AC = Mod1_DSM.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 C1 C2 D1 D2 D4
MO NX=25 NK=4 LX=FU,FI TD=DLFR PH=ID
FR LX 9 1 LX 10 1 LX 11 1 LX 12 1 LX 13 1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1 LX 19 1 LX 20 1
FR LX 1 2 LX 2 2 LX 3 2 LX 4 2 LX 5 2 LX 6 2 LX 7 2 LX 8 2 LX 22 2
FR LX 21 3 LX 23 3 LX 24 3 LX 25 3
FR LX 1 4 LX 2 4 LX 3 4 LX 4 4 LX 5 4 LX 6 4 LX 7 4 LX 8 4 LX 9 4 LX 10 4 LX 11 4 LX 12 4 LX 13 4
LX 14 4 LX 15 4 LX 16 4 LX 17 4 LX 18 4 LX 19 4 LX 20 4 LX 21 4 LX 22 4 LX 23 4 LX 24 4 LX 25 4
PD
OU ALL RS MI ND=3 ME=DWLS
Appendix F: LISREL Syntax for Module 3, One-factor Model

TI module3 1fact, nomiss
DA NI=24 NO=689 NG=1 MA=PM
PM = Mod3_1fac.poly
AC = Mod3_1fac.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 A9 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 C1 D1 D2 D4 D5
MO NX=24 NK=1 LX=FU,FI TD=D1,FR PH=ST
FR LX 1 1 LX 2 1 LX 3 1 LX 4 1 LX 5 1 LX 6 1 LX 7 1 LX 8 1 LX 9 1 LX 10 1 LX 11 1 LX 12 1 LX 13
1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1 LX 19 1 LX 20 1 LX 21 1 LX 22 1 LX 23 1 LX 24 1
PDOU ALL RS MI ND=3 ME=DWLS
Appendix G: LISREL Syntax for Module 3, Two-factor Model

TI module3 2fact, nomiss
DA NI=24 NO=689 NG=1 MA=PM
PM = Mod3_2fac.poly
AC = Mod3_2fac.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 A9 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 C1 D1 D2 D4 D5
MO NX=24 NK=2 LX=FU,FI TD=DLFR PH=ST
FR LX 1 1 LX 2 1 LX 3 1 LX 4 1 LX 5 1 LX 6 1 LX 7 1 LX 8 1 LX 9 1 LX 10 1 LX 11 1 LX 12 1 LX 13
1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1 LX 19 1 LX 20 1
FR LX 21 2 LX 22 2 LX 23 2 LX 24 2
PD
OU ALL RS MI ND=3 ME=DWLS
Appendix H: LISREL Syntax for Module 3, DSM-IV Model

TI module3 DSM, nomiss
DA NI=24 NO=689 NG=1 MA=PM
PM = Mod3_DSM.poly
AC = Mod3_DSM.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 A9 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 C1 D1 D2 D4 D5
MO NX=24 NK=3 LX=FU,F1 TD=DLFR PH=ST
FR LX 10 1 LX 11 1 LX 12 1 LX 13 1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1 LX 19 1
FR LX 1 2 LX 2 2 LX 3 2 LX 4 2 LX 5 2 LX 6 2 LX 7 2 LX 8 2 LX 9 2 LX 20 2
FR LX 21 3 LX 22 3 LX 23 3 LX 24 3
PD
OU ALL RS MI ND=3 ME=DWLS
Appendix I: LISREL Syntax for Module 3, DSM-V Model

TI module3 vmodel, nomiss
DA NI=24 NO=689 NG=1 MA=PM
PM = Mod3_vmodel.poly
AC = Mod3_vmodel.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 A9 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 C1 D1 D2 D4 D5
MO NX=24 NK=2 LX=FU,F1 TD=D1,FR PH=ST
FR LX 5 1 LX 6 1 LX 7 1 LX 8 1 LX 9 1 LX 10 1 LX 11 1 LX 12 1 LX 13 1 LX 14 1 LX 15 1 LX 16 1
LX 17 1 LX 18 1 LX 19 1 LX 20 1
FR LX 1 2 LX 2 2 LX 3 2 LX 4 2 LX 21 2 LX 22 2 LX 23 2 LX 24 2
PD
OU ALL RS MI ND=3 ME=DWLS
Appendix J: LISREL Syntax for Module 3, DSM-IV Bi-factor Model

TI module3 DSM_bi, nomiss
DA NI=24 NO=689 NG=1 MA=PM
PM = Mod3_DSM.poly
AC = Mod3_DSM.acov
LA
A1 A2 A3 A4 A5 A6 A7 A8 A9 B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 C1 D1 D2 D4 D5
MO NX=24 NK=4 LX=FU,FI TD=DLFR PH=ID
FR LX 10 1 LX 11 1 LX 12 1 LX 13 1 LX 14 1 LX 15 1 LX 16 1 LX 17 1 LX 18 1 LX 19 1
FR LX 1 2 LX 2 2 LX 3 2 LX 4 2 LX 5 2 LX 6 2 LX 7 2 LX 8 2 LX 9 2 LX 20 2
FR LX 21 3 LX 22 3 LX 23 3 LX 24 3
FR LX 1 4 LX 2 4 LX 3 4 LX 4 4 LX 5 4 LX 6 4 LX 7 4 LX 8 4 LX 9 4 LX 10 4 LX 11 4 LX 12 4 LX 13
4 LX 14 4 LX 15 4 LX 16 4 LX 17 4 LX 18 4 LX 19 4 LX 20 4 LX 21 4 LX 22 4 LX 23 4 LX 24 4
PD
OU ALL RS MI ND=3 ME=DWLS