5-4-2022

The Roles of the ADOS-2 and Cognition in Measuring and Diagnosing Autism Spectrum Disorder

Phebe Albert

The Roles of the ADOS-2 and Cognition in Measuring and Diagnosing Autism Spectrum Disorder

by

Phebe Albert

Under the Direction of MaryAnn Romski, Ph.D.

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2021

ABSTRACT

Autism is a complex and heterogenous neurodevelopmental disorder characterized by impairments in social communication skills and the presence of restricted and repetitive behaviors or interests (RRB's). While prevalence rates of autism have increased rapidly in the past decade, it remains a challenging disorder to measure and diagnose. Cognitive functioning can influence the presentation of autism symptoms and children's performances on standardized autism measures like the *Autism Diagnostic Observation Schedule, Second Edition* (ADOS-2), which contributes to misdiagnosis. This study examined retrospective chart review data from 188 children referred for clinical autism evaluations who were administered Module 3 of the ADOS-2. Findings demonstrated the superior fit and utility of a bifactor model to characterize autism symptomatology on the ADOS-2, Module 3. Results revealed non-significant effects of cognitive functioning on ADOS-2 performance and confirmed the measure's ability to assess autism overall. However, nuanced RRB's were not fully captured by the ADOS-2 in this sample of children. Regarding clinical diagnosis, a general autism trait measured by the ADOS-2 had the largest influence on diagnostic outcomes, and cognitive functioning had little influence. Lastly, this study revealed several small differences in the magnitude and direction of relations between verbal cognition and autism symptomatology in Black compared to White children.

INDEX WORDS: Autism spectrum disorder, Verbal cognition, Nonverbal cognition, Clinical best-estimate diagnosis, Autism Diagnostic Observation Schedule, Second Edition

The Roles of the ADOS-2 and Cognition in Measuring and Diagnosing Autism Spectrum Disorder

by

Phebe Albert

Committee Chair: MaryAnn Romski

Committee: Laura Dilly

Robin Morris

Rose Sevcik

Erin Tully

Electronic Version Approved:

Office of Graduate Services

College of Arts and Sciences

Georgia State University

May 2022

DEDICATION

For my sisters, Vivian and Lydia Albert, and my parents, Ken and Terri Albert, whose unconditional love and support have been invaluable throughout my life and in navigating all my academic pursuits. Also, to my friends for their compassion and humor that has carried me through my graduate school journey.

ACKNOWLEDGEMENTS

I would like to thank my committee members and academic advisors, Dr. MaryAnn Romski, Dr. Robin Morris, Dr. Rose Sevcik, Dr. Erin Tully and Dr. Laura Dilly for their guidance and instrumental contributions to the development and completion of this project, as well as their support throughout my graduate career. I would also like to thank the research team at Marcus Autism Center for their assistance with the data entry for this project, and Gal Kaldes for her invaluable contributions to the analyses conducted in this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

Autism spectrum disorder (ASD), also known as autism, is a heterogenous neurodevelopmental disorder characterized by impairments in social communication skills and the presence of restricted and repetitive behaviors or interests (RRB's) (American Psychiatric Association, 2013). The use of the term autism has changed over time, from describing a specific category of pervasive and developmental disorders, to a broader umbrella term for autism spectrum disorders. For clarity in this paper and respect for self-advocates' preferences (Lord et al., 2020), the term autism will be used throughout to broadly describe ASD's. According to the Centers for Disease Control and Prevention (CDC) and the Autism and Developmental Disabilities Monitoring (ADDM) Network, prevalence rates of autism in the U.S. have increased in recent decades from 1 in 110 children in 2006, to 1 in 54 8-year-old children in 2016 (Maenner et al., 2020). The reasons why prevalence rates have risen dramatically remains under debate. It is unclear if this recent increase is due to a rise in individuals who have autism, or if it reflects other factors including changes in diagnostic criteria, earlier age of identification, greater awareness of autism, or current trends in U.S. healthcare that prioritize autism when multiple possible diagnoses are present (reviewed by Matson & Kozlowski, 2011). Despite which factors best explain the current upward trajectory of autism diagnoses in the U.S., more and more clinicians are tasked with answering the question of whether or not a child has autism. However, making a diagnosis of autism is often complicated by the presence of co-occurring developmental problems, as well as substantial heterogeneity in autism symptom severity and overall cognitive and adaptive functioning in this population (American Psychiatric Association, 2013; Wiggins et al., 2019).

Cognitive abilities are one domain of developmental functioning that play a particularly significant role in the overall quality of life for children with autism. Cognitive deficits can affect autism symptom severity, i.e., children with autism and intellectual disability (ID) tend to experience more severe problems with social communication and demonstrate worse RRB's (Goldstein & Ozonoff, 2018; Ozonoff et al., 2005; Shulman et al., 2020). Cognitive functioning can also directly influence how a child performs on autism assessments, which contributes to misdiagnosis (Goldstein & Ozonoff, 2018; Ozonoff et al., 2005; Shulman et al., 2020). Thus, it is imperative to have a clear understanding of what our autism assessments are actually measuring. To ultimately improve our accuracy in evaluating and diagnosing autism, more research is needed to elucidate the structure of autism measured by autism assessments, and to provide information about how cognitive abilities influence autism symptomatology and how these variables work together to influence diagnostic outcomes. Therefore, the purpose of this study is twofold: 1) to examine the structure of autism that is measured by a gold-standard autism assessment tool, and 2) to investigate the relations between cognitive abilities, autism symptomatology, and diagnostic outcomes.

1.1 Measuring Autism Symptomatology

The assessment methods we use to evaluate autism play an integral role in diagnostic outcomes. In addition to an extensive review of a child's background and developmental history, a structured observation focused on assessing social communication and interactions and RRB's is an important component of a thorough autism evaluation (Hyman et al., 2020; Volkmar et al., 2014). Standardized autism-specific, clinician-administered assessment tools can provide valuable diagnostic information both directly, through their cutoff scores for autism, and indirectly by providing opportunities for structured observation by a trained clinician (Lord et al.,

2012; Molloy et al., 2011). Accurately characterizing the structure of autism traits measured by these assessment tools is vital to understanding how to interpret and use the information they provide for differential diagnosis and making appropriate treatment referrals (Kim et al., 2018; Plomin et al., 2009).

Prior research utilizing different methodologies of investigation, including factor analytic and genetic heritability twin studies is inconsistent, but in general has supported a two-factor model of autism symptomatology (i.e., social/communication deficits and RRB's) in clinical and general population samples (Bishop et al., 2006, 2013; Cuccaro et al., 2003; Leekam et al., 2011; Richler et al., 2010). In their 2008 review, Mandy and Skuse found that among seven confirmatory and exploratory factor analytic studies of autism, although specific factors differed between studies, there was a consistent pattern of at least one factor relating to social communication and a separate factor relating to "nonsocial" symptoms that broadly described RRB's. These findings were further supported in their review of genetics studies, which also suggested that despite some discrepancies, overall there is strong support for distinct biological etiologies of the social and nonsocial characteristics of autism (Mandy & Skuse, 2008). However, Mandy and Skuse underscored that despite clear distinctions between these social and nonsocial symptom clusters, a moderate correlation still exists between the two domains. This finding has been replicated in other studies that have also identified separate dimensions of autism symptoms, but moderate to large correlations between social communication and RRB-related symptoms using data from different measures including the *Child and Adolescent Symptom Inventory-4R* (CASI-4R; Gadow & Sprafkin, 2005), the *Broader Phenotype Autism Symptoms Scale* (BPASS; Dawson et al., 2007), the *Social Responsiveness Scale* (SRS;

Constantino, & Gruber, 2005), and the *Social Communication Questionnaire* (SCQ; Rutter, Bailey, et al., 2003) (Bishop et al., 2013; Cuccaro et al., 2003; Hus et al., 2007).

1.1.1 The Autism Diagnostic Observation Schedule

The *Autism Diagnostic Observation Schedule* (ADOS; Lord et al., 1999), and its most recently updated version, the *Autism Diagnostic Observation Schedule, Second Editio*n (ADOS-2; Lord et al., 2012) is the most commonly used "gold-standard" observational tool for assessing autism in research and clinical settings (Dworzynski et al., 2009; Frazier et al., 2012; Ronald et al., 2005, 2010). The ADOS-2 is a semi-structured observational measure of social interaction and communication, play skills and restricted and repetitive behaviors. The ADOS-2 has five modules that are administered according to age and expressive language level, which is determined by clinician observation of language samples during play and unstructured interactions at the start of test administration. See Table 1.1 for selection criteria for the administration of each module. Importantly, while all the modules are designed to assess the same components of autism symptomatology across different age groups and language levels, the items within each module vary, and thus findings associated with children's performances on one module are not directly comparable to findings from other modules.

Table 1.1 ADOS-2 Criteria for Module Selection

Module	**Age**	**Expressive Language Level**
T (Toddler)	12-30 months	No speech up to simple phrases
Module 1	+31 months	No speech up to simple phrases
Module 2	Children of any age	Phrase speech, i.e., "flexible use of non-echoed, three-word utterances that sometimes involve a verb and that are spontaneous, meaningful word combinations"[a]
Module 3	Best for 4-16 years	Verbally fluent, i.e., "producing a range of flexible sentence types, providing language beyond the immediate context, and describing logical connections within a sentence"[a]
Module 4	Best for +16 years	Same as Module 3

Note. [a](Figure 1, p. 10, ADOS-2 manual).

1.1.2 Updated ADOS algorithms

In response to concerns about the influence of age, language and cognitive abilities on the original ADOS algorithm scoring that are used to determine an autism diagnosis, new algorithms were developed in 2007 (Gotham et al., 2007). The new algorithm scoring achieved several important goals with regard to improved psychometric functioning including freedom from the influence of age, improved freedom from the influence of verbal IQ (with the exception of the Module 1 algorithms) and improved predictive and diagnostic validity across modules. Reflecting this improvement, a recent study by Havdahl et al. (2016) demonstrated that compared to the SRS and the *Autism Diagnostic Interview, Revised* (ADI-R; Rutter et al., 2003) the discriminative threshold of the ADOS-2 was influenced less by nonverbal IQ in a sample of children ages 2 to 13 years of age with autism and other neurodevelopment disorders (e.g., Attention-Deficit/Hyperactivity Disorder [ADHD], language disorders, ID). However, the authors caution that while the ADOS-2 is a strong tool for measuring autism symptomatology, a child's functioning in other domains such as cognition still need to be accounted for when interpreting ADOS-2 performance and making clinical decisions regarding the appropriateness of an autism diagnosis (Havdahl et al., 2016).

1.1.3 Dimensionality of the ADOS-2

Using the updated algorithm items, total scores in two domains, Social Affect (SA) and Restricted and Repetitive Behavior (RRB) are calculated when scoring the ADOS-2 (Gotham et al., 2007). This is in line with the current consensus that there are two domains or dimensions of autism symptomatology, a social and nonsocial or RRB-related domain. However, research on the dimensionality of the ADOS has been complicated by changes in what is behaviorally described as autism (Charman & Gotham, 2013), changes in the structure of the DSM criteria

(i.e., merging of social and communication domains from the DSM-IV to the DSM-5), and changes to the ADOS itself (i.e., Social Interaction and Communication domains in the ADOS restructured as Social Affect and RRB domains in the ADOS-2). According to the ADOS-2 manual, ADOS-2 algorithm items for all modules measure two distinct dimensions of autism symptoms (i.e., Social Affect and RRB's). Prior research has been inconsistent regarding the best fitting factor structure of the ADOS, with findings ranging from one to six factors (see Kamp-Becker et al., 2009; Norris et al., 2012 for reviews). However, study characteristics including sample size, which ADOS module(s) were administered, which items (i.e., algorithm and/or non-algorithm items) were included in the analyses, and participant demographics including age and language level vary greatly across studies. Norris et al. (2012) directly compared unidimensional, two- and three-correlated factor models using a large sample (N = 720) of children ages 3 to 18 years with autism. They found that for children who were assessed with Module 3, a two-factor model performed slightly better than a unidimensional and three-factor model. Regardless of exact factor structure, many studies report moderate to large correlations between factors (Mandy & Skuse, 2008). This finding is suggestive of a general autism factor that may be explaining a large proportion of common variance amongst component domains like SA and RRB. Additionally, when using the ADOS-2 clinically, an Overall Total score is derived by combining SA and RRB Total scores, and that Overall Total is compared to cutoff scores for autism. To appropriately use an overall composite score like this to determine cutoffs for autism, the ADOS-2 should demonstrate evidence that it is measuring a general autism factor. Yet, many prior factor analytic studies do not model a general autism factor.

One way to address many of the challenges and inconsistent findings regarding the ADOS-2 dimensionality is with a bifactor model. In contrast to the more traditionally used

correlated-factors and second-order models, bifactor models allow us to measure the unique variance of a general factor (e.g., autism) and specific factors (e.g., SA and RRB) on item performance; and thus to answer questions about 1) the permissibility of using a composite score like the ADOS-2 Overall Total to determine cutoffs for autism, 2) to what degree a general factor dominates item responses, and 3) to what degree specific factors make substantive contributions above and beyond a general factor (Hammer & Toland, 2016; Kline, 2015; Reise et al., 2010). To date, no other studies have investigated a bifactor model of the ADOS-2.

1.1.4 Subdivision of the RRB domain of autism

There is some existing literature that supports further subdivision of the RRB domain of autism using data collected from the ADOS-2 and other autism symptom measures. The RRB domain has most commonly been divided into two general subdomains, one often described as "repetitive sensorimotor" (RSM) that encompasses repetitive behaviors and motor mannerisms and sensory-related symptoms such as unusual sensory interests, and the other capturing symptoms related to resistance to change and ritualistic behaviors, described as "insistence on sameness" (IS) (Bishop et al., 2013; Leekam et al., 2011). In a large longitudinal study of children with autism and other neurodevelopmental disabilities at ages 2 through 9 years, Richler et al. (2010) demonstrated that other child characteristics including NVIQ were differentially associated with concurrent RSM and IS symptoms, and differentially related to these symptoms over time, providing strong evidence for a distinction between them.

RSM behaviors have been characterized as "lower-level" RRB's that are associated with general developmental level such that they occur more frequently in younger children and children with lower cognitive functioning; whereas IS RRB's are considered "higher-level" (reviewed in Leekam et al., 2011). The relationship between IS and other child variables is

slightly less clear in the existing literature. Some studies suggest that this subdomain of RRB's is relatively independent of age, verbal and nonverbal IQ, and autism symptom severity as measured by the ADI-R and the ADOS (Kuhfeld & Sturm, 2018). In contrast, Richler et al. (2010) found that social/communicative abilities as measured by the ADOS were negatively related to IS RRB's, such that children with less social communication impairment demonstrated more IS behaviors. Bishop et al. (2006) found that in a sample of children ages 15 months to 11 years, circumscribed interests and compulsions/rituals, typically categorized as IS behaviors, were more common in children with higher NVIQ's compared to those with lower NVIQ's, but other studies have not replicated this finding (e.g., Richler et al., 2010).

A major challenge in consistently replicating the structure of these RRB subdomains is that many factor analytic studies utilize items from the ADI-R (Kuhfeld & Sturm, 2018), a semi-structured parent interview that provides a thorough assessment of RRB's but is very time-intensive. Administering the ADI-R may not be feasible for all study designs and is often not feasible for frequent use in clinic-settings where time is limited. A more recent study by Bishop et al. (2013) demonstrated evidence confirming the construct validity of the RSM and IS factor structure of RRB's using the *Repetitive Behavior Scale-Revised* (RBS-R; Bodfish et al., 2000), which is a detailed measure focusing specifically on RRB's. They also included an ADOS-based RSM score in their analyses that was comprised of items D1: Unusual Sensory Interests and D2: Hand, Finger, and Other Complex Mannerisms. They confirmed that this ADOS-based RSM factor was moderately correlated with the RSM factors measured by the RBS-R and the ADI-R, $r = .33$, $r = .39$, respectively. However, they were unable to create an ADOS-based IS factor because the only conceptually "pure" IS item is D5: Compulsions or Rituals that is included in Modules 3 and 4 but not Modules 1 or 2 of the ADOS (and their sample was administered all

four modules). Since the ADOS-2 is frequently administered in clinical settings to assist with diagnosing autism, it is important to examine if the RRB subdomains of RSM and IS are measured by the ADOS-2, which has not yet been demonstrated. The present study will address this gap in the literature by investigating RSM and IS subdomains measured by the ADOS-2, Module 3.

Lastly, a recent item response theory (IRT) investigation of the precision of the ADOS in measuring two domains of autism symptomatology (i.e., SA and RRB) revealed that while the current algorithm items for the SA domain demonstrated sufficient reliability across a range of autism symptom severity, they did not reliably measure the RRB domain (Kuhfeld & Sturm, 2018). Kuhfeld and Strum (2018) demonstrated slightly improved reliability when three non-algorithm ADOS items were added to the measurement of the RRB domain, i.e., A2: Speech Abnormalities Associated with Autism, A3: Immediate Echolalia, and D5: Compulsions or Rituals. Their confirmatory factor analysis also revealed that item A2: Speech Abnormalities Associated with Autism was the most strongly related item to the latent RRB factor. Therefore, it is important that other studies continue to explore the utility of these non-algorithm items in representing autism symptomatology measured by the ADOS-2.

1.2 Cognition and Autism Symptomatology

Cognitive abilities are important to consider when assessing for autism because of the significant role that they play in the degree of functional impairment that a child with autism may experience, and the influence that they can have on the presentation and severity of autism symptoms (American Psychiatric Association, 2013; Bill & Geschwind, 2009; Mayes & Calhoun, 2011; Wiggins et al., 2019). There is also strong evidence in the extant literature that cognitive abilities influence performance on autism-specific assessment measures, which can

complicate autism diagnosis (Goldstein & Ozonoff, 2018; Ozonoff et al., 2005; Shulman et al., 2020). For example, verbal and nonverbal cognitive deficits may impair individuals' ability to describe their mental state, daily life experiences and/or to engage in the back and forth reciprocal conversation that is assessed during an autism evaluation (Leyfer et al., 2006). Individuals with autism and ID often demonstrate more severe deficits in several areas of social skills than those with ID or autism alone (Charman et al., 2011; Kanne et al., 2014; LoVullo & Matson, 2009; Smith & Matson, 2010). Further, prior research has demonstrated that in children without autism, cognitive impairments are associated with higher autism symptom severity scores on autism-specific assessments (Hus et al., 2013; Mayes & Calhoun, 2011). Taken together, these findings highlight the challenges of characterizing autism in children who present with concomitant cognitive deficits and symptoms of autism. It is crucial that research efforts to build our understanding of how cognition and autism symptomatology relate be continued to improve diagnostic accuracy and ensure that children receive appropriate intervention services (e.g., Parent et al., 2016).

1.2.1 Verbal and nonverbal cognitive abilities in autism

Children with autism demonstrate widely variable verbal and nonverbal cognitive abilities, ranging from above average to severely impaired (Farley et al., 2009; Filipek et al., 2000). Some research suggests that uneven cognitive profiles, i.e. significant discrepancies between verbal IQ (VIQ) and nonverbal IQ (NVIQ) are more common in children with autism than their typically developing peers (Courchesne et al., 2019; Joseph et al., 2002). However, research findings are mixed regarding patterns of cognitive strengths and weakness, which likely reflects differences across studies in test instruments, sample sizes, age and overall level of cognitive functioning (see Ankenman et al., 2014 for a review). Some studies suggest strengths

in expressive language and verbal reasoning (Weismer et al., 2010), while others suggest stronger nonverbal reasoning than verbal reasoning abilities (Courchesne et al., 2019; Filipek et al., 2000), e.g. a significant relative strength on the *Wechsler Intelligence Scale for Children, 4th edition* (WISC-IV) Perceptual Reasoning Index (PRI) compared to other intelligence index scores (Nader et al., 2016). Strengths in performances on specific subtests within broader intelligence indexes have also been identified, e.g. significant relative strengths on WISC-IV Matrix Reasoning and Similarities subtests, with weakness on Comprehension and processing speed subtests (i.e., Coding and Symbol Search) (Oliveras-Rentas et al., 2012). Additionally, there is a longstanding finding of a significant relative strength in performance on the WISC Block Design subtest in children with autism (see Oliveras-Rentas et al., 2012 for a review). However, these subtest performance patterns have not always been consistently replicated (Charman et al., 2011).

1.2.2 Relations between VIQ, NVIQ and autism symptoms

Findings from prior studies have demonstrated that in children with autism, greater severity of social and communication deficits were moderately and significantly related to lower verbal reasoning abilities, but were not significantly related nonverbal reasoning abilities (Joseph et al., 2002; Klin et al., 2007; Oliveras-Rentas et al., 2012). Contrastingly, in a sample of children with autism, Dawson et al. (2007) identified significant relations between NVIQ and multiple subscales of the BPASS, i.e., Social Motivation, Expressiveness, and Conversational Skills, r = -.21, -.41, -.51, respectively. Black et al. (2009) also showed a significant negative relation between NVIQ and autism communication symptoms measured by an aggregate score comprised of items from the ADI-R and the ADOS, but did not find a significant relation between NVIQ and autism social symptoms, or between VIQ or NVIQ and RRB's. Further,

Dawson et al. (2007) did not find any significant correlations between cognitive abilities and the total composite score on the flexibility/range of interests domain on the BPASS. However, in contrast, other research has demonstrated significant relations between level of cognitive functioning in children with autism and the prevalence and severity of RRB's (Bishop et al., 2006). Taken together, these findings suggest that there is likely some effect of verbal reasoning abilities on autism symptom severity, with slightly weaker evidence for an effect of nonverbal reasoning abilities on autism symptoms.

1.2.3 Racial disparities in autism and cognitive functioning

While investigating the relations between cognition and autism symptomatology in children, it is imperative to acknowledge what we know about racial disparities in these domains. It is well documented that relative to children who are White and Non-Hispanic White, children who are Black are diagnosed with autism at older ages are at greater risk for misdiagnosis, experience more difficulty accessing quality services, and have higher rates of co-morbid intellectual disability (Broder-Fingert et al., 2013; Constantino et al., 2020; Klaiman et al., 2015; Mandell et al., 2009). There is some recent evidence, however, to suggest this identification disparity is narrowing (Maenner et al., 2020). Less information is known about how specific aspects of autism symptomatology present across racial groups. Some previous studies suggest that race is not a significant predictor of autism symptom severity and that total scores on autism symptom measures do not differ significantly between white and nonwhite children (Hus & Lord, 2013; Mayes & Calhoun, 2011). It is also important to acknowledge that race as a variable may often be influenced or biased by other confounding socioenvironmental variables (e.g., parental education, parental academic expectations, and SES), and thus it is crucial not to interpret racial group differences without consideration of these possible mediating factors

(Broder-Fingert et al., 2013; Weiss & Saklofske, 2020). Overall, it is unclear from the current literature how cognitive abilities relate to the presentation of autism symptomatology similarly or uniquely across racial groups. A better understanding of these relations could provide information about the mechanisms contributing to disparities in age of identification and misdiagnosis.

1.3 Diagnosing Autism

Making a diagnosis of autism is a challenging and nuanced process that should integrate multi-method assessment approaches including standardized tests of autism symptoms and cognitive abilities, as well as expert clinical judgment to interpret test findings with consideration of important contextual factors (Frazier et al., 2014). This approach to diagnosing autism is important for accurate diagnosis and for making appropriate treatment referrals and recommendations (Shulman et al., 2020).

1.3.1 Clinical judgment in diagnostic decision-making

Clinical judgment can be incorporated into the diagnostic decision-making process in various ways. For example, in a longitudinal study of clinic-referred children being evaluated for autism, Lord et al. (2006) implemented different diagnostic processes across time points (i.e., at ages 2, 5 and 9 years) by varying the number of clinicians involved in assessment, how many of those clinicians were also involved in providing clinical insight, and by what information each clinician had access to prior to making a diagnosis. The addition of clinical judgment in the diagnostic decision-making process has been shown to improve the sensitivity and longitudinal stability of autism diagnosis (Lord et al., 2006). In contrast, diagnostic classification based solely on test cutoffs, using the ADI-R for example, has demonstrated diagnostic inconsistency between children at ages 2 compared to age 7 (Visser et al., 2017). In a sample of children 2- to 13-years-

old with diagnoses of autism or other neurodevelopmental disorders, Havdahl et al. (2016) demonstrated that judgment by experienced clinicians added to the accuracy of autism diagnosis initially (at age 2 years) and over time (by age 9 years) above and beyond the contributions of standardized test scores.

1.3.2 Clinical best-estimate diagnosis

The use of clinical judgment to interpret assessment findings and make a diagnosis is sometimes referred to as clinical best-estimate (CBE) diagnosis. CBE diagnosis typically involves a process of integrating information about a child's background, standardized test performances across multiple domains of functioning (e.g., cognitive, language, behavioral), and information from caregiver report and/or structured interviews to make a diagnosis (e.g., Lord et al., 2012). This process should involve trained, expert clinicians with experience in childhood development and developmental disabilities, who have either administered or observed administration of standardized testing, observed the child, and then reached a consensus diagnosis with one or more other clinicians (e.g. Lord et al., 2006; Moore & Goodson, 2003; Shulman et al., 2020; Stone et al., 1999). While CBE diagnosis is highly valuable in improving diagnostic accuracy (e.g., Havdahl et al., 2016), it is not without its limitations. For example, in their large, multi-site study, Lord et al. (2012) found that across sites there were significant differences in CBE diagnosis of autism, but not in diagnostic classifications according to standardized test cutoff scores.

1.3.3 Inconsistencies between CBE and standardized test diagnoses

Prior research has demonstrated that performance on the ADOS and CBE diagnosis are not consistently correlated over time, which likely reflects that CBE diagnosis, unlike performance on a standardized autism measure, is influenced by the integration of information

from other sources and about other co-occurring problems (Frazier et al., 2014). One consideration is that discrepancies between CBE diagnosis and standardized test performances may vary in relation to specific child characteristics that complicate the diagnostic picture. For example, Lord et al. (2006) found that for 40% of cases where clinicians disagreed with diagnostic classifications generated by standardized tests (i.e., on the ADI-R and the ADOS), children demonstrated either significant ID or high levels of behavioral problems, and in the remainder of those cases, autism symptoms were mild, and clinicians rated high levels of diagnostic uncertainty. In line with this, a review by Woolfenden et al. (2012) revealed that although several studies demonstrated strong stability of autism diagnosis over time (i.e., 88% to 89% stability), rates of diagnostic stability were substantially lower for children with cognitive impairment. Further complicating our understanding of diagnostic accuracy, is the presence of inconsistencies in how autism diagnosis is determined across research samples, which can vary from classification according solely to performance cutoffs on standardized measures, to clinical diagnosis where clinician judgment plays a large role (see Abbeduto et al., 2014). Relatedly, it has been suggested that research sites also differ in how, to what degree and at what time during the diagnostic decision-making process information about functioning in other developmental domains, like cognition, is accounted for (Lord et al., 2012). Overall, these findings highlight that children's functioning in domains like cognition can influence clinical judgment during the diagnostic decision-making process, and that autism diagnosis is not always consistent between standardized assessments and clinical judgment. Therefore, it is important to investigate how standardized test performances on autism-specific measures and cognitive abilities concomitantly influence diagnostic outcomes.

1.4 Summary and Research Questions and Hypotheses

In summary, there is clear evidence in the extant literature that children's performances on standardized autism-specific measures and their cognitive functioning influence clinical decision-making and diagnostic outcomes. However, how these variables relate during the clinical best-estimate diagnosis process is lacking, and concerningly, diagnostic conclusions have been shown to vary across sites. These issues are further complicated by inconsistencies in proposed models of the structure of autism traits measured by gold standard tools like the ADOS-2 across studies. Thus, more research is needed to clarify the dimensionality of autism traits measured by the ADOS-2, and to describe the relations between cognition, ADOS-2 test performances, and CBE diagnosis of autism. Additionally, there is some evidence in the literature to suggest that race may play a role in misidentification and misdiagnosis of autism, which also requires continued investigation. To address these gaps in the literature, the current study will investigate: 1) the structure of autism measured by the ADOS-2 with a novel approach, i.e., a bifactor model, 2) the influence of cognition on autism symptomatology, as measured by the ADOS-2, 3) the concomitant influences of cognition and autism symptomatology on CBE diagnosis in a clinic-referred sample of children, and 4) the relations between cognitive abilities and autism symptomatology in Black compared to White children (while there are many other racial identities, analyses were limited to these categories based upon available data in the retrospective chart review).

1.4.1 Research Question 1

Question 1: What is the dimensionality of the ADOS-2, Module 3?

I will investigate four possible underlying structures of the ADOS-2: unidimensional, correlated-factors (two-factor and three-factor), and bifactor. The three-factor correlated-factors

model will include two specific factors (i.e., the IS and RSM factors described earlier). Initially, these models will be comprised of algorithm and non-algorithm RRB items based on evidence in the existing literature of two subdomains of RRB's, and potential improvement of RRB measurement with the inclusion of non-algorithm items (see Kuhfeld et al., 2018). I will also investigate the best fitting factor structure without the inclusion of the non-algorithm items to further ensure that the best fitting model has been identified. I hypothesized that the bifactor model with the additional non-algorithm items would demonstrate superior fit.

1.4.2 Research Question 2

Question 2: Do VIQ and NVIQ predict autism symptomology measured by the ADOS-2?

Hypotheses for this research question vary based on which model of the ADOS-2 dimensionality is retained from research question 1. If the unidimensional model is retained, I hypothesized that lower VIQ would predict more severe autism symptomatology, but NVIQ would not be a significant predictor. If the two-factor model is retained, I hypothesized that lower VIQ would predict higher RRB and SA symptom severity, and lower NVIQ would predict higher RRB severity, but would not significantly predict SA severity. If the three-factor model is retained, I hypothesized that lower VIQ would significantly predict higher SA and RSM symptom severity, but would not significantly predict IS severity, and that lower NVIQ would predict higher RSM severity and lower IS severity but would not significantly predict SA severity. Because the bifactor model is a novel approach to modeling the dimensionality of the ADOS-2, there was less support for initial hypotheses. However, based on patterns in the existing literature using other factor structures, I hypothesized that VIQ would negatively predict SA and any RRB-related specific factors, and NVIQ would negatively predict any RRB-related specific factors.

1.4.3 Research Question 3

Question 3: What are the concomitant influences of VIQ, NVIQ and autism symptomatology on the outcome of clinical best-estimate diagnosis (i.e., autism vs. non-autism)?

I hypothesized that more autism symptoms, as measured by the ADOS-2, would predict a higher likelihood of receiving an autism diagnosis. Due to significant variability in the current literature regarding profiles of cognitive functioning in children with autism, I hypothesized that VIQ and NVIQ would not significantly predict diagnostic category above and beyond the influence of ADOS-2 scores.

1.4.4 Research Question 4

Question 4: Do VIQ and NVIQ relate to autism symptomatology measured by the ADOS-2 (i.e., the SA, RRB, and Overall Total scores) similarly or differently between two racial groups (i.e., children identified by parents as Black or White)?

Recent literature has underscored the significant mediating influences of several socioenvironmental variables (e.g., parental education, parental academic expectations, and SES) on cognitive functioning across racial groups, and how inappropriate conclusions can be drawn from simple group comparison analyses on the basis of race (Broder-Fingert et al., 2013; Weiss & Saklofske, 2020). Because these critical socioenvironmental factors were beyond the scope of the available data for the present analyses, I did not utilize race as a predictor of cognitive functioning, but rather conducted descriptive analyses to investigate how cognitive functioning relates to autism symptoms within each racial group. These relations are important to characterize as they may elucidate possible factors that contribute to misidentification and misdiagnosis across racial groups. This research question is largely exploratory due to the paucity of current literature on the topic. However, based on some prior studies that have found

no significant relationship between race and autism symptoms, I hypothesized that VIQ and NVIQ would correlate with ADOS-2 SA, RRB and Total scores with a similar direction and magnitude in both racial groups.

2 METHODS

2.1 Participants

This study includes data collected through a retrospective chart review of 188 children who were clinically referred by pediatricians to a large autism clinic in the Southeast for concerns about possible autism. Children are only seen in this clinic for evaluation of a new diagnosis of autism or when an autism diagnosis is unclear, but not for confirmatory or second opinion autism diagnoses. Families were primarily from the state of Georgia including the metro Atlanta area as well as other cities outside the metro area, and three families were from neighboring states including Alabama and Tennessee. The assessments in this study were conducted from 2016 through 2020. The children included in this study ranged from 4.2 to 17.6 years of age (mean age = 8.9 years, *SD* = 2.8), 68% male (N = 128), 32% female (N = 60), and all received the Module 3 of the ADOS-2. Children administered Module 3 must have fluent speech, i.e., are able to generate a range of sentences with varied sentence structures and include information beyond the immediate context. To be included children also had to have available cognitive testing scores. For 174 children (92.5%), cognitive ability scores were derived from either the WISC-V, WASI-II, DAS-II or SB-5 that was conducted by a clinician within the autism clinic on the same day that the ADOS-2 was administered. A subset of 14 children (7.4%) were assessed with one of the same aforementioned IQ tests prior to, but close to, the administration of their ADOS-2 by an independent clinician. For these children, it was clinically determined that their recently administered IQ test was a valid measure of their cognitive functioning, and therefore gathering additional IQ data was unnecessary. The majority of the children in this sample, N = 115 (61.2%), had one or more developmental and/or psychiatric diagnoses prior to being evaluated at the autism clinic. The most common pre-existing diagnosis

was ADHD (N = 85), followed by an anxiety disorder (N = 29), and Oppositional Defiant Disorder (ODD; N = 19). Other diagnoses included depressive disorders, speech and language impairment, Obsessive-Compulsive Disorder (OCD), intellectual and developmental disability, learning disability, trichotillomania, developmental coordination disorder, sensory integration disorder, adjustment and other behavioral/conduct disorders. There were also 21 (11.2%) children who had been previously diagnosed with either "rule-out autism," "provisional autism," or an autism spectrum disorder. The majority of children (N = 149, 79.3%) had received and/or were currently receiving intervention services at the time of their evaluation. These intervention services included being followed by a psychiatrist, psychological/behavioral interventions, play therapy, social skills therapy, speech-language, occupational and/or physical therapies. Additionally, 68 children (36.2%) were prescribed one or more psychotropic medications at the time of their evaluation. The present sample consisted of the following racial backgrounds which were identified on an intake questionnaire completed by children's caregivers: 45 (23.9%) Black, 120 (63.8%) White, 15 (8.0%) Mixed-Race, 3 (1.6%) Asian, 1 (0.5%) American Indian, and 4 (2.1%) unknown or declined to respond. Approximately 60-70% of this sample was insured by Medicaid for their evaluations. Years of education information was only available for 91 mothers (48%) and ranged from 8th grade to doctorate degrees (e.g., JH/PhD/MD), and for 80 fathers (43%) ranging from 8th grade to Master's degrees.

2.2 Procedures

The Children's Healthcare of Atlanta's Institutional Review Board approved the study procedures for this retrospective chart review. Children's Healthcare of Atlanta (CHOA) has a data sharing agreement with Georgia State University (GSU) that allows procedures for this study conducted by affiliates of GSU to be approved by CHOA's IRB. All caregivers consented

for their children to be evaluated at the Marcus Autism Center. The retrospective chart review IRB allows for the collection of information obtained through clinical evaluations to be analyzed in a de-identified format for research purposes without direct consent from patients and their caregivers.

Initially, caregivers participated in a 1-hour diagnostic interview with a psychologist and provided information about children's medical, developmental, and academic histories as well as their social, emotional, and behavioral functioning. Following this interview, caregivers and children participated in on-site evaluations lasting approximately 3 hours conducted by a team of at least two psychologists at an autism clinic. A cognitive assessment and Module 3 of the ADOS-2 were administered to the children.

2.3 Measures

2.3.1 Cognitive functioning

Verbal and nonverbal reasoning abilities were assessed using one of four cognitive assessments selected according to age and clinical judgment regarding level of difficulty and appropriateness for individual cases. The *Differential Ability Scales, Second Edition* (DAS-II; Elliot, 2007), the *Wechsler Intelligence Scale for Children, Fifth Edition* (WISC-V; Wechsler, 2014), the *Wechsler Abbreviated Scale of Intelligence, Second Edition* (WASI-II; Wechsler, 2011), or the *Stanford-Binet Intelligence Scales, Fifth Edition* (SB-5; Roid, 2003) was administered to each child.

The DAS-II measures cognitive abilities across multiple domains of cognitive functioning. It produces a General Conceptual Composite (GAC) score of overall cognitive functioning, as well as composite *cluster* scores (i.e., Verbal Ability, Nonverbal Ability, Spatial Ability, and a Special Nonverbal Composite [SNC]), and ability-specific subtest scores for each

of the core subtests that comprise the cluster scores. There are two separate batteries of the DAS-II, the Early Years battery (Upper and Lower levels) appropriate for children ages 2:6 to 6:11, or children ages 7:0 to 8:11 of low ability, and the School-Age battery appropriate for children ages 7:0 to 17:11, or children ages 5:0 to 6:11 of high ability. Children in this study were administered either the Early Years Lower or Upper level form, or the School-Age form. Verbal and Nonverbal Ability cluster standard scores (M = 100, SD = 15) were used in analyses to represent VIQ and NVIQ, respectively, for children who received the DAS-II. According to the DAS-II manual, average internal reliability coefficients (IRT-based, see manual, p. 125) for the Early Years battery Verbal and Nonverbal Reasoning clusters were .90 and .89, respectively, and for the School-Age battery, .89 and .92, respectively. Average internal reliability remained high when examined in special population samples including children with a range of developmental disabilities (e.g., ADHD, language disorder, intellectual disability, etc.). Test-retest stability for the Verbal and Nonverbal Reasoning composite scores were generally excellent (r's greater than .80), except for the Nonverbal Reasoning composite for children ages 5:0 to 8:11, r = .77, which was still well within the adequate range. Regarding validity, the Verbal and Nonverbal Reasoning cluster scores also meet criteria, according to Kaufman (1994), for ample evidence of specificity, supporting the use of these scores as measures of distinct cognitive abilities.

The WISC-V is a comprehensive assessment of cognitive functioning appropriate for children ages 6:0 to 16:11. It yields index scores (M = 100, SD = 15) representing several areas of cognitive abilities, i.e., the Verbal Comprehension Index (VCI), Visual Spatial Index (VSI), Fluid Reasoning Index (FRI), Working Memory Index (WMI) and Processing Speed Index (PSI), and a measure of overall cognitive functioning, Full Scale IQ (FSIQ). Previous versions of the WISC (Wechsler, 1991; 2003) have traditionally included fewer index scores. An update of

the WISC-V is the replacement of the Perceptual Reasoning Index (PRI) from the WISC-IV (Wechsler, 2003), with VSI and FRI, which provides more fine-tuned information across specific components of nonverbal cognitive abilities. For the purpose of the current study, scores from the FRI were used to represent NVIQ because it is conceptually most closely aligned with the NVIQ scores of the other cognitive tests that participants were administered. The VCI was used to represent VIQ. According to the WISC-V manual, split-half reliability coefficients for all index scores ranged from .88 to .93 for children ages 6 to 16 years in the standardization sample. Average test-retest reliability for index scores across all age groups also ranged from good to excellent range, with the exception of the FRI, which was still within the acceptable range (r = .75). Evidence is also provided within the manual supporting strong convergent and discriminant validity across subtests and index scores.

The WASI-II is a brief cognitive assessment appropriate for individuals ages 6:0 to 90:11 that provides a quick but reliable measure of general verbal and nonverbal intellectual abilities. The WASI-II has two- and four-subtest versions that provide standard scores (M = 100, SD = 15) on several indices, a Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI) and a Full Scale IQ (FSIQ-4 or FSIQ-2 depending on how many subtests are administered). Test items are designed to parallel those on the WISC-V and administration instructions are similar but more streamlined. The VCI and PRI scores were used to represent VIQ and NVIQ, respectively, for children who were administered the WASI-II in this study. According to the WASI-II manual (Wechsler, 2011), average split-half reliability coefficients for the VCI and PRI in children ages 6 to 16 years were .94 and .92, respectively. Test-retest reliability for two age groups, children 6 to 11 and 12 to 16 years, were also high for the VCI, r = .96 and .93, respectively, and were adequate for the PRI, r = .87 and .86, respectively. With regard to

validity, the internal structure of the WASI-II has been supported by strong intercorrelations between the subtests and composite scores, via factor analyses, and concurrent validity with other measures of intelligence ranging from acceptable to excellent (McCrimmon & Smith, 2013).

The SB-5 is a test of intellectual ability and cognitive functioning that can be administered to individuals ages 2 to 85+ years of age. The SB-5 yields a Full Scale IQ (FSIQ) that is comprised of 10 subtests (M = 10, SD = 3), five of which comprise a Nonverbal IQ scale and the remaining five comprise a Verbal IQ scale. In a large nationally representative sample of individuals ages 2 to 85 years, all SB-5 composite scores (i.e., FSIQ, Nonverbal IQ and Verbal IQ) demonstrated strong reliability, .91 to .98 (see the SB-5 manual, Roid, 2003). According to the manual (Roid, 2003), criterion and construct validity analyses demonstrated that the SB-5 is a valid measure of cognitive functioning for individuals across a range of ages, races and cognitive abilities. The Verbal IQ scale and Nonverbal IQ scale were used to represent VIQ and NVIQ, respectively, for children who were administered the SB-5.

2.3.2 Autism symptoms

The ADOS-2 is a semi-structured, standardized assessment of autism that evaluates an individual's skills in the areas of socialization, communication, play and the presence of RRB's. There are five modules of the ADOS-2 that are administered according to an individual's expressive language level and chronological age. Only children who were administered a Module 3 were included in this study. Each module can be administered in approximately 40 to 60 minutes and involves a series of "presses," i.e., social activities presented by the examiner that allow for the evaluation of an individual's social-communicative strengths and weaknesses, as well as elicit certain behaviors that are diagnostically relevant to autism. After observations

are made during the administration of the ADOS-2 activities, ratings ("codes") are assigned to individual items on a scale of 0 (indicating no evidence of abnormality) to 3 (strong evidence of abnormality) in the domains of language and communication, reciprocal social interaction, play/imagination, stereotyped behaviors and restricted interests, and other abnormal behaviors. In addition, for some items a code of 7 can be assigned when behavior is abnormal, but in a way not indicated by the protocol, or an 8 when a rating is not applicable to a particular child. These codes are converted to algorithm scores within two domains, Social Affect (SA) and Restricted and Repetitive Behavior (RRB), and then added to produce an Overall Total score, which is used to determine ADOS-2 classification (i.e., autism, autism spectrum, or non-spectrum) based on cutoff scores for each module. Algorithm scoring required that all scores of a "3" be converted to a "2," however in analyses for the present study I used the initial scores ranging from 0 to 3 to increase the variability in performances that were measured across children. All ADOS-2 administrations in the present study were scored by research reliable licensed psychologists and were administered by psychology trainees under the supervision of a licensed psychologist.

According to the ADOS-2 manual, both interrater item reliability and test-retest reliability for the SA, RRB and Overall Total scores was high across all five ADOS-2 modules, with generally stronger reliability in the SA than the RRB domain. Exploratory and confirmatory factor analyses supported a two-factor model (i.e., an SA and RRB factor). Item-total correlations, i.e., correlations between items and their assigned domain, ranged from r = .33 to .78 for the SA domain, and from r = .27 to .55 for the RRB domain, which was expected to be lower due to greater heterogeneity among items in this domain. For Module 3, only one SA algorithm item, A7: Reporting of Events, was strongly correlated with verbal mental age, r = -.48. The ADOS-2 manual indicates good sensitivity and specificity of the ADOS-2, Module 3,

i.e., 91% and 84%, respectively. The ADOS-2 has also demonstrated good sensitivity (83% to 91%) and specificity (86% to 94%) across each of the five modules in other independent samples (Shulman et al., 2020).

2.4 Data Analysis

Descriptive statistics and analyses were conducted with Mplus (Version 8; Muthén & Muthén, 2005) and IBM SPSS Statistics (Version 27) predictive analytics software. First, multiple confirmatory factor analyses (CFA)'s were specified and evaluated to determine the best fitting model to represent the dimensionality of the ADOS-2; these models are depicted in Figure 2.1 and Figure 2.2. Next, a structural regression model based upon the best fitting CFA was specified and evaluated. Before conducting these analyses, I inspected the current data set for amount and type of missing data. Missing data can be characterized in three ways according to the data loss mechanism (see Brown, 2006; Kline, 2015). The most preferable scenario is when data are missing completely at random (MCAR), meaning that missingness is not related to any study variables. Missing at random (MAR) is when missing data is predictable and related to another study variable, e.g., children with a particular NVIQ level were less likely to complete certain ADOS-2 items. Lastly, data can be missing not at random (MNAR), which indicates that the data loss is "nonignorable" and missingness is directly related to a study variable itself, e.g., children with a certain NVIQ could not complete certain ADOS-2 items. Only 0.2% of the current data was missing, and those missing data points were investigated and determined to be MCAR. Kline (2015) notes that different approaches to handling missing data are not likely to make a significant difference in results when <5% of the dataset is missing. Therefore, analyses were run with the default Mplus approach of pairwise deletion for the mean- and variance-adjusted weighted least squares (WLSMV) estimation that was used.

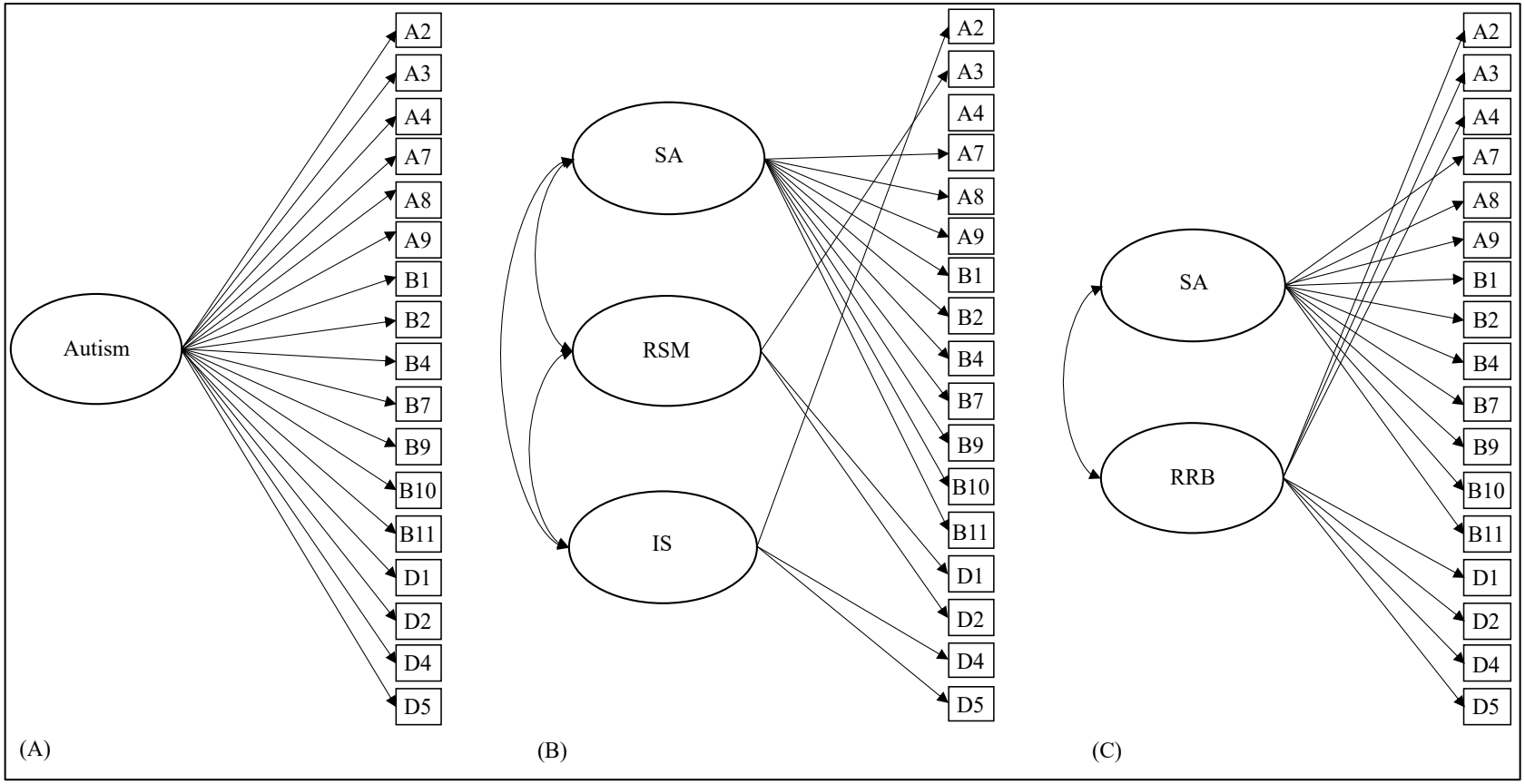

Figure 2.1 Unidimensional and Correlated-Factors CFA Models

Note. SA = Social Affect, RRB= Restricted and Repetitive Behavior, RSM = Repetitive Sensorimotor, IS = Insistence on Sameness. (A) Unidimensional model of autism represented by 17 algorithm and non-algorithm (i.e., A2, A3 and D5) ADOS-2 items, (B) Three-factor correlated-factors model with two RRB subdomains, IS and RSM, (C) Two-factor correlated-factors model. All indicators represent categorical items on the ADOS-2 scored on scales of 0-2 or 0-3.

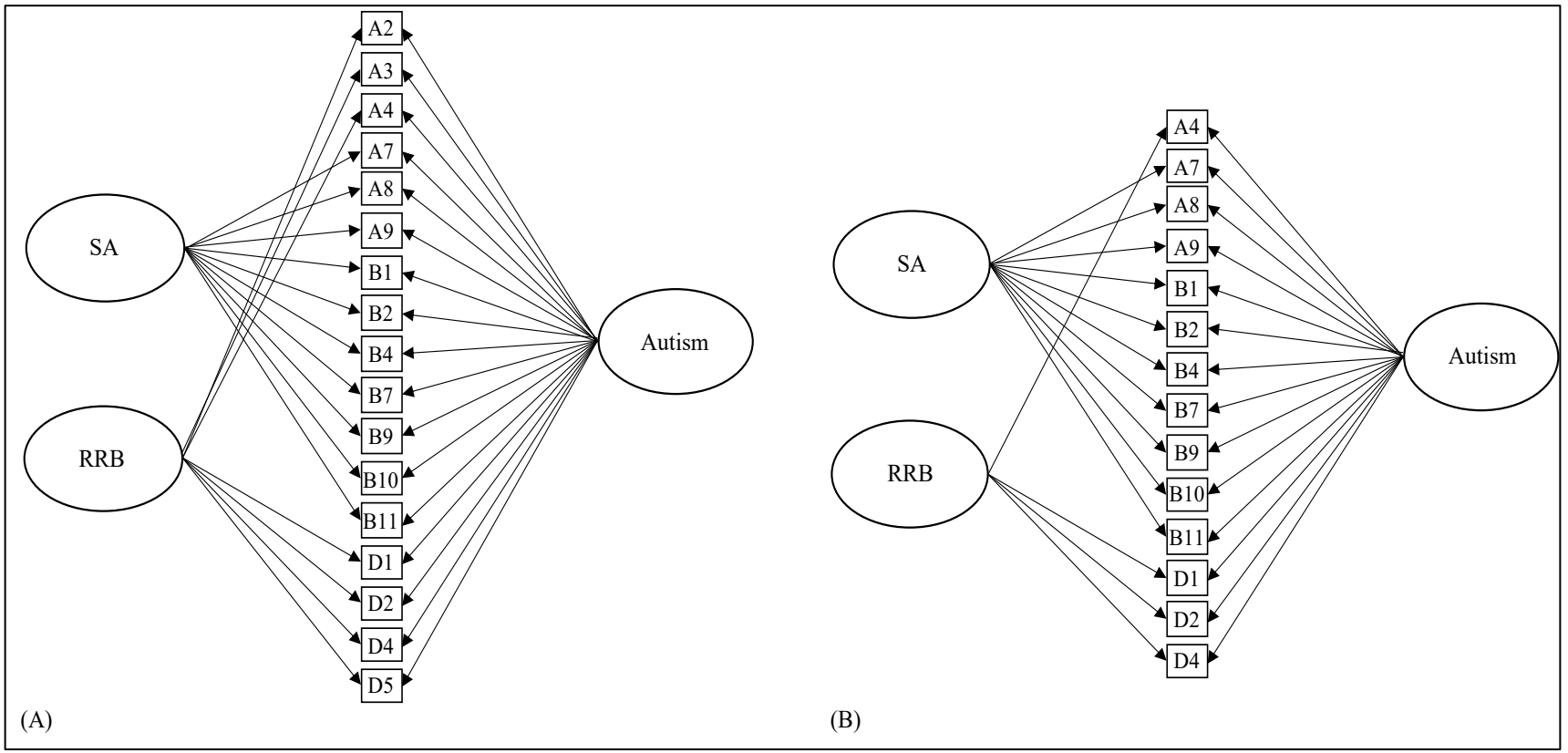

Figure 2.2 Bifactor Model 1 and Bifactor Model 2

Note. SA = Social Affect, RRB = Restricted and Repetitive Behavior. (A) Bifactor Model 1 with 17 algorithm and non-algorithm (i.e., A2, A3 and D5) ADOS-2 items, (B) Bifactor Model 2 with only the 14 algorithm items. All indicators represent categorical items on the ADOS-2 scored on scales of 0-2 or 0-3.

2.4.1 Research Question 1

Question 1: What is the dimensionality of the ADOS-2, Module 3?

Multiple CFA's were run to compare four potential structures of the ADOS-2, Module 3 dimensionality, 1) a unidimensional model with autism as the general latent factor, 2) a two-factor correlated-factors model with the latent factors SA and RRB (as represented in the ADOS-2 scoring algorithm), 3) a three-factor correlated-factors model with an SA factor and two RRB subdomain latent factors, i.e., RSM and IS, and 4) a bifactor model with a general autism factor and either two or three specific factors depending on the model fit of the two- compared to three-factor correlated-factors models. Each CFA initially included 17 indicators that represented individual ADOS-2 algorithm and non-algorithm items that have been suggested in prior research to add substantially to the measurement of autism with the ADOS-2 (i.e., items A2, A3, and D5; see Kuhfeld & Strum, 2018). Raw scores were used for each item, and items scored a 7 or an 8 were converted to 0 as is suggested in the ADOS-2 algorithm scoring. This resulted in scores ranging from 0-2 or 0-3 on all items. Lastly, the best fitting model was re-run with only the ADOS-2 algorithm items.

Due to the ordinal nature of the factor indicators, a categorical variable methodology (CVM) approach was used to analyze the data (see Finney & DiStefano, 2013; Kline, 2015). This approach comprises two components to address categorical indictors, 1) analysis input that accounts for categorical data, and 2) weighted least squares estimation. Regarding the first component, it is theorized that observed ordinal indicators are associated with an underlying latent response variable, i.e., a normal distribution of the amount of the indicator required for each of the categorical responses. Accordingly, the models in this study included latent response variables, as well as "threshold" values that represent the point of each latent response variable

where one response category was selected over another. I ran models with the default parameterization, which is delta. Latent factor metrics for the unidimensional and correlated-factors models were set using the Mplus default marker indicator approach, which automatically sets the first indicator for each latent factor as the reference indicator, fixing that indicator's factor loading to 1.00. For the bifactor model, the first indicators were freed, and the variance of each factor was set to 1.00, which has advantages for bifactor models (see Hammer et al., 2016). Regarding the second component of the CVM approach, I employed WLSMV estimation, which has been shown to outperform the other robust WLS approaches (Finney & DiStefano, 2013; Kline, 2015).

To evaluate the proposed CFA models, I examined: 1) overall goodness of fit indices, and 2) attributes such as size and significance of the model's parameter estimates including covariances, standardized loadings, and residuals. To assess overall goodness of fit I examined the chi-square test of model fit, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit index (CFI), and Tucker-Lewis index (TLI) values. RMSEA values < 0.06 suggest good model fit, values < 0.08 suggest adequate fit, and values > 0.1 indicate mediocre fit or should be rejected (Bollen & Long, 1993; Hu & Bentler, 1999; MacCallum et al., 1996). SRMR values less than .08, and CFI and TLI values from .90 - .95 suggest acceptable model fit (Bentler, 1990; Brown, 2006; Hu & Bentler, 1999). According to Kline (2015), model misfit is also suggested by residuals less than 0 and standardized loadings or covariances greater than 1. Nested models with good fit were compared using the chi-square different test. This test compares the less restrictive model (i.e., the model with more free parameters) to the more restrictive model. A significant difference test indicates that the less restrictive model demonstrates significantly better fit.

Due to the less restrictive nature of bifactor models compared to unidimensional and correlated-factors models, they almost always demonstrate superior fit; therefore, I computed additional indexes to characterize the bifactor models (Hammer & Toland, 2016; Rodriguez et al., 2016a, 2016b). To assess overall model-based reliability, the coefficient omega hierarchical (omegaH) was computed. OmegaH measures the systematic variance in total scores that is attributable to individual levels of the general factor, therefore, large omegaH values (> .80) are indicative of a total test score that can be considered "essentially unidimensional" (Rodriguez, 2016, p. 144). To directly evaluate the dimensionality of the ADOS-2, I calculated the explained common variance of the general factor (ECV), and the percent of uncontaminated correlations (PUC), which represents the relative amount of a model's correlation matrices that reflect a general factor. It is recommended that the ECV and PUC be used in conjunction with one another because while ECV directly influences parameter bias, this relation is moderated by the PUC. When both ECV and PUC values are > .70, the common variance among items in a bifactor model can generally be considered unidimensional (Rodriguez et al., 2016b, 2016a). Lastly, to assess the amount of bias in parameter estimates that may occur by specifying a unidimensional model from a multidimensional test like the ADOS-2, I examined the average relative parameter bias (ARPB). It has been suggested that parameter bias less than 10% to 15% indicates little difference between the unidimensional and bifactor model estimates of the general factor. This would suggest that a unidimensional model is an appropriate structure to specify the dimensionality of the ADOS-2, even when a bifactor model demonstrates superior statistical fit according to traditional fit indices (Rodriguez et al., 2016b).

2.4.2 Research Question 2

Question 2: Do VIQ and NVIQ predict autism symptomology measured by the ADOS-2?

A structural regression model, depicted in Figure 2.3, was conducted to address this research question. The best fitting CFA from the prior analyses was entered into the structural regression model with VIQ and NVIQ added as predictors of the ADOS-2 based latent factors, along with sex and age at covariates (which act essentially as additional predictors in SEM). Age is an important covariate because children included in the study spanned meaningfully different developmental age groups, i.e., early childhood through adolescence. With regard to sex, it is well documented that sex plays a role in the severity and presentation of autism symptomatology (see Rubenstein et al., 2015 for a review). In the structural regression model, VIQ was a standard score (M = 100, SD = 15) represented by either the DAS-II Verbal Ability cluster score, the WISC-V VCI, the WASI-II VCI, or the SB-5 Verbal IQ scale, based on which test a child was administered. NVIQ was also a standard score (M = 100, SD = 15) represented by either the DAS-II Nonverbal Ability cluster score, the WISC-V FRI, the WASI-II PRI, or the SB-5 Nonverbal IQ scale. To evaluate this model, I examined the same aspects of model findings that were evaluated for the CFA models: 1) overall goodness of fit indices, and 2) attributes such as size and significance of the model's parameter estimates including covariances, standardized loadings, and residuals.

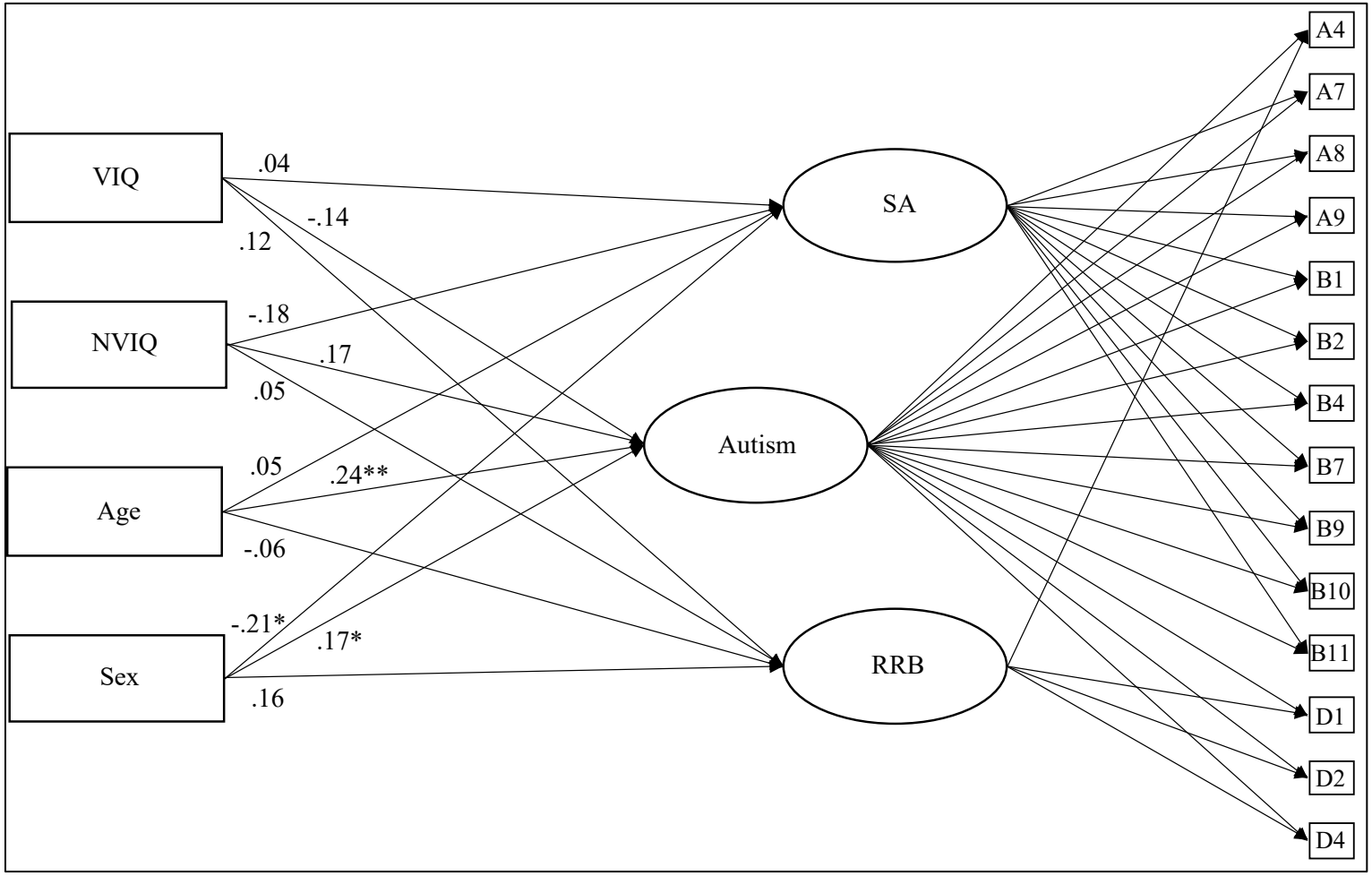

Figure 2.3 Structural Regression Model of VIQ, NVIQ, Sex and Age Predicting ADOS-2 Based Latent Factors

Note. Parameter estimates are standardized. VIQ = Verbal IQ, NVIQ = Nonverbal IQ, SA = Social Affect, RRB = Restricted and Repetitive Behavior. *$p < .05$. **$p <. 001$.

2.4.3 Research Question 3

Question 3: What are the concomitant influences of VIQ, NVIQ and autism symptomatology on the outcome of clinical best-estimate diagnosis (i.e., autism vs. non-autism)?

For this analysis, CBE diagnosis was a binary categorical variable obtained from final diagnostic reports for each child. CBE diagnosis was labeled a "1" for autism, which included children who received an autism diagnosis only, and children who received an autism diagnosis with one or more co-occurring developmental and/or psychiatric disorders, or a "0" for non-autism, which included children who received any non-autism diagnosis or no diagnoses. Because of the significant positive skew of this outcome variable, i.e., 69% of children received an autism diagnosis (N = 130), and the relatively small size of the overall sample, a model including all predictors of interest (i.e., VIQ, NVIQ, and the ADOS-2 based latent factors) on CBE diagnosis would not converge in Mplus. Therefore, a binary logistic regression was conducted in SPSS with VIQ, NVIQ, and factor scores entered as predictors of CBE diagnosis. Factor scores were calculated from the latent factors of the best-fitting CFA produced in Mplus and then imputed into SPSS to be used in the regression analysis as observed variables. All predictors were converted to z-scores due to significantly different scales (i.e., ranges of 79 and 87 for VIQ and NVIQ, compared to ranges of 3.5 to 4.1 for factor scores). Additional diagnostic statistics to check for signs of bias and multicollinearity were also calculated.

2.4.4 Research Question 4

Question 4: Do VIQ and NVIQ relate to autism symptomatology measured by the ADOS-2 (i.e., the SA, RRB, and Overall Total scores) similarly or differently between two racial groups (i.e., children identified by parents as Black or White)?

Pearson's *r* correlations were conducted between VIQ, NVIQ and ADOS-2 Total scores (including the SA, RRB and Overall Total scores) in the two available racial groups, White and Black children, separately.

3 RESULTS

3.1 Descriptive Statistics

Descriptive statistics for all study variables are reported in Table 3.1. Spearman's rho correlations between ADOS-2 Total scores and cognitive functioning revealed significant, moderate to large correlations between ADOS-2 Total scores, and between VIQ and NVIQ, but small, non-significant correlations between cognitive functioning and ADOS-2 Total scores, see Table 3.2. Regarding missing data, one child was missing a VIQ score. This child's IQ data was provided by an outside clinician and only the NVIQ score was available. Three children were missing at least one ADOS-2 item. Item B2: Facial Expressions to Examiner was not coded for one child who was wearing a mask due to COVID-19 precautions. Items A2: Abnormal Speech, A3: Immediate Echolalia, A7: Reporting of Events, and B2: Facial Expressions to Examiner were not coded for one child due to language challenges (i.e., primary language was Spanish, and speech was difficult to understand). Lastly, item D2: Hand/Finger/Other Complex Mannerisms was not coded for one child who demonstrated complex mannerisms related to a diagnosis of Tourette's syndrome.

Table 3.1 Descriptive Statistics for ADOS-2 Items, VIQ and NVIQ

	Range	Mean	Standard Deviation
ADOS-2 Item			
A2: Abnormal Speech	0 – 2	0.89	0.71
A3: Immediate Echolalia	0 – 3	0.17	0.44
A4: Stereotyped Words/Phrases	0 – 3	0.69	0.61
A7: Reporting of Events	0 – 3	0.61	0.65
A8: Conversation	0 – 3	0.91	0.75
A9: Gestures	0 – 3	0.42	0.56
B1: Unusual Eye Contact	0 – 2	1.35	0.94
B2: Facial Expressions to Examiner	0 – 2	0.75	0.59
B4: Shared Enjoyment	0 – 3	0.55	0.73
B7: Quality Social Overtures	0 – 3	0.82	0.56
B9: Quality Social Response	0 – 3	0.87	0.46
B10: Reciprocal Social Comm	0 – 3	0.76	0.66
B11: Overall Quality of Rapport	0 – 3	0.82	0.62
D1: Unusual Sensory Interest	0 – 3	0.40	0.62
D2: Complex Mannerisms	0 – 3	0.61	0.86
D4: Excessive Interest	0 – 3	0.94	0.83
D5: Compulsions or Rituals	0 – 2	0.60	0.63
VIQ	63 – 142	95.93	14.85
NVIQ	61 – 148	97.71	15.83

Note. VIQ = Verbal IQ, NVIQ = Nonverbal IQ. VIQ and NVIQ were derived from either the WISC-V, WASI-II, SB-5, or DAS-II. VIQ and NVIQ scores from all measures represent standard scores (M = 100, SD = 15), see methods section for details regarding which standard scores were used to represent VIQ and NVIQ for each measure. ADOS-2 items are raw scores.

Table 3.2 Correlations Between VIQ, NVIQ, and ADOS-2 Totals

	1	2	3	4	5
1. VIQ	1.00				
2. NVIQ	.62**	1.00			
3. ADOS-2 SA Total	-.08	.01	1.00		
4. ADOS-2 RRB Total	.06	.14	.48**	1.00	
5. ADOS-2 Overall Total	-.05	.04	.94**	.73**	1.00

Note. N = 187 for VIQ, and 188 for all other variables. VIQ = Verbal IQ, NVIQ = Nonverbal IQ, SA = Social Affect, RRB = Restricted and Repetitive Behavior. ADOS-2 Total scores are raw scores, and VIQ and NVIQ are standard scores (M = 100, SD = 15).

**$p < .001$.

3.2 RQ 1: Dimensionality of the ADOS-2

Fit indices for each CFA model are presented in Table 3.3. All four models produced significant chi-square tests, which can indicate poor absolute fit (Kline, 2015). However, this test is not sufficient to retain or reject a model and is influenced by sample size such that larger sample sizes are more likely to produce significant chi-square tests than smaller sizes (Kline, 2015). Values for RMSEA, SRMR, CFI and TLI fit indices revealed adequate to excellent fit for all four models. Additionally, there were no standardized loadings or covariances greater than 1, or residual variances less than 0 in any model. Therefore, I proceeded with chi-square difference testing of nested models. The three-factor correlated-factors model with RRB's subdivided into two factors (i.e., RSM and IS) was compared to the two-factor model (i.e., SA and RRB). A chi-square test revealed that the two-factor model demonstrated better fit than three-factor model, $\Delta\chi 2 = 0.90$, $\Delta df = 2$, $p = .64$. The two-factor correlated-factors model also fit better than a unidimensional model, $\Delta\chi 2 = 35.23$, $\Delta df = 1$, $p < .001$.

Next, I compared the two-factor correlated-factors model with the novel bifactor model that included ADOS-2 algorithm and non-algorithm items (i.e., Bifactor Model 1, see Figure 2.2 [A]). Bifactor Model 1 demonstrated significantly better fit than the two-factor correlated-factors model, $\Delta\chi 2 = 43.38$, $\Delta df = 16$, $p < .001$. Lastly, I tested a bifactor model with only the ADOS-2 algorithm items (i.e., Bifactor Model 2, see Figure 2.2 [B]). Bifactor Model 2 that included only algorithm items demonstrated better fit than Bifactor Model 1, which included non-algorithm items A2, A3 and D5. Fit statistics for Bifactor Model 1 and 2 are reported in Table 3.4. In summary, Bifactor Model 2 demonstrated the best fit.

Standardized loadings for the best fitting Bifactor Model 2 are reported in Table 3.5. All items loaded positively and significantly on the general autism factor and ranged from .34 to .89.

All items loaded more highly on the general autism factor than their respective specific factors except for items A8: Conversation and D1: Unusual Sensory Interest in Play Material/Person. For the RRB specific factor, all items also loaded significantly and positively (N = 4). This suggests that children who scored high overall on the RRB domain were more likely to be rated higher, i.e., more symptomatic, on each RRB item. For the SA specific factor, six items loaded positively (60%), five of which were statistically significant, and four items loaded negatively (40%), none of which were statistically significant. This suggests that children who scored high overall on the SA domain of the ADOS-2 were more likely to score higher on most SA items. However, they were more likely to score lower, i.e., less symptomatic, on other items, although these items did not reach statistical significance.

Table 3.3 Fit Statistics for CFA Models

	χ^2	*df*	RMSEA	SRMR	CFI	TLI
Unidimensional	258.56**	119	0.08	0.08	0.96	0.95
Two-factor	206.70**	118	0.06	0.08	0.97	0.97
Three-factor	207.58**	116	0.07	0.08	0.97	0.97
Bifactor	165.74**	102	0.06	0.06	0.98	0.97

Note. RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, CFI = Comparative Fit Index, and TLI = Tucker-Lewis Index.

**$p < .001$.

Table 3.4 Fit Statistics for Non-Nested Bifactor Models 1 and 2

	χ^2	*df*	RMSEA	SRMR	CFI	AIC/BIC
Bifactor Model 1	165.74**	102	0.06	0.06	0.98	4809.72/5045.98
Bifactor Model 2	95.66**	63	0.05	0.05	0.99	3999.86/4189.041

Note. Bifactor Model 1 includes three non-algorithm items (A2, A3 and D5), Bifactor Model 2 includes only algorithm items. RMSEA = Root Mean Square Error of Approximation, SRMR = Standardized Root Mean Square Residual, CFI = Comparative Fit Index, AIC = Akaike's Information Criteria, and BIC = Bayesian Information Criteria.

**$p < .001$.

Table 3.5 Factor Loadings for Bifactor Model 2

ADOS-2 Items	**Estimates (SE)**		
	Autism[a]	SA	RRB
A4	.60 (.05)		.55 (.12)**
A7	.41 (.08)	.19 (.11)	
A8	.55 (.10)	.71 (.12)**	
A9	.50 (.07)	-.01 (.12)	
B1	.89 (.05)	-.02 (.13)	
B2	.80 (.05)	-.20 (.11)	
B4	.78 (.04)	-.02 (0.10)	
B7	.82 (.04)	.23 (.10)*	
B9	.86 (.05)	.25 (.12)*	
B10	.78 (.04)	.20 (.10)*	
B11	.85 (.06)	.39 (.12)**	
D1	.34 (.08)		.37 (.12)*
D2	.43 (.07)		.43 (.12)**
D4	.44 (.06)		.41 (.10)**

Note. Bifactor Model 2 includes only ADOS-2 algorithm items. SE = standard error, Autism = general autism factor, SA = Social Affect specific factor, RRB = Restricted and Repetitive Behavior specific factor.

[a]All factor loadings on the general autism factor were significant at $p < .001$.

*$p < .05$, **$p < .001$.

3.2.1 Ancillary bifactor analyses

Additional analyses were conducted to further characterize Bifactor Model 1 and 2. For Bifactor Model 1, I calculated an omegaH of .75. OmegaH subscale (omegaHS) was .41 for the SA specific factor, and .01 for the RRB specific factor. This model produced a PUC of .54. The ECV was .64, indicating that 64% of the item variance in Bifactor Model 1 could be accounted for by the general autism factor. Average relative parameter bias (ARBP) was 27%, which indicates that the parameter estimates for the item loadings on the general autism factor produced by Bifactor Model 1 were quite different from those yielded by a unidimensional model. For Bifactor Model 2, I calculated an omegaH of .87. OmegaHS was .05 for the SA specific factor, and .35 for the RRB specific factor. This model produced a PUC of .44 and the ECV was .79,

indicating that 79% of item variance was explained by the general autism factor. ARBP was 7%, suggesting small differences among item loadings on the general autism factor produced by Bifactor Model 2 compared to a unidimensional model.

Lastly, I inspected IECV values for Bifactor Model 1 and 2, results are reported in Table 3.6. These values represent the amount of variance in an item attributable solely to the general factor. Values larger than 0.80 to 0.85 suggest that an item largely reflects the characteristics of the general factor (Stucky & Edelen, 2014). IECV values revealed a greater proportion of items in Bifactor Model 2 compared to Bifactor Model 1 that primarily represented the general autism factor. In summary, the ADOS-2 dimensionality represented by Bifactor Model 2 could be considered essentially unidimensional, whereas analyses for Bifactor Model 1 exhibited weaker evidence for the appropriateness of a unidimensional structure.

Table 3.6 IECV Values for Bifactor Model 1 and 2

ADOS-2 Item	Bifactor Model 1	Bifactor Model 2
A2	**0.92**	--
A3	0.61	--
A4	**1.00**	0.55
A7	0.25	**0.82**
A8	0.28	0.38
A9	0.42	**1.00**
B1	0.75	**1.00**
B2	0.76	**0.94**
B4	0.54	**1.00**
B7	0.70	**0.93**
B9	0.66	**0.92**
B10	0.59	**0.94**
B11	0.49	**0.82**
D1	**0.93**	0.46
D2	**0.89**	0.50
D4	**0.86**	0.54
D5	0.35	--

Note. Bifactor Model 1 includes three non-algorithm items (A2, A3 and D5), Bifactor Model 2 includes only algorithm items. All values in the Bifactor Model 1 and Bifactor Model 2 columns

represent IECV values, values > 0.80 are bolded to indicate that these values largely represent the content of the general factor.

3.3 RQ 2: Relations between Cognition, Sex, Age and Autism Symptomatology

The structural regression model for research question 2 is depicted in Figure 2.3. Model fit indices revealed adequate to excellent fit for this model, RMSEA = .05, CFI/TLI = 0.98/0.98, and SRMR = 0.11. Additionally, the model did not produce any standardized loadings or covariances greater than 1, or residual variances less than 0. Standardized parameter estimates were only significant for sex and age. Sex significantly predicted the general autism factor (R^2 = .03) and the SA specific factor (R^2 = .04), such that being male was associated with greater overall autism symptom severity and being female was associated with greater SA symptom severity. Age significantly predicted the general autism factor (R^2 = .06), such that older age was associated with greater overall autism symptom severity. VIQ and NVIQ were not significant predictors of any of the ADOS-2 based latent factors. Due to the non-significant direct effects of VIQ and NVIQ on the autism latent factors, investigation of indirect effects of VIQ and NVIQ on CBE diagnosis through the autism factors were not pursued in further analyses.

3.4 RQ 3: Predicting CBE Diagnosis from VIQ, NVIQ and ADOS-2 Factor Scores

Prior to conducting the logistic regression, data were checked for potential biases. Inspection for outliers revealed no significant outliers in performances on VIQ, NVIQ, or any of the ADOS-2 based factor scores. Analysis of potentially influential cases revealed three cases with problematic residual statistics, i.e., Cook's distance > 1, and large standardized residuals (i.e., -43.29, -6.41, and 2.71). For these cases, predicted group membership (i.e., CBE diagnosis) by the model was inconsistent with observed CBE diagnosis, which likely contributed to the problematic residual statistic values. There were also 15 cases with large leverage values (i.e.,

three times the expected value of 0.032, see Stevens, 2002). Because leverage values are measured on the outcome and do not influence the regression coefficients, and all other residual statistics were within normal limits for these 15 cases, they were not of significant concern. Collinearity statistics revealed that all tolerance values were > .01 and all Variance Inflation Factor (VIF; Belsley, 1984) values were < 10, indicating no major problems with multicollinearity of the predictors (Field, 2013). Therefore, I proceeded with the logistic regression.

A binary logistic regression predicting CBE diagnosis from VIQ, NVIQ, general autism factor scores, and SA and RRB factor scores, was significant, χ^2 (5) = 165.68, $p < .001$, see Table 3.7. Effect sizes were large, $R^2 = .59$ (Cox & Snell) and $R^2 = .83$ (Nagelkerke). The overall accuracy of diagnostic classification from the model was 96.3%. Four children without autism were inaccurately predicted by the model as having autism, and three with autism were inaccurately predicted as not having autism. There was a significant effect for general autism factor scores (*OR* = 102.42 [95% CI: 21.97, 477.44], $p < .001$), and for RRB factor scores (*OR* = 5.60 [95% CI: 1.98, 15.81], $p = .001$), suggesting that children with higher general autism and RRB symptom severity were more likely to receive a diagnosis of autism. SA factor scores, VIQ and NVIQ did not make significant contributions to the likelihood of receiving an autism diagnosis above and beyond the general autism and RRB factor scores, *OR* = 0.74 (95% CI: 0.27, 2.03), *OR* = 1.01 (95% CI: 0.95, 1.08), $p = .68$, and *OR* = 1.00 (95% CI: 0.94, 1.06), $p = .92$, respectively. Additionally, the agreement between autism diagnosis according to ADOS-2 cutoff scores and CBE diagnosis by a clinician was 95%. Eight children scored above the cutoff for autism on the ADOS-2 and did not receive a final CBE diagnosis of autism, and one child scored below the ADOS-2 cutoff for autism but did receive a final CBE diagnosis of autism.

Table 3.7 Regression Predicting CBE Diagnosis from Cognition and Autism Factor Scores

Predictor	***β***	***df***	***p*-value**
General Autism Factor	4.63	1	<.001
SA Factor	-0.31	1	.55
RRB Factor	1.72	1	.001
VIQ	-0.003	1	.92
NVIQ	0.01	1	.68

Note. N = 187 for VIQ, 188 for all other variables. VIQ = Verbal IQ, NVIQ = Nonverbal IQ. The outcome variable for the logistic regression was clinical best-estimate (CBE) diagnosis, where 0 = non-autism and 1 = autism.

3.5 RQ 4: Relations between Cognition and Autism across Racial Groups

Spearman's rho correlations revealed significant, moderate correlations between VIQ and NVIQ in both racial groups. There were no statistically significant correlations between VIQ or NVIQ and any of the ADOS-2 total scores in either racial group, see Table 3.8. Overall, correlations were similar in magnitude and direction between racial groups. There were two notable, albeit small and non-significant differences between groups. Firstly, ADOS-2 SA Total scores and NVIQ were negatively correlated for Black children but positively correlated for White children, suggesting that as nonverbal cognitive abilities decrease, social affect symptom severity tends to increase for Black children but decrease for White children. Secondly, correlations between VIQ and all ADOS-2 total scores were slightly larger for Black compared to White children. There were no significant group differences between White and Black children with regard to age, NVIQ, ADOS-2 Overall Total, SA Total, or RRB Total scores. In contrast, VIQ was significantly lower for Black compared to White children in this sample, $t(163) = -2.23$, $p = .03$.

Table 3.8 Correlations Between VIQ, NVIQ, and ADOS-2 Totals Across Racial Groups

	1	2	3	4	5
Black (N = 45)					
1. VIQ	1.00				
2. NVIQ	.62**	1.00			
3. ADOS-2 SA Total	-.17	-.09	1.00		
4. ADOS-2 RRB Total	.11	.17	.43*	1.00	
5. ADOS-2 Overall Total	-.09	.01	.94**	.69**	1.00
White (N = 120)					
1. VIQ	1.00				
2. NVIQ	.57**	1.00			
3. ADOS-2 SA Total	-.01	.04	1.00		
4. ADOS-2 RRB Total	.05	.15	.51**	1.00	
5. ADOS-2 Overall Total	-.01	.05	.94**	.75**	1.00

Note. VIQ = Verbal IQ, NVIQ = Nonverbal IQ, SA = Social Affect, RRB = Restricted and Repetitive Behavior. ADOS-2 Total scores are raw scores, VIQ and NVIQ are standard scores (M = 100, SD = 15).

*$p < .05$, **$p < .001$.

4 DISCUSSION

The purpose of the current study was to investigate several aspects of the measurement and diagnosis of autism in a clinic-referred sample of children being evaluated for possible autism. Specifically, I examined the dimensionality of the ADOS-2, relations between cognition, autism symptomatology and clinical best-estimate diagnosis of autism, and differences in relations between cognition and autism symptomatology between two racial groups. Results confirmed some, but not all initial hypotheses. I hypothesized that a bifactor model with additional non-algorithm ADOS-2 items would demonstrate the best fit. Results revealed that a bifactor model fit best, but with algorithm items only. In contrast to my initial hypotheses, VIQ and NVIQ did not significantly predict any ADOS-2 based latent factors. Alternatively, in support of my hypotheses, increased autism symptom severity as measured by the ADOS-2 predicted a higher likelihood of receiving an autism diagnosis, and cognitive abilities did not

influence diagnostic outcomes above and beyond ADOS-2 performances. Lastly, in support of hypotheses, most relations between cognition and ADOS-2 performances were similar in magnitude and direction for Black and White children.

4.1 Dimensionality of the ADOS-2

A foundational step in improving the diagnostic accuracy of autism is confirming the structure of autism traits measured by autism-specific assessment tools. The ADOS-2 is considered a gold standard clinician-administered assessment for autism and is one of the most regularly utilized autism measures across clinical and research settings. Thus, it is critical to confirm the optimal dimensionality of autism traits measured by the ADOS-2 to inform how to interpret ADOS-2 performances for differential diagnosis, and ultimately for treatment planning (Kim et al., 2018; Plomin et al., 2009). The present study examined several possible factor structures to represent the dimensionality of the ADOS-2, Module 3. Comparison of a two- and three-factor correlated-factors model revealed that a single RRB factor fit the data better than when RRB items were subdivided into insistence on sameness (IS) and repetitive sensorimotor (RSM) factors. This is disappointing considering evidence in the literature that RRB's can be conceptually divided into IS and RSM subdomains, which exhibit unique relations with other child characteristics including cognitive abilities (Bishop et al., 2013; Leekam et al., 2011; Richler et al., 2010). However, it is not entirely surprising that data from ADOS-2 performances alone did not provide a strong measurement of RRB subdomains; the ADOS-2 is administered in a time-limited environment where RRB behaviors that can be more context-dependent, e.g., hand flapping when excited, may not be observed during administration (Hus et al., 2014). Additionally, prior studies have suggested that ADOS-2 algorithm items do not measure RRB's well (Kuhfeld & Sturm, 2018). Findings from the present study extend those by Kuhfeld et al.

(2018) and suggest that subdomains of RRB's may not be captured well by the ADOS-2, at least for children who are assessed with Module 3.

Comparisons of correlated-factors, unidimensional and bifactor models revealed that a bifactor model with only ADOS-2 algorithm items (Bifactor Model 2) was the best-fitting structure for the data. While the application of bifactor models to psychological and behavioral constructs is a relatively new approach (Rodriguez et al., 2016a), these models have demonstrated utility in resolving questions about the multidimensionality of these constructs as a result of their unique ability to simultaneously measure variances of a general trait, and variances of trait subcomponents having accounted for variance explained by a general trait (e.g., Ebesutani et al., 2011; Reise et al., 2007). Previous literature has revealed inconsistencies in the best-fitting factor structure of ADOS data, as well as issues with multidimensionality such as high correlations between domain factors (see ADOS-2 manual, p. 244; Mandy & Skuse, 2008). Thus, the bifactor model offers a possible solution to these challenges. The superior fit of the bifactor model in the present study suggests that there is a general, overarching autism factor that explains most of the variance in item performances, which would explain why correlations between domain factors are often high. This also fits with previous studies identifying a single, continuous autism trait (Constantino et al., 2004). Yet, the strong fit of this model also shows that autism has two unique subcomponents (i.e., the SA and RRB specific factors), which supports our current conceptualization of autism as comprising deficits in two separate behavioral domains, social communication and RRB's (DSM-5, 2013). The present finding of meaningful, unique components of autism traits fits with evidence that social communication deficits and RRB's likely arise from different genetic pathways and neurobiological mechanisms,

and relate differently to other child characteristics like cognition (Happé & Ronald, 2008; Mandy & Skuse, 2008; Ronald et al., 2010).

Ancillary analyses of Bifactor Model 2 also revealed that ADOS-2 algorithm item scores can be considered essentially unidimensional; the specific factors accounted for only a small proportion of item variance (21%) compared to the general autism factor (79%), and the average difference between parameter estimates for the general factor yielded by the unidimensional model compared to Bifactor Model 2 was small (7%). This finding supports the permissibility of utilizing the ADOS-2 Overall Total score (i.e., the sum of the SA and RRB Total scores) as a composite score. In current clinical practice, the ADOS-2 Overall Total is used to calculate cutoff scores for autism; findings from the present study confirm that this is an appropriate use of the Overall Total score, and that the ADOS-2 is a valid measure of overall autism.

An unexpected pattern was revealed in comparing bifactor models with and without ADOS-2 non-algorithm items. According to ancillary analyses, Bifactor Model 2 (only algorithm items) demonstrated consistent evidence of being essentially unidimensional, whereas Bifactor Model 1 (with non-algorithm items A2, A3 and D5) did not. The general autism factor accounted for 79% of the variance in item performance for Bifactor Model 2, and only 64% for Bifactor Model 1. In line with this finding, investigation of IECV values revealed that a greater proportion of items in Bifactor Model 2 (9 out of 14 items; 64%) largely represented the general autism factor than in Bifactor Model 1 (5 out of 17 items; 29%). In looking at specific items, I found that variances for all ADOS-2 RRB algorithm items (i.e., A4, D1, D2, and D4) changed from being largely attributable to the specific RRB factor in Bifactor Model 2, to primarily reflecting the general autism factor in Bifactor Model 1. Interestingly, the only RRB item that was more representative of the RRB specific factor than the general autism factor in Bifactor

Model 1 was item D5: Compulsions or Rituals. Item D5 is a non-algorithm item that has been conceptualized as a more direct measure of the RRB subdomain of insistence on sameness compared to other items (e.g., Bishop et al., 2013). One interpretation of this pattern of findings is that item D5 represents a conceptually different subdomain of RRB behaviors (i.e., insistence on sameness) than the RRB algorithm items, which may alternatively represent the other common subdomain of RRB's, repetitive sensorimotor behaviors. If this is the case, it could be concluded that the RRB Total score on the ADOS-2 is more closely akin to a measure of repetitive sensorimotor behaviors than RRB's in general, underscoring a specific area of autism symptomatology that is not well represented in the ADOS-2 algorithm scoring.

4.2 Relations between Cognition, Sex, Age and Autism Symptomatology

In contrast to initial hypotheses, results revealed no significant prediction of either VIQ or NVIQ on any of the ADOS-2 based latent factors. This finding is somewhat surprising considering the extensive body of literature describing the influence of child characteristics, like cognition, on the presentation and measurement of autism symptoms on several autism-specific assessments including the ADOS (American Psychiatric Association, 2013; Bill & Geschwind, 2009; Goldstein & Ozonoff, 2018; Mayes & Calhoun, 2011; Moldin & Rubenstein, 2006; Ozonoff et al., 2005; Shulman et al., 2020; Wiggins et al., 2019). Though, the current findings are consistent with a study by Havdahl et al. (2016) that demonstrated less influence of cognitive variables on ADOS performances compared to other autism symptom measures. The current findings appear to positively reflect the psychometric modifications that were made to the updated ADOS algorithms (Gotham et al., 2007) to better control for the effects of child characteristics including cognitive abilities. However, even with the updated algorithms, some previous studies have still found that ADOS raw total scores were significantly predicted by VIQ

(De Bildt et al., 2011; Gotham et al., 2009; Shumway et al., 2012), and in fewer cases also by NVIQ (Hus et al., 2014). Comparison of study sample characteristics between prior studies and the present study may explain differences in findings. For example, Shumway et al. (2012) found that in a sample of children with and without autism, when children with non-spectrum delays were included in analyses, the amount of variance in ADOS raw total scores that was accounted for by verbal developmental quotient (VDQ) decreased. Shumway et al. (2012) also found that correlations between language measures (e.g., expressive, and receptive vocabulary) and ADOS total scores were weaker for children with non-spectrum delays than for children with autism. Thus, it is possible that the non-significant relation between VIQ on ADOS-2 latent factors in the present study was partly driven by weaker relations in the non-spectrum group (N = 58). However, this may not be the most likely explanation, as the proportion of non-spectrum children in this sample was relatively small (i.e., 31%). It is more likely that discrepant findings between this study and prior studies are related to more notable differences in study design, e.g., I used latent factors instead of observed raw scores to model the prediction of cognitive variables on autism symptomatology. Also, only children who were administered Module 3 of the ADOS-2 were included, whereas many prior studies have included children who were assessed with multiple modules. In the ADOS-2 validation sample, verbal mental age was more strongly associated with ADOS-2 Total scores for children who were administered Module 1 than for children who were administered other modules; this supports the non-significant prediction of VIQ on ADOS-2-based latent factors in the present sample of children assessed with only Module 3. Children assessed with Module 3 were older and had higher language levels than children administered lower modules, so it is important that the present conclusions of non-significant relations between cognition and ADOS-2 performances be restricted to this specific

subsample of children on the spectrum that excludes very young children, children with more significant language and cognitive impairments, and adults.

Investigation of sex and age as covariates revealed significant effects of both on the general autism and SA latent factors, but not on the RRB latent factor. Age was positively associated with the general autism factor, suggesting that older children exhibited more severe general autism symptom severity. This finding is surprising considering prior research demonstrating that more severe autism symptom severity is associated with a younger age (ADOS-2 manual; Havdahl et al., 2016; Mayes & Calhoun, 2011). One possible explanation is that this trend is unique to the present sample that was restricted in age (i.e., age 4 years and older) due to the inclusion of only children administered Module 3 of the ADOS-2. This would then suggest that in a group of older, generally school aged children, symptom severity tends to be greater for those who are older, whereas when children under the age of 4 are included, younger age is associated with greater symptom severity. In line with this idea, Shumway et al. (2012) found that age negatively predicted ADOS Total raw scores in a sample of children as young as age 2 years who were assessed with ADOS Modules 1 to 3. Prior research has also demonstrated that the inclusion of children without autism increases the effect of age on ADOS scores, such that age is a significant predictor in samples with and without children with autism but is not a significant predictor in samples of children with autism only (Gotham et al., 2009). Thus, the inclusion of children with and without autism in the current sample could also partly explain the significant effect of age on the general autism factor.

Regarding participant sex, being male compared to being female was associated with higher levels of the general autism factor. There was also a larger proportion of males (N = 128, 68%) than females included in the present study. These results are in line with well-replicated

findings in the literature of higher rates of diagnosis, and more severe overall autism symptoms for males than females (Constantino et al., 2020; Kirkovski et al., 2013; Loomes et al., 2017; Maenner et al., 2020; McFayden et al., 2020). Previous findings are more mixed regarding specific differences in RRB and social communication symptoms between boys and girls, with the most consistent pattern being greater RRB symptom severity in boys compared to girls (McFayden et al., 2020; Rubenstein et al., 2015; Van Wijngaarden-Cremers et al., 2014). In the present study I only found a significant effect of sex on the SA, and not the RRB latent factor. Despite some literature to the contrary (Lai et al., 2017; Rubenstein et al., 2015; Sedgewick et al., 2016), a recent study by Kaat et al. (2021) did demonstrate more severe social communication symptoms reported on the SRS by parents of girls than boys. However, ADOS-based social communication scores did not vary by sex in this same study. Notably, Kaat et al. (2021) found only very small effect sizes for sex differences on some but not all domains across multiple standardized measures of autism symptoms. Similarly, effect sizes for sex on the autism factors in the present study were very small; sex accounted for only 3% and 4% of the variance in the general autism factor and SA factor, respectively. In sum, the results from this study suggested that being male was associated with greater general autism symptom severity, and being female was associated with greater SA symptom severity, although these differences were small and not clinically meaningful.

4.3 Predicting CBE Diagnosis from VIQ, NVIQ and ADOS-2 Factor Scores

In support of initial hypotheses, more autism symptoms as measured by the ADOS-2 predicted a higher likelihood of receiving an autism diagnosis, and VIQ and NVIQ did not significantly predict diagnostic category above and beyond the influence of ADOS-2 scores. Results revealed that children's scores on the general autism factor had the largest influence on

their diagnostic outcomes (i.e., relative to contributions by SA and RRB specific factors and cognitive abilities). This finding is consistent with prior research that has demonstrated a strong association between overall performance on standardized autism assessments and clinical diagnosis of autism (Constantino et al., 2003; Frazier et al., 2014).

Regarding SA and RRB specific factor scores, results revealed that even accounting for general autism factor scores, RRB scores still made a significant, unique contribution to CBE diagnosis, while SA scores did not. In the ADOS-2 validation sample, raw scores on SA and RRB domains both made significant contributions to the prediction of autism diagnosis; however, coefficients and z-scores were larger for the SA domain. The discrepancy between the ADOS-2 validation study and the present study suggests that SA domain scores may contribute more strongly to meeting diagnostic cutoffs for autism on the ADOS-2, whereas when clinical judgment is incorporated and CBE diagnosis is the outcome, RRB behaviors seem to have a stronger influence on final diagnosis. This finding may be related to the older age of the current sample, and previous findings that some RRB's are observed less in older children (Esbensen et al., 2009) and lessen in degree of severity over time, particularly following early intervention services (Leekam et al., 2011). It is thus possible that the persistence of RRB's at older ages in the current sample may have signaled to clinicians potentially more severe or untreated autism, resulting in greater RRB symptom severity playing a particularly influential role in diagnostic decision making.

To further explore relations between ADOS-2 performances and CBE diagnosis, I investigated the agreement between autism diagnosis according to ADOS-2 cutoffs and CBE diagnosis. Results revealed very high agreement (95%), which is slightly higher than what has been reported in prior studies, e.g., approximately 75% in a study by Mazefsky & Oswald

(2006). One explanation for this finding is that the present sample is from an autism clinic where the same clinicians who administered the ADOS-2 also made the final clinical diagnosis. Prior research suggests that autism clinic samples tend to produce greater predictive validity of the ADOS in matching clinical diagnoses (Charman & Gotham, 2013). Despite the high agreement, there were several cases where ADOS-2 cutoffs and CBE diagnosis did not align (N = 9). Consistent with previous studies, most of these cases were false positives on the ADOS-2, i.e., children met criteria on the ADOS-2 but were not given a CBE diagnosis of autism (Kamp-Becker et al., 2018). Also consistent with prior research, most of these children had low levels of autism symptom severity (i.e., ADOS-2 comparison scores in the "low" range), suggesting that clinicians may feel less confident in diagnosing autism for children who present with milder symptoms (Bölte & Poustka, 2004; Kamp-Becker et al., 2018; Lord et al., 2006). Alternatively, there were three children whose symptom severity did fall within in the "moderate" to "high" ranges but who were not given a CBE diagnosis of autism. These children all demonstrated a large split in their SA and RRB domain scores such that their Overall Total scores were almost exclusively driven by high scores in the SA domain. This pattern likely explains why these children met criteria for autism on the ADOS-2, i.e., they had high total scores. However, the extreme unevenness between domain scores signaled to clinicians that those children's social communication challenges were related to other disorders, as the presence of RRB symptoms in addition to social communication deficits is required for a diagnosis of autism. The majority of these children also demonstrated significant emotional and/or behavioral difficulties during their evaluations (e.g., anxiety, hyperactivity), which is in line with prior research demonstrating lower clinician confidence regarding an autism diagnosis for children who demonstrate significant emotional and behavioral challenges (Lord et al., 2006).

Lastly, the results of the present study revealed that in this clinic-referred sample of children, cognitive abilities did not play a strong role in diagnostic decision-making, at least not above and beyond ADOS-2 performance. On one hand, this finding could be interpreted to suggest that children's cognitive abilities were not heavily factored into clinical judgment during the diagnostic decision-making process. An alternative interpretation is that due to heterogeneity in the range of cognitive profiles of children with autism, there was no one linear trend of higher or lower cognitive performances that increased the likelihood of clinicians making a diagnosis of autism. This interpretation is consistent with previous literature highlighting variability in cognitive profiles within autism and compared to other typical and neurodevelopmental populations (Courchesne et al., 2019; Farley et al., 2009; Filipek et al., 2000; Joseph et al., 2002).

4.4 Relations between Cognition and Autism across Racial Groups

Initial hypotheses that cognitive abilities would relate to ADOS total score performances similarly across two racial groups (Black and White children) were largely confirmed. Overall, correlations between cognitive abilities and ADOS scores were small and non-significant in both groups. Although relations between cognitive abilities and autism symptoms have not yet been investigated between racial groups, the present findings are in line with prior research that race is not a significant predictor of autism symptomatology, and thus substantial differences between cognition and autism symptoms wouldn't be expected by race (Hus & Lord, 2013; Mayes & Calhoun, 2011). However, two small, qualitative differences in relations between cognitive abilities and autism symptoms between the two racial groups were observed. First, as NVIQ performances decreased, Black children tended to demonstrate higher, or worse social affect symptom severity, whereas White children demonstrated less. Second, VIQ was generally more

strongly related to ADOS-2 total scores for Black compared to White children. Additionally, average VIQ was significantly lower for Black than White children in this sample. These results are consistent with findings from prior research that relative to White and Non-Hispanic White children, Black children with autism tend to have higher rates of co-morbid ID (Broder-Fingert et al., 2013; Constantino et al., 2020; Klaiman et al., 2015; Mandell et al., 2009). The negative relation between NVIQ and SA Total scores illustrates that for Black children in this study, those with lower nonverbal cognition were characterized as having greater problems with social affect, which would be consistent with higher rates of co-morbid intellectual disabilities in Black compared to White children diagnosed with autism.

4.5 Limitations

There are several limitations in the present study that are important to acknowledge. First, when the structural regression model with VIQ, NVIQ, and all ADOS-2 based latent factors predicting CBE diagnosis was initially conducted in Mplus, CBE diagnosis produced a negative residual variance. Multiple modifications to the model failed to resolve this issue, which suggests model misspecification or poor model fit. This issue could have resulted from one of several potential limitations of the data. The CBE diagnosis outcome variable was very positively skewed; 69% (N = 130) of the children received an autism diagnosis, i.e., were labeled a "1." This left a relatively small number of children labeled "0" for non-autism (N = 58 children, 31%). Additionally, within this already small subgroup of children labeled non-autism, diagnoses spanned several different types of other neurodevelopmental and mood disorders, as well as three children who did not receive any mood or behavioral diagnoses. Thus, a larger and/or more homogenous non-autism group may have improved the fit of the structural regression model with CBE diagnosis as the outcome variable. Another possible cause of model misspecification may

have been related to the exclusion of other important predictors of CBE diagnosis in the model. Prior research has demonstrated that externalizing and internalizing symptoms frequently co-occur with autism and influence the presentation and measurement of autism symptomatology (Frazier et al., 2014; Hus & Lord, 2013; Klaiman et al., 2015; M. C. Lai et al., 2019; Simonoff et al., 2008). Emotional and behavioral difficulties are also especially relevant to school-age children with autism who often have high rates of co-occurring disorders, increased awareness of their differences and experience higher incidences of bullying (Dawson, 2008; Shulman et al., 2020; Thomas & Karmiloff-Smith, 2002; Vivanti et al., 2013). Therefore, it is possible that the structural regression model might have been strengthened by the addition of internalizing and externalizing symptoms as predictors. Ideally, I would have included these variables in the present study, but this was not possible due to limited availability of emotional and behavioral data in the retrospective chart review.

A second major limitation was that the current sample included a very specific subset of children who were referred for autism evaluations at relatively "later" ages and who were assessed with a Module 3 of the ADOS-2 that requires children to be "verbally fluent." This inclusion criteria restricted the range of verbal abilities in the sample, and subsequently also restricted the range of cognitive abilities. While a range of cognitive abilities were still represented in the sample, most children performed in the average to above average range. Only 12% (N = 23) of children had VIQ or NVIQ scores ≤ 75. The purpose of restricting the study sample to only children assessed with Module 3 was to include item D5 (not included in Modules 1 or 2). Item D5 has been considered one of the only conceptually "pure" items of the IS subdomain of RRB's (Bishop et al., 2013) included across ADOS-2 modules. Thus, including item D5 was important for investigating if the ADOS-2 dimensionality could comprise two

separable factors of RRB subdomains. Unfortunately, the consequence of this study design was a restricted representation of children referred for possible autism that was biased towards children with higher verbal and cognitive abilities. Thus, it is possible that trends found in the present study regarding relations between cognition, autism symptomatology and CBE diagnosis may be different for children with more significant cognitive and/or language deficits and are not generalizable to children across the whole autism spectrum.

Another limitation of this study was that racial group analyses were limited to only children identified by their parents as Black or White. Ideally, a more diverse range of racial and ethnic identities would have been included. However, the retrospective chart review data used in this study did not have enough participants in other racial and ethnic groups to be included in statistical analyses. The resulting sample size for racial group analyses was also relatively small, particularly for Black children (N = 46, 24%), which may have limited our ability to detect significant difference in the relations between cognition and autism symptomatology between the two groups. Additionally, due to data limitations the present study did not control for salient socioenvironmental variables that are known to mediate the roles between race and cognition when analyzing racial group differences, so the findings from this study are preliminary at best and should be interpreted as a starting point for future investigations rather than definitive conclusions about how cognition and autism symptomatology relate within these groups.

4.6 Implications and Future Directions

Despite these limitations, findings from the present study have several important implications. First, regarding autism measurement, findings from this study underscored previous concerns that the ADOS-2 is not a strong measure of RRB's, particularly in terms of its ability to detect unique subdomains of RRB's. Clinically, this highlights the importance of using

other measures to collect additional information about restricted and repetitive behaviors, e.g., the ADI-R, beyond what is gathered on the ADOS-2 during an autism evaluation. Failing to fully describe a child's RRB's could result in missing important treatment referrals, as RRB's are known to have a significant impact on functioning in daily life and are often barriers to learning (Leekam et al., 2011).

Secondly, support for a bifactor model of autism on the ADOS-2, Module 3 in the present study has exciting implications for future research. Use of a bifactor model in future studies may help to reduce previous inconsistencies in the literature regarding the factor structure of the ADOS. The bifactor model also provides a way for future studies to investigate relations between important demographic and environmental variables with both a general overarching autism trait and its unique subcomponents. This is particularly relevant for studies of autism biomarkers and genetic etiologies, which are often hypothesized to relate to specific domains of autism symptomatology rather than a unitary autism trait (Hus et al., 2014). Clinically, the unidimensionalilty of the best-fitting bifactor model indicates that when clinicians rate a child higher on most ADOS-2 algorithm items, they have observed deficits that are attributable to an overarching, general autism trait. However, for some items, particularly item A8: Conversations and D1: Unusual Sensory Interests in Play Materials/Persons, assigning a higher score is more indictive of difficulties associated with specific SA or RRB traits. Importantly, because these specific traits are represented as independent by the bifactor model, it could suggest that performance on items that relate more to specific SA and RRB components than the general autism factor have unique etiological pathways, and thus difficulties measured by these items could respond differently to interventions. Future studies should investigate the fit of bifactor

models for the other ADOS-2 modules and explore if SA and RRB specific factors are associated with different biomarkers or treatment outcomes.

Findings regarding the non-significant prediction of VIQ and NVIQ on ADOS-2 performances provides continued support for the ADOS's internal validity in this unique clinic-referred sample of children who were assessed with Module 3. Unfortunately, due to data limitations, i.e., the negative residual variance for CBE diagnosis described earlier, I was not able to analyze the concomitant influences of VIQ, NVIQ, and autism symptomatology using SEM as originally planned. Thus, the benefits of this statistical approach such as modeling measurement error were lost. Therefore, future studies should replicate these analyses with larger sample sizes that comprise larger groups of children with non-spectrum disorders and/or typical development. This may allow for the use of SEM that was not feasible in the present study. Additionally, it would be interesting to examine if emotional and behavioral functioning predict CBE diagnosis and contribute to the clinical decision-making process. Future studies could also examine more homogenous non-autism groups, e.g., children with ADHD or mood disorders alone, to explore if there are unique predictors of CBE diagnosis when differential diagnosis is between autism and a specific neurodevelopmental or mood disorder. Lastly, replication of these analyses with the other ADOS-2 modules would provide helpful information about how the relations between cognition, autism symptomatology, and CBE diagnosis may differ for children of all ages and language abilities.

Finally, despite the small sample size and limited diversity of racial groups, the results of this study have implications for racial disparities in autism diagnosis. The stronger relation between cognitive and social communication deficits that was identified for Black compared to White children in this study may indicate that deficits in these two domains are being

characterized as less distinct in Black compared to White children. This phenomenon could be related to biases in how performances are scored, or to other mediating socioenvironmental variables (e.g., parental education, parental academic expectations, and SES) that have already been shown to influence cognitive functioning differentially across racial groups (Broder-Fingert et al., 2013; Weiss & Saklofske, 2020). Relations between cognition and autism symptomatology across racial groups should be examined in future studies with larger and more racially and ethnically diverse sample sizes to build upon these findings.

4.7 Conclusion

In conclusion, results from this study make several important contributions to understanding and improving the measurement and diagnosis of autism. This study demonstrated the excellent fit and utility of a bifactor model to characterize autism symptomatology on the ADOS-2, Module 3. Bifactor models should be employed in future studies investigating relations between ADOS-2 performances and other variables of interest. The internal validity of the ADOS-2 algorithm items to measure autism with little impact from a children's cognitive functioning was also confirmed. Alternatively, this study identified that nuanced RRB's are not well-measured by the ADOS-2. Thus, it is important that clinicians gather additional data from other measures about children's functioning in this domain during an autism evaluation. Regarding the clinical diagnostic decision-making process, in a sample of children referred for possible autism, I found that a general autism trait measured by the ADOS-2 had the largest influence on diagnostic outcomes, and cognitive functioning had little influence. However, it will be important to replicate these findings in larger and more diagnostically diverse samples. Lastly, this study revealed preliminary evidence for a possible mechanism contributing to racial

disparities in the identification of autism in Black compared to White children that warrants continued investigation.

REFERENCES

Abbeduto, L., McDuffie, A., & Thurman, A. J. (2014). The fragile X syndromeâ€“autism comorbidity: What do we really know? *Frontiers in Genetics*, *5*(SEP), 355–355. https://doi.org/10.3389/fgene.2014.00355

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders. *Arlington: American Psychiatric Publishing*.

Ankenman, K., Elgin, J., Sullivan, K., Vincent, L., & Bernier, R. (2014). Nonverbal and verbal cognitive discrepancy profiles in autism spectrum disorders: Influence of age and gender. *American Journal on Intellectual and Developmental Disabilities*, *119*(1), 84–99. https://doi.org/10.1352/1944-7558-119.1.84

Belsley, D. A. (1984). Demeaning conditioning diagnostics through centering. *The American Statistician*, 38, 73–93.

Bill, B. R., & Geschwind, D. H. (2009). Genetic advances in autism: Heterogeneity and convergence on shared pathways. *Current Opinion in Genetics & Development*, *19*(3), 271–278.

Bishop, S. L., Hus, V., Duncan, A., Huerta, M., Gotham, K., Pickles, A., Kreiger, A., Buja, A., Lund, S., & Lord, C. (2013). Subcategories of restricted and repetitive behaviors in children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(6), 1287–1297. https://doi.org/10.1007/s10803-012-1671-0

Bishop, S. L., Richler, J., & Lord, C. (2006). Association between restricted and repetitive behaviors and nonverbal IQ in children with autism spectrum disorders. *Child Neuropsychology : A Journal on Normal and Abnormal Development in Childhood and Adolescence*, *12*(4–5), 247–267. https://doi.org/10.1080/09297040600630288

Black, D. O., Wallace, G. L., Sokoloff, J. L., & Kenworthy, L. (2009). Brief report: IQ split predicts social symptoms and communication abilities in high-functioning children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(11), 1613–1619. https://doi.org/10.1007/s10803-009-0795-3

Bodfish, J. W., Symons, F. J., Parker, D. E., & Lewis, M. H. (2000). Varieties of repetitive behavior in autism: Comparisons to mental retardation. *Journal of Autism and Developmental Disorders*, *30*(3), 237–243.

Bollen, K. A. (1989). *Structural equations with latent variables* (Atlanta Library North 5 HA29 .B732 1989). Wiley. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=cat06552a&AN=gsu.996622243402952&site=eds-live&scope=site&custid=gsu1

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models* (Atlanta Library North 5 H61.25 .T47 1993). Sage Publications. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=cat06552a&AN=gsu.997243213402952&site=eds-live&scope=site&custid=gsu1

Bölte, S., & Poustka, F. (2004). Diagnostic Observation Scale for Autistic Disorders: Initial results of reliability and validity. *Zeitschrift Fur Kinder-Und Jugendpsychiatrie Und Psychotherapie*, *32*(1), 45–50.

Broder-Fingert, S., Shui, A., Pulcini, C. D., Kurowski, D., & Perrin, J. M. (2013). Racial and ethnic differences in subspecialty service use by children with autism. *Pediatrics*, *132*(1), 94–100.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. Guilford publications.

Charman, T., & Gotham, K. (2013). Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders—Lessons from research and practise. *Child and Adolescent Mental Health*, *18*(1), 52–63. https://doi.org/10.1111/j.1475-3588.2012.00664.x

Charman, T., Pickles, A., Simonoff, E., Chandler, S., Loucas, T., & Baird, G. (2011). IQ in children with autism spectrum disorders: Data from the Special Needs and Autism Project (SNAP). *Psychological Medicine*, *41*(03), 619–627.

Constantino, J. N., Abbacchi, A. M., Saulnier, C., Klaiman, C., Mandell, D. S., Zhang, Y., Hawks, Z., Bates, J., Klin, A., Shattuck, P., Molholm, S., Fitzgerald, R., Roux, A., Lowe, J. K., & Geschwind, D. H. (2020). Timing of the diagnosis of autism in african American children. *Pediatrics*, *146*(3). https://doi.org/10.1542/PEDS.2019-3629

Constantino, J. N., Davis, S. A., Todd, R. D., Schindler, M. K., Gross, M. M., Brophy, S. L., Metzger, L. M., Shoushtari, C. S., Splinter, R., & Reich, W. (2003). Validation of a brief quantitative measure of autistic traits: Comparison of the social responsiveness scale with the autism diagnostic interview-revised. *Journal of Autism and Developmental Disorders*, *33*(4), 427–433.

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

Constantino JN, Gruber CP. (2005). *Social Responsiveness Scale (SRS).* Western Psychological Services.

Courchesne, V., Girard, D., Jacques, C., & Soulières, I. (2019). Assessing intelligence at autism diagnosis: Mission impossible? Testability and cognitive profile of autistic preschoolers.

Journal of Autism and Developmental Disorders, *49*(3), 845–856. https://doi.org/10.1007/s10803-018-3786-4

Cuccaro, M. L., Shao, Y., Grubber, J., Slifer, M., Wolpert, C. M., Donnelly, S. L., Abramson, R. K., Ravan, S. A., Wright, H. H., & DeLong, G. R. (2003). Factor analysis of restricted and repetitive behaviors in autism using the Autism Diagnostic Interview-R. *Child Psychiatry and Human Development*, *34*(1), 3–17.

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology*, *20*(3), 775–803. https://doi.org/10.1017/S0954579408000370

Dawson, G., Estes, A., Munson, J., Schellenberg, G., Bernier, R., & Abbott, R. (2007). Quantitative Assessment of Autism Symptom-related Traits in Probands and Parents: Broader Phenotype Autism Symptom Scale. *Journal of Autism and Developmental Disorders*, *37*(3), 523–536. https://doi.org/10.1007/s10803-006-0182-2

De Bildt, A., Oosterling, I. J., Van Lang, N. D. J., Sytema, S., Minderaa, R. B., Van Engeland, H., Roos, S., Buitelaar, J. K., Van Der Gaag, R. J., & De Jonge, M. V. (2011). Standardized ADOS scores: Measuring severity of autism spectrum disorders in a Dutch sample. *Journal of Autism and Developmental Disorders*, *41*(3), 311–319. https://doi.org/10.1007/s10803-010-1057-0

Dueber, D. M. (2017). Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models. https://doi.org/10.13023/edp.tool.01 [Available at http://sites.education.uky.edu/apslab/resources]

Dworzynski, K., Happé, F., Bolton, P., & Ronald, A. (2009). Relationship between symptom domains in autism spectrum disorders: A population based twin study. *Journal of Autism and Developmental Disorders*, *39*(8), 1197–1210.

Ebesutani, C., Smith, A., Bernstein, A., Chorpita, B. F., Higa-McMillan, C., & Nakamura, B. (2011). A bifactor model of negative affectivity: Fear and distress components among younger and older youth. *Psychological Assessment*, *23*(3), 679.

Farley, M. A., McMahon, W. M., Fombonne, E., Jenson, W. R., Miller, J., Gardner, M., Block, H., Pingree, C. B., Ritvo, E. R., Ritvo, R. A., & Coon, H. (2009). Twenty-year outcome for individuals with autism and average or near-average cognitive abilities. *Autism Research*, *2*(2), 109–118. https://doi.org/10.1002/aur.69

Filipek, P. A., Accardo, P. J., Ashwal, S., Baranek, G. T., Cook, E. H., Dawson, G., Gordon, B., Gravel, J. S., Johnson, C. P., Kallen, R. J., Levy, S. E., Minshew, N. J., Ozonoff, S., Prizant, B. M., Rapin, I., Rogers, S. J., Stone, W. L., Teplin, S. W., Tuchman, R. F., & Volkmar, F. R. (2000). Practice parameter: Screening and diagnosis of autism. Report of the quality standards subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology*, *55*(4), 468–479. https://doi.org/10.1212/WNL.55.4.468

Finney, S. J. & DiStefano, C. (2013). Nonnormal and Categorical Data in Structural Equation Modeling. In R. O. Mueller & G. R. Hancock (Eds.), *Structural Equation Modeling: A Second Course* (p. 439-492). Information Age Publishing.

Frazier, T. W., Youngstrom, E. A., Embacher, R., Hardan, A. Y., Constantino, J. N., Law, P., Findling, R. L., & Eng, C. (2014). Demographic and clinical correlates of autism

symptom domains and autism spectrum diagnosis. *Autism : The International Journal of Research and Practice*, *18*(5), 571–582. https://doi.org/10.1177/1362361313481506

Frazier, T. W., Youngstrom, E. A., Speer, L., Embacher, R., Law, P., Constantino, J., Findling, R. L., Hardan, A. Y., & Eng, C. (2012). Validation of proposed DSM-5 criteria for autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(1), 28-40.e3. https://doi.org/10.1016/j.jaac.2011.09.021

Gadow, K., & Sprafkin, J. (2005). Child and adolescent symptom inventory-4R. *Stony Brook, NY: Checkmate Plus*.

Goldstein, S., & Ozonoff, S. (2018). *Assessment of autism spectrum disorder*. Guilford Publications.

Gotham, K., Pickles, A., & Lord, C. (2009). Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *39*(5), 693–705. https://doi.org/10.1007/s10803-008-0674-3

Gotham, K., Risi, S., Pickles, A., & Lord, C. (2007). The autism diagnostic observation schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*, *37*(4), 613–627. https://doi.org/10.1007/s10803-006-0280-1

Hammer, J. H., & Toland, M. D. (2016, November). Bifactor analysis in Mplus. [Video file]. Retrieved from http://sites.education.uky.edu/apslab/upcoming-events/

Happé, F., & Ronald, A. (2008). The 'Fractionable Autism Triad': A Review of Evidence from Behavioural, Genetic, Cognitive and Neural Research. *Neuropsychology Review*, *18*(4), 287–304. https://doi.org/10.1007/s11065-008-9076-8

Havdahl, K. A., Hus Bal, V., Huerta, M., Pickles, A., Øyen, A.-S., Stoltenberg, C., Lord, C., & Bishop, S. L. (2016). Multidimensional Influences on Autism Symptom Measures:

Implications for Use in Etiological Research. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(12), 1054-1063.e3. https://doi.org/10.1016/J.JAAC.2016.09.490

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55.

Hus, V., Bishop, S., Gotham, K., Huerta, M., & Lord, C. (2013). Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *54*(2), 216–224. https://doi.org/10.1111/j.1469-7610.2012.02589.x

Hus, V., Gotham, K., & Lord, C. (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders*, *44*(10), 2400–2412. https://doi.org/10.1007/s10803-012-1719-1

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

Hus, V., Pickles, A., Cook Jr, E. H., Risi, S., & Lord, C. (2007). Using the autism diagnostic interview—Revised to increase phenotypic homogeneity in genetic studies of autism. *Biological Psychiatry*, *61*(4), 438–448.

Hyman, S. L., Levy, S. E., & Myers, S. M. (2020). Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. *Pediatrics*, *145*(1). https://doi.org/10.1542/peds.2019-3447

Joseph, R. M., Tager-Flusberg, H., & Lord, C. (2002). Cognitive profiles and social-communicative functioning in children with autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *43*(6), 807–821.

Kaat, A. J., Shui, A. M., Ghods, S. S., Farmer, C. A., Esler, A. N., Thurm, A., Georgiades, S., Kanne, S. M., Lord, C., & Kim, Y. S. (2021). Sex differences in scores on standardized measures of autism symptoms: A multisite integrative data analysis. *Journal of Child Psychology and Psychiatry*, *62*(1), 97–106.

Kamp-Becker, I, Albertowski, K., Becker, J., Ghahreman, M., Langmann, A., Mingebach, T., Poustka, L., Weber, L., Schmidt, H., Smidt, J., Stehr, T., Roessner, V., Kucharczyk, K., Wolff, N., & Stroth, S. (2018). Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *European Child & Adolescent Psychiatry*, *27*(9), 1193–1207. https://doi.org/10.1007/s00787-018-1143-y

Kamp-Becker, Inge, Ghahreman, M., Smidt, J., & Remschmidt, H. (2009). Dimensional structure of the autism phenotype: Relations between early development and current presentation. *Journal of Autism and Developmental Disorders*, *39*(4), 557–571. https://doi.org/10.1007/s10803-008-0656-5

Kanne, S. M., Mazurek, M. O., Sikora, D., Bellando, J., Branum-Martin, L., Handen, B., Katz, T., Freedman, B., Powell, M. P., & Warren, Z. (2014). The Autism Impact Measure (AIM): Initial development of a new tool for treatment outcome measurement. *Journal of Autism and Developmental Disorders*, *44*(1), 168–179.

Kaufman, A. S. (1994). *Intelligence testing the WISC-III*. Wiley.

Kim, H., Keifer, C., Rodriguez-Seijas, C., Eaton, N., Lerner, M., & Gadow, K. (2018). Quantifying the Optimal Structure of the Autism Phenotype: A Comprehensive

Comparison of Dimensional, Categorical, and Hybrid Models. *Journal of the American Academy of Child & Adolescent Psychiatry*. https://doi.org/10.1016/J.JAAC.2018.09.431

Kirkovski, M., Enticott, P. G., & Fitzgerald, P. B. (2013). A review of the role of female gender in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *43*(11), 2584–2603.

Klaiman, C., Fernandez-Carriba, S., Hall, C., & Saulnier, C. (2015). Assessment of Autism Across the Lifespan: A Way Forward. *Current Developmental Disorders Reports*, *2*(1), 84–92. https://doi.org/10.1007/s40474-014-0031-5

Klin, A., Saulnier, C. A., Sparrow, S. S., Cicchetti, D. V., Volkmar, F. R., & Lord, C. (2007). Social and communication abilities and disabilities in higher functioning individuals with autism spectrum disorders: The Vineland and the ADOS. *Journal of Autism and Developmental Disorders*, *37*(4), 748–759. https://doi.org/10.1007/s10803-006-0229-4

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Kuhfeld, M., & Sturm, A. (2018). An examination of the precision of the Autism Diagnostic Observation Schedule using item response theory. *Psychological Assessment*, *30*(5), 656–668. https://doi.org/10.1037/pas0000512

Lai, M. C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., Szatmari, P., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *The Lancet Psychiatry*, *6*(10), 819–829. https://doi.org/10.1016/S2215-0366(19)30289-5

Lai, M.-C., Lombardo, M. V., Ruigrok, A. N., Chakrabarti, B., Auyeung, B., Szatmari, P., Happé, F., Baron-Cohen, S., & MRC AIMS Consortium. (2017). Quantifying and exploring camouflaging in men and women with autism. *Autism*, *21*(6), 690–702.

Leekam, S. R., Prior, M. R., & Uljarevic, M. (2011). Restricted and repetitive behaviors in autism spectrum disorders: A review of research in the last decade. *Psychological Bulletin*, *137*(4), 562–593. https://doi.org/10.1037/a0023341

Leyfer, O. T., Folstein, S. E., Bacalman, S., Davis, N. O., Dinh, E., Morgan, J., Tager-Flusberg, H., & Lainhart, J. E. (2006). Comorbid psychiatric disorders in children with autism: Interview development and rates of disorders. *Journal of Autism and Developmental Disorders*, *36*(7), 849–861.

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(6), 466–474.

Lord, C, Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule–2nd edition (ADOS-2). *Los Angeles, CA: Western Psychological Corporation*.

Lord, Catherine, Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J. H., Jones, R. M., Pickles, A., State, M. W., Taylor, J. L., & Veenstra-VanderWeele, J. (2020). Autism spectrum disorder. *Nature Reviews Disease Primers*, *6*(1), 1–23. https://doi.org/10.1038/s41572-019-0138-4

Lord, Catherine, Petkova, E., Hus, V., Gan, W., Lu, F., Martin, D. M., Ousley, O., Guy, L., Bernier, R., Gerdts, J., Algermissen, M., Whitaker, A., Sutcliffe, J. S., Warren, Z., Klin, A., Saulnier, C., Hanson, E., Hundley, R., Piggot, J., … Risi, S. (2012). A multisite study

of the clinical diagnosis of different autism spectrum disorders. *Archives of General Psychiatry*, *69*(3), 306–313. https://doi.org/10.1001/archgenpsychiatry.2011.148

Lord, Catherine, Risi, S., DiLavore, P. S., Shulman, C., Thurm, A., & Pickles, A. (2006). Autism from 2 to 9 years of age. *Archives of General Psychiatry*, *63*(6), 694–701. https://doi.org/10.1001/archpsyc.63.6.694

Lord, Catherine, Rutter, M., DiLavore, P. C., & Risi, S. (1999). Autism diagnostic observation schedule-WPS (ADOS-WPS). *Los Angeles, CA: Western Psychological Services*.

Lord, Catherine, Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685. https://doi.org/10.1007/BF02172145

LoVullo, S. V., & Matson, J. L. (2009). Comorbid psychopathology in adults with autism spectrum disorders and intellectual disabilities. *Research in Developmental Disabilities*, *30*(6), 1288–1296.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–130.

Maenner, M. J., Shaw, K. A., & Baio, J. (2020). Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries*, *69*(4), 1–1.

Mandell, D. S., Wiggins, L. D., Carpenter, L. A., Daniels, J., DiGuiseppi, C., Durkin, M. S., Giarelli, E., Morrier, M. J., Nicholas, J. S., & Pinto-Martin, J. A. (2009). Racial/ethnic

disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health*, *99*(3), 493–498.

Mandy, W. P. L., & Skuse, D. H. (2008). Research Review: What is the association between the social-communication element of autism and repetitive interests, behaviours and activities? *Journal of Child Psychology and Psychiatry*, *49*(8), 795–808. https://doi.org/10.1111/j.1469-7610.2008.01911.x

Matson, J. L., & Kozlowski, A. M. (2011). The increasing prevalence of autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(1), 418–425. https://doi.org/10.1016/j.rasd.2010.06.004

Mayes, S. D., & Calhoun, S. L. (2011). Impact of IQ, age, SES, gender, and race on autistic symptoms. *Research in Autism Spectrum Disorders*, *5*(2), 749–757. https://doi.org/10.1016/j.rasd.2010.09.002

Mazefsky, C. A., & Oswald, D. P. (2006). The discriminative ability and diagnostic utility of the ADOS-G, ADI-R, and GARS for children in a clinical setting. *Autism*, *10*(6), 533–549.

McCrimmon, A. W., & Smith, A. D. (2013). Review of the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II). *Journal of Psychoeducational Assessment*, *31*(3), 337–341. https://doi.org/10.1177/0734282912467756

McFayden, T. C., Antezana, L., Albright, J., Muskett, A., & Scarpa, A. (2020). Sex Differences in an Autism Spectrum Disorder Diagnosis: Are Restricted Repetitive Behaviors and Interests the Key? *Review Journal of Autism and Developmental Disorders*, *7*(2), 119–126. https://doi.org/10.1007/s40489-019-00183-w

Moldin, S. O., & Rubenstein, J. L. R. (2006). *Understanding autism: From basic neuroscience to treatment.* (Atlanta Library North 3 RC553.A88 U53 2006). CRC/Taylor & Frances;

GSU GIL-Find Catalog. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=cat06552a&AN=gsu.9915114593402952&site=eds-live&scope=site&custid=gsu1

Molloy, C. A., Murray, D. S., Akers, R., Mitchell, T., & Manning-Courtney, P. (2011). Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism*, *15*(2), 143–162. https://doi.org/10.1177/1362361310379241

Moore, V., & Goodson, S. (2003). How well does early diagnosis of autism stand the test of time? Follow-up study of children assessed for autism at age 2 and development of an early diagnostic service. *Autism : The International Journal of Research and Practice*, *7*(1), 47–63. https://doi.org/10.1177/1362361303007001005

Muthén, L. K., & Muthén, B. O. (2005). *Mplus: Statistical Analysis with Latent Variables: User's Guide* (pp. 1998-2012). Los Angeles: Muthén & Muthén.

Nader, A. M., Courchesne, V., Dawson, M., & Soulières, I. (2016). Does WISC-IV Underestimate the Intelligence of Autistic Children? *Journal of Autism and Developmental Disorders*, *46*(5), 1582–1589. https://doi.org/10.1007/s10803-014-2270-z

Norris, M., Lecavalier, L., & Edwards, M. C. (2012). The structure of autism symptoms as measured by the autism diagnostic observation schedule. *Journal of Autism and Developmental Disorders*, *42*(6), 1075–1086. https://doi.org/10.1007/s10803-011-1348-0

Oh, M.-Y. (2016). Psychological Assessment in Neurodevelopmental Disorders. *Hanyang Medical Reviews*, *36*(1), 72–72. https://doi.org/10.7599/hmr.2016.36.1.72

Oliveras-Rentas, R. E., Kenworthy, L., Roberson, R. B., Martin, A., & Wallace, G. L. (2012). WISC-IV profile in high-functioning autism spectrum disorders: Impaired processing speed is associated with increased autism communication symptoms and decreased

adaptive communication abilities. *Journal of Autism and Developmental Disorders*, *42*(5), 655–664. https://doi.org/10.1007/s10803-011-1289-7

Ozonoff, S., Goodlin-Jones, B. L., & Solomon, M. (2005). Evidence-based assessment of autism spectrum disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*(3), 523–540. https://doi.org/10.1207/s15374424jccp3403_8

Parent, V., Birtwell, K. B., Lambright, N., & DuBard, M. (2016). Combining CBT and Behavior-Analytic Approaches to Target Severe Emotion Dysregulation in Verbal Youth With ASD and ID. *Journal of Mental Health Research in Intellectual Disabilities*, *9*(1–2), 60–82. https://doi.org/10.1080/19315864.2016.1166301

Pallant, J. (2013). SPSS survival manual. McGraw-Hill Education (UK).

Plomin, R., Haworth, C. M., & Davis, O. S. (2009). Common disorders are quantitative traits. *Nature Reviews Genetics*, *10*(12), 872–878.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor Models and Rotations: Exploring the Extent to Which Multidimensional Data Yield Univocal Scale Scores. *Journal of Personality Assessment*, *92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*(1), 19–31.

Richler, J., Huerta, M., Bishop, S. L., & Lord, C. (2010). Developmental trajectories of restricted and repetitive behaviors and interests in children with autism spectrum disorders. *Development and Psychopathology*, *22*(1), 55–69. https://doi.org/10.1017/S0954579409990265

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Roid, G. H. (2003). *Standford-Binet Intelligence Scales*. Riverside Publishing.

Ronald, A., Butcher, L. M., Docherty, S., Davis, O. S. P., Schalkwyk, L. C., Craig, I. W., & Plomin, R. (2010). A genome-wide association study of social and non-social autistic-like traits in the general population using pooled DNA, 500 K SNP microarrays and both community and diagnosed autism replication samples. *Behavior Genetics*, *40*(1), 31–31.

Ronald, A., Happé, F., & Plomin, R. (2005). The genetic relationship between individual differences in social and nonsocial behaviours characteristic of autism. *Developmental Science*, *8*(5), 444–458. https://doi.org/10.1111/j.1467-7687.2005.00433.x

Rubenstein, E., Wiggins, L. D., & Lee, L.-C. (2015). A review of the differences in developmental, psychiatric, and medical endophenotypes between males and females with autism spectrum disorder. *Journal of Developmental and Physical Disabilities*, *27*(1), 119–139.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*.

Rutter, M., LeCouteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised (ADI-R)*. Western Psychological Services.

Sedgewick, F., Hill, V., Yates, R., Pickering, L., & Pellicano, E. (2016). Gender differences in the social motivation and friendship experiences of autistic and non-autistic adolescents. *Journal of Autism and Developmental Disorders*, *46*(4), 1297–1306.

Shulman, C., Rice, C. E., Morrier, M. J., & Esler, A. (2020). The Role of Diagnostic Instruments in Dual and Differential Diagnosis in Autism Spectrum Disorder Across the Lifespan. *Child and Adolescent Psychiatric Clinics*.

Shumway, S., Farmer, C., Thurm, A., Joseph, L., Black, D., & Golden, C. (2012). The ADOS Calibrated Severity Score: Relationship to Phenotypic Variables and Stability over Time. *Autism Research*, *5*(4), 267–276. https://doi.org/10.1002/aur.1238

Simonoff, E., Pickles, A., Charman, T., Chandler, S., Loucas, T., & Baird, G. (2008). Psychiatric disorders in children with autism spectrum disorders: Prevalence, comorbidity, and associated factors in a population-derived sample. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(8), 921–929. https://doi.org/10.1097/CHI.0b013e318179964f

Smith, K. R. M., & Matson, J. L. (2010). Social skills: Differences among adults with intellectual disabilities, co-morbid autism spectrum disorders and epilepsy. *Research in Developmental Disabilities*, *31*(6), 1366–1372.

Stevens, J. (2002). *Applied Multivariate Statistics for the Social Sciences: Vol. 4th ed*. Lawrence Erlbaum Associates, Inc. https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,shib&db=nlebk&AN=63477&site=eds-live&scope=site&custid=gsu1

Stone, W. L., Lee, E. B., Ashford, L., Brissie, J., Hepburn, S. L., Coonrod, E. E., & Weiss, B. H. (1999). Can Autism Be Diagnosed Accurately in Children Under 3 Years? *Journal of*

Child Psychology and Psychiatry, *40*(2), 219–226. https://doi.org/10.1111/1469-7610.00435

Stucky, B. D., & Edelen, M. O. (2014). Using hierarchical IRT models to create unidimensional measures from multidimensional data. *Handbook of Item Response Theory Modeling: Applications to Typical Performance Assessment*, 183–206.

Thomas, M., & Karmiloff-Smith, A. (2002). Are developmental disorders like cases of adult brain damage? Implications from connectionist modelling. *Behavioral and Brain Sciences*, *25*(6), 727–750. https://doi.org/10.1017/S0140525X02000134

Van Wijngaarden-Cremers, P. J. M., van Eeten, E., Groen, W. B., Van Deurzen, P. A., Oosterling, I. J., & Van der Gaag, R. J. (2014). Gender and Age Differences in the Core Triad of Impairments in Autism Spectrum Disorders: A Systematic Review and Meta-analysis. *Journal of Autism and Developmental Disorders*, *44*(3), 627–635. https://doi.org/10.1007/s10803-013-1913-9

Visser, J. C., Rommelse, N. N. J., Lappenschaar, M., Servatius-Oosterling, I. J., Greven, C. U., & Buitelaar, J. K. (2017). Variation in the Early Trajectories of Autism Symptoms Is Related to the Development of Language, Cognition, and Behavior Problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, *56*(8), 659–668. https://doi.org/10.1016/J.JAAC.2017.05.022

Vivanti, G., Barbaro, J., Hudry, K., Dissanayake, C., & Prior, M. (2013). Intellectual Development in Autism Spectrum Disorders: New Insights from Longitudinal Studies. *Frontiers in Human Neuroscience*, *7*(JUL), 354–354. https://doi.org/10.3389/fnhum.2013.00354

Volkmar, F., Siegel, M., Woodbury-Smith, M., King, B., McCracken, J., & State, M. (2014). Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *53*(2), 237–257. https://doi.org/10.1016/j.jaac.2013.10.013

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children (3rd ed.) (WISC-III): Manual*. The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children (4th ed.).* The Psychological Corporation.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence–Second Edition (WASI-II).* NCS Pearson.

Wechsler, D. (2014). *WISC-V: Technical and Interpretive Manual*. Pearson.

Weismer, S. E., Lord, C., & Esler, A. (2010). Early language patterns of toddlers on the autism spectrum compared to toddlers with developmental delay. *Journal of Autism and Developmental Disorders*, *40*(10), 1259–1273.

Weiss, L. G., & Saklofske, D. H. (2020). Mediators of IQ test score differences across racial and ethnic groups: The case for environmental and social justice. *Personality and Individual Differences*, *161*, 109962–109962.

Wiggins, L. D., Barger, B., Moody, E., Soke, G., Pandey, J., & Levy, S. (2019). Brief Report: The ADOS Calibrated Severity Score Best Measures Autism Diagnostic Symptom Severity in Pre-School Children. *Journal of Autism and Developmental Disorders*, *49*(7), 2999–3006. https://doi.org/10.1007/s10803-017-3072-x

Woolfenden, S., Sarkozy, V., Ridley, G., & Williams, K. (2012). A systematic review of the diagnostic stability of Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *6*(1), 345–354. https://doi.org/10.1016/j.rasd.2011.06.008