

12-17-2013

Sex Specific Behavioral Profiles in Toddlers At Risk for Autism Spectrum Disorders (ASD)

Natasha Ludwig

Follow this and additional works at: https://scholarworks.gsu.edu/psych_theses

Recommended Citation

Ludwig, Natasha, "Sex Specific Behavioral Profiles in Toddlers At Risk for Autism Spectrum Disorders (ASD)." Thesis, Georgia State University, 2013.

https://scholarworks.gsu.edu/psych_theses/113

This Thesis is brought to you for free and open access by the Department of Psychology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Psychology Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

SEX SPECIFIC BEHAVIORAL PROFILES IN TODDLERS AT RISK FOR AUTISM SPECTRUM DISORDERS
(ASD)

by

Natasha N. Ludwig

Under the Direction of Diana L. Robins

ABSTRACT

The Positive Predictive Value (PPV) of the Modified Checklist for Autism in Toddlers (M-CHAT), a parent report autism screening tool, is higher for males than for females (Ludwig et al., IMFAR 2011). Given the long waitlists and high costs for ASD evaluations, there is a need to reduce the number of false positive females on the M-CHAT. The current study examined the sex specific clinical profiles of toddlers who received an ASD evaluation based on M-CHAT screen positive status in order to explore potential differences that may contribute to the differential PPV of the M-CHAT in boys and girls. The sample included 250 males and 106 females (mean age=25.3 months, $SD=4.6$) who were evaluated based on screen positive status on the M-CHAT. Although children with ASD demonstrated greater ASD symptoms, lower IQ and weaker language and motor skills, minimal sex differences were discovered.

INDEX WORDS: Autism, Sex differences, M-CHAT, Screening

SEX SPECIFIC BEHAVIORAL PROFILES IN TODDLERS AT RISK FOR AUTISM SPECTRUM DISORDERS
(ASD)

by

NATASHA N. LUDWIG

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

Master of Arts

in the College of Arts and Sciences

Georgia State University

2013

Copyright by
Natasha Nerissa Ludwig
2013

SEX SPECIFIC BEHAVIORAL PROFILES IN TODDLERS AT RISK FOR AUTISM SPECTRUM DISORDERS
(ASD)

by

NATASHA N. LUDWIG

Committee Chair: Diana Robins

Committee: Lauren Adamson

Erin Tully

Lisa Wiggins

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

December 2013

DEDICATION

This work is dedicated to Gianni; my little brother, my biggest inspiration.

ACKNOWLEDGEMENTS

I would like to acknowledge all of the families who so generously participated in this research.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
1 INTRODUCTION.....	1
1.1 Sex differences in ASD symptomology	2
<i>1.1.1 Understanding equivocal findings in the sex specific presentation of ASD.....</i>	<i>2</i>
<i>1.1.2 Sex differences specific to the repetitive, restricted and stereotyped patterns of behavior domain in individuals with ASD.</i>	<i>5</i>
<i>1.1.3 Sex differences within the social and communication domains in individuals with ASD.</i>	<i>5</i>
<i>1.1.4 Sex differences in ASD symptoms in toddlers with ASD.....</i>	<i>7</i>
<i>1.1.5 Sex differences in ASD symptomology in non-ASD samples.</i>	<i>7</i>
1.2 Sex differences in cognitive ability/intellectual functioning in individuals with ASD. ..	9
<i>1.2.1 Sex-based IQ differences in preschoolers, school-age children, adolescents and adults with ASD.</i>	<i>9</i>
<i>1.2.2 Sex-based IQ differences in toddlers with ASD.</i>	<i>10</i>
<i>1.2.3 Explanations of discrepant findings between toddlers and older individuals with ASD regarding IQ-based sex differences.</i>	<i>11</i>
1.3 Sex differences in developmental skills: Language and motor domains	11
1.4 Including IQ as a covariate	12
1.5 Addressing the gaps	13
2 HYPOTHESES.....	15

2.1	Hypothesis 1: ASD symptoms.....	15
2.1.1	<i>Hypothesis 1a: Domain symptoms.</i>	<i>16</i>
2.1.2	<i>Hypothesis 1b: Global symptom severity.</i>	<i>16</i>
2.1.3	<i>Hypothesis 1c: Including IQ as a covariate.</i>	<i>16</i>
2.2	Hypothesis 2: Intellectual functioning (IQ).....	17
2.3	Hypothesis 3: Developmental skills.....	17
2.3.1	<i>Hypothesis 3a: Language and motor skills.</i>	<i>17</i>
2.3.2	<i>Hypothesis 3b: Including IQ as a covariate.....</i>	<i>18</i>
3	METHODS	18
3.1	Participants.....	18
3.2	Measures	19
3.2.1	<i>ASD Screening Measures.....</i>	<i>19</i>
3.2.2	<i>Autism diagnostic measures.</i>	<i>22</i>
3.2.3	<i>Measures of cognitive and developmental functioning.</i>	<i>25</i>
3.2.4	<i>Measures not analyzed.</i>	<i>25</i>
3.3	Procedure	27
3.4	Data analyses.....	29
4	RESULTS	30
3.5	Missing data.....	30
3.6	Overall Results	30

3.7	Results for hypothesis 1: ASD symptoms	31
3.7.1	<i>Overall Results for hypothesis 1.</i>	32
3.7.2	<i>Results for Hypothesis 1a: Domain Symptoms.</i>	32
3.7.3	<i>Results for hypothesis 1b: Global symptom severity.</i>	34
3.7.4	<i>Results for hypothesis 1c: Including IQ as a covariate.</i>	35
3.8	Results for hypothesis 2: Intellectual functioning (IQ)	36
3.8.1	<i>Overall results for hypothesis 2.</i>	36
3.8.2	<i>IQ-based prevalence.</i>	36
3.8.3	<i>Mean nonverbal IQ (NVIQ).</i>	37
3.9	Results for hypothesis 3: Developmental skills	38
3.9.1	<i>Overall results for hypothesis 3.</i>	39
3.9.2	<i>Results for hypothesis 3a: Language and motor skills.</i>	39
3.9.3	<i>Results for hypothesis 3b: Including IQ as a covariate.</i>	41
5	DISCUSSION	42
3.10	Discussion of findings: ASD symptoms	42
3.10.1	<i>Methodological differences that may have contributed to inconsistent findings in ASD symptoms across toddler studies</i>	44
3.11	Discussion of findings: IQ	47
3.11.1	<i>Discussion about the lack of sex differences in IQ in toddler samples.</i>	48
3.12	Discussion of findings: Developmental skills	49
3.12.1	<i>Language skills.</i>	49

3.12.2	<i>Motor skills.</i>	49
3.12.3	<i>Interpretation of scores with large standard errors.</i>	50
3.13	Including an IQ as a covariate: Does it change anything?	51
3.14	Conclusion	52
REFERENCES		54
APPENDICES		63
	Appendix A	63
	Appendix B	64
	Appendix C	75

1 INTRODUCTION

Autism Spectrum Disorders (ASDs) are a set of neurodevelopmental disorders characterized by deficits in social interaction and communication, as well as restricted, repetitive, and stereotyped patterns of behavior. There are three disorders that fall into the category of ASD: Autistic Disorder, Asperger's Disorder and Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS; American Psychiatric Association [DSM-IV-TR], 2000). ASDs are more prevalent in males than females with current estimates of the male to female ratio ranging from 2.7/1 to 7.2/1 with an average of 4.6/1 (Centers for Disease Control and Prevention [CDC], 2012). Evidence suggests that this ratio may be dependent upon cognitive functioning, such that comorbid Intellectual Disability (ID) is more prevalent in females (CDC, 2012; Ehlers, Gillberg, & Wing 1993; Honda, Shimizu, Imai, & Nitto, 2005; Lord, Schopler, & Rivicki, 1982; Volkmar, Szatmari, & Sparrow, 1993; Wing, 1981; Yeargin-Allsopp et al., 2003).

Given that the formulation of our current diagnostic criteria (DSM-IV) is primarily based on the male presentation, and that the higher male prevalence of ASD skews the sex distribution of research samples in favor of males, little is known about the female phenotype of ASD. Furthermore, there are few studies specifically examining sex differences in the clinical presentation of ASD, and even fewer focused on differences in toddlers. Enhancing our knowledge of potential behavioral differences within this age group is important given the implications for early detection and diagnosis.

The present study addressed this gap by examining potential sex differences in behavioral profiles of toddlers considered at risk for an ASD based on a widely used ASD screening tool, the Modified Checklist for Autism in Toddlers (M-CHAT; Robins, Fein, & Barton, 1999a). Several aspects of the clinical phenotype were analyzed, including ASD symptoms, nonverbal intellectual functioning, and general developmental skills, including language and motor functioning. Analyses explored behavioral sex differences in toddlers with and without an ASD, in order to elucidate whether possible sex differences are

specific to an ASD diagnosis or to all children initially considered at risk based on the M-CHAT. Studying potential sex differences within this population is integral to improving the clinical utility of the M-CHAT, which is more than three times better at predicting an ASD diagnosis in males versus females (Ludwig, Robins & Fein, IMFAR, 2011).

1.1 Sex differences in ASD symptomology

ASD is a behaviorally defined disorder characterized by impairment within the areas of social interaction, communication, and repetitive, restricted, and stereotyped patterns of behavior; however the specific set of diagnostic symptoms endorsed, and the severity of these symptoms (the ASD symptom profile), differs based on an individual's unique set of strengths and weaknesses. Some studies have demonstrated differences in the manifestation of ASD symptoms based on sex, but findings are equivocal. Due to these inconsistencies, it is important to review not only findings within this body of research, but also to note the methodological differences across studies, such as symptom assessment, sample characteristics, and covariates as these factors may contribute to disparate findings.

1.1.1 Understanding equivocal findings in the sex specific presentation of ASD

The literature examining sex differences in ASD is characterized by inconsistent findings across studies within all areas of functioning, which limits the ability to draw general conclusions that have theoretical and clinical utility. Therefore, it is important to explore factors that may be contributing to these inconsistencies, primarily based in methodological approach. Thinking critically about how methodological techniques strengthen or limit results is important not only for the interpretation of previously reported data, but also in developing a study that will maximize the utility of findings. Table 1 in Appendix B includes some of the major methodological characteristics of the comprehensive studies of behavioral sex differences in ASD cited in this proposal.

The first methodological variables to consider are inclusion and exclusion criteria. Given chang-

ing diagnostic criteria since autism was added to the DSM-III in 1980, as well as the improvement of diagnostic tools, diagnostic inclusion criteria have changed over the years making it difficult to synthesize findings from studies over a broad time period. Some have included individuals with a clinical diagnosis of ASD based on the Diagnostic and Statistical Manual (DSM) criteria only (Hartley & Sikora, 2009; Lai et al., 2011; Rivet & Matson, 2011; Zwaigenbaum et al., 2012), others required an “ASD” classification based on autism specific diagnostic measures such as the CARS, ADOS and/or the ADI (Lord et al., 1982), and some required both (Carter et al., 2007). Furthermore, the inclusion of the various disorders on the ASD spectrum (i.e., Autistic Disorder, Asperger’s Disorder, Pervasive Developmental Disorder-Not Otherwise specified (PPD-NOS)), and individuals with comorbid MR varied by study. Some have excluded individuals with comorbid genetic disorders, and other health complications (Carter et al., 2007), which are common in ASD; however, excluding these children may reduce generalizability of findings to the broad range of children affected by ASD. Variability in age range is another source of inconsistency as many of the studies within this literature have utilized a broad age range, including both very young children and adults (Holtman et al., 2007; Pilowsky et al., 1988). Given that the behavioral sequelae of ASD may change over the course of development, combined with the increased measurement error associated with the use of a range of tests to assess similar constructs across ages, generalizability of findings to specific age ranges is compromised.

It is also important to consider the different methods used to measure behavior; ASD symptoms, intellectual ability, and general development across studies. For example, some measures quantify current behavior, while others utilize retrospective report of childhood behavior. In addition, some researchers have decided to use both parent report of behavior and direct observation methods, but others have restricted measurement to parent report of behavior (McLennan et al., 1993; Park et al., 2012, Rivet & Matson, 2011, Sipes et al., 2011). When using parent report based measures of outcome it is important to consider that previous knowledge of a child’s diagnosis may influence the way in which a

parent interprets their child's behavior. For example, a parent who is unaware of a diagnosis may be less likely to consider a certain behavior as atypical, whereas a parent who is familiar with ASD and has been told by a professional that their child has an ASD, they may be more sensitive to atypical behaviors consistent with their child's diagnosis. This becomes especially important when a study relies solely on parent report measures and few studies make note of parent knowledge of diagnosis at the time of measurement.

One of the most important aspects of the potential impact of methodological approach on findings within this body of research is the decision to partial out variability accounted for by IQ. Given that IQ may be associated with clinical symptomology as well as general developmental functioning (i.e., motor and language skills) and adaptive behavior (Carter et al., 2007; Lord et al., 1982; Zwaigenbaum, 2012), many researchers have decided to control for cognitive functioning in order to understand the differences in behavior that can be attributed to the unique effect of sex. Some studies have accounted for differences in cognitive functioning by recruiting an IQ-matched sample (Holtman et al., 2007; Lai et al., 2011; Pilowsky et al., 1998), whereas others including Carter and colleagues (2007), and Hartley and Sikora (2009), have used an estimate of IQ as a covariate in statistical analyses. Despite this, there is debate in the literature as to whether it is appropriate to control for IQ in studies exploring neurodevelopment (See section entitled "Including IQ as a covariate" for a more detailed review).

There are many factors that may contribute to the equivocal findings that characterize this body of research. However, upon close examination of these factors and outcomes, minimal themes emerge, which makes it difficult to synthesize findings and gain clinically useful information. Subsequent subsections will describe mixed findings pertaining to sex differences in ASD symptom domains in individuals across the lifespan, and focused subsections specific to toddler and non-ASD samples, as these populations are particularly important to the present study.

1.1.2 Sex differences specific to the repetitive, restricted and stereotyped patterns of behavior domain in individuals with ASD.

Within the repetitive, restricted, and stereotyped patterns of behavior domain, studies conducted in toddlers (Hartley & Sikora, 2009), preschoolers, school-aged children, and young adults (Lord et al., 1982; Mandy et al., 2012; Park et al., 2012; Szatmari, 2010) support that males demonstrate more symptoms than females based on both direct observation and parent report measures. In Hartley and Sikora's (2009) sample of toddlers, males showed more repetitive, restricted and stereotyped behaviors than girls during a standardized play-based assessment for autism symptoms, the Autism Diagnostic Observation Schedule (ADOS; Lord et al., 2000). Lord and colleagues (1982) found that males exhibited less appropriate/more stereotypical play and more severe unusual visual interests compared to females, and Sipes and colleagues (2011) demonstrated that females with an average Developmental Quotient (DQ; an estimate of IQ in toddlers) on a measure of developmental skills, the Battelle Developmental Inventory, Second Edition (Newborg, 2005), endorsed fewer items on an ASD symptom checklist (Baby and Infant Screen for Children with Autism Traits; BISCUIT; Matson et al. 2009), related to restrictive and repetitive behaviors compared to males; however, no sex differences were found in children with a below average DQ.

1.1.3 Sex differences within the social and communication domains in individuals with ASD.

Although several studies have demonstrated that boys with ASD tend to exhibit more repetitive, restricted and stereotyped patterns of behavior than girls, findings of differences within other symptom domains are less consistent. For example, Lai and colleagues (2011) found that adults with ASD did not differ on the number of total childhood symptoms on core ASD domains based on retrospective parent report on the ADI-R (Lord, Rutter & Le Couteur, 1994). However, females did endorse more lifetime abnormal sensory interests on the ADI-R, and more ASD symptoms on a self-report measure of ASD traits, the Autism Quotient (AQ; Baron-Cohen et al., 2001). In contrast, when direct observation of current be-

havior was examined, males demonstrated significantly more ASD symptoms within the social and communication domains on the ADOS. Consistently, Park and colleagues (2012) found that males demonstrated significantly more symptoms within the social and communication domains based on parent report (ADI-R) in a sample of children with ASD. Inconsistent with the lack of differences in childhood symptoms observed by Lai and colleague's (2011), McLennan, Lord & Schopler (1993) found sex differences in childhood ASD symptoms in high functioning children and adults with ASD. Males in McLennan and colleagues' sample demonstrated more social and communication symptoms in early childhood, than girls as reported on the ADI-R. With regard to current symptoms, McLennan and colleagues found that parents of females reported more impairment in current friendships than males. Another study of high functioning children and adolescents with ASD (Holtman et. al., 2007), found no differences in total domain scores on the ADI-R or the ADOS.

Whereas some groups report significant sex differences in ASD symptom presentation, there have also been a number of studies that have reported no differences. In Rivet and Matson's (2011) sample of toddlers with ASD, no sex differences were found in the number of symptoms endorsed on the BISCUIT. When the subset of toddlers with an estimated IQ above 70 were analyzed, findings were consistent. Similarly, Volkmar and colleagues (1993) found no significant differences between boys and girls on a checklist of autism symptoms, the Autism Behavior Checklist (ABC; Krug, Arick, & Almond, 1980), or in the total number of symptoms endorsed on the ICD-10 diagnostic criteria for autism (World Health Organization, 1990), both when IQ was controlled, and when it was not. In a sample of children with autism ranging from 20 months to 34 years old, with boys and girls matched on chronological and mental age, no sex differences were observed in parent reported ASD symptoms as measured by ASD symptom domain scores on the ADI-R or on clinician ratings of ASD symptoms on the CARS (Pilowsky et al., 1998). Several studies of high functioning children and adults with ASD, matched for IQ, demonstrat-

ed no sex differences in ASD symptom domain scores on the ADI-R and the ADOS (Holtman et al., 2007; Lai et al., 2010).

1.1.4 Sex differences in ASD symptoms in toddlers with ASD.

When considering only the toddler literature exploring sex differences in children with ASD, inconsistencies persist. Carter and colleagues (2007) found no sex differences in ASD symptomatology based on the ADOS and the ADI-R; however, Hartley and Sikora (2009) found that boys demonstrated greater restricted, repetitive and stereotyped behaviors compared to boys and girls demonstrated greater communication deficits than boys on the ADOS. In contrast to both of these studies, Zwaigenbaum and colleagues (2012) study of sex differences in high-risk siblings with ASD revealed that boys demonstrated greater social and communication deficits and greater overall symptom severity as measures by the ADOS2.

Overall, there are inconsistencies in the research examining the sex specific behavioral presentation of ASD symptoms, making it difficult to draw conclusions about whether differences exist among the kinds of symptoms endorsed, and the severity of these symptoms in males and females.

1.1.5 Sex differences in ASD symptomatology in non-ASD samples.

Thus far, literature examining sex differences strictly within ASD samples has been reviewed. However, the proposed study will broaden the study of sex differences beyond children with a confirmed diagnosis, and will include children who are initially at risk for an ASD based on the M-CHAT, regardless of subsequent diagnosis. There are no known studies to date that have examined the sex specific behavioral profiles of all young children who are at risk for an ASD based on early screening practices. However, it is important to explore whether these ASD sex differences are specific to ASD, or can be generalized to all children who are at risk in order to generate knowledge that can be used to improve early screening and diagnostic practices.

Child, adolescent and adult screening tools, including the Social Responsiveness Scale (SRS; Constantino & Todd, 2003), Autism Spectrum Quotient (AQ; Auyeung et al., 2008; Baron-Cohen et al., 2001; Baron Cohen et al., 2006), Systemizing Quotient (SQ; Auyeung et al., 2009) and the Childhood Autism Spectrum Test (Williams et al., 2008), yield significantly higher scores in boys compared to girls. Similarly, toddler screening tools including the Qualitative- Checklist for Autism in Toddlers (Q-CHAT; Allison et al., 2008) and the M-CHAT (Ludwig et al., IMFAR, 2012) also demonstrate this pattern. Furthermore, the proportion of males who screen positive on the M-CHAT is higher than the proportion of females who screen positive (Ludwig et al., IMFAR, 2011). This suggests that in the general population, males endorse more autistic traits than females.

Some have also explored the behavioral presentation of children without ASD using ASD specific diagnostic measures. In a study examining unaffected siblings of children with ASD aged 4-15, Park and colleagues (2012) demonstrated that males endorsed higher scores on both the communication and social interaction domains based on parent report using the ADI-R. Similarly, Zwaigenbaum and colleagues (2012) found that males demonstrated increased ASD symptom endorsement and symptom severity compared to females on both the ADOS and ADI, in high-risk siblings without ASD, and in a typical controls. In contrast to these non-developmentally delayed populations, studies in children with developmental disabilities other than ASD, have not found sex differences. For example, Rivet and Matson (2011) found no differences in the number of autism symptoms endorsed on an autism symptom checklist, the BISCUIT, in a sample with developmental disabilities.

Although sex differences in ASD diagnostic symptoms are to examine, there are other areas of functioning that may affect our ability to identify ASD in toddlers, such as cognitive ability and developmental skills. Therefore, it is important to study potential sex differences in these areas of functioning in order to better understand how these features may influence screening and diagnostic practices as well as treatment efforts in boys and girls on the autism spectrum.

1.2 Sex differences in cognitive ability/intellectual functioning in individuals with ASD.

Findings from recent population surveillance studies suggested a high prevalence of comorbid cognitive impairment in ASD, with estimates of 38-45% of children presenting with an Intellectual Quotient (IQ) of 70 or below (CDC, 2012; CDC, 2007; Giarelli et al., 2010). Evidence also has suggested that cognitive ability is correlated to ASD symptom severity in toddlers (Carter et al., 2007; Zwaigenbaum, 2012), and that early measures of cognitive ability are valuable predictor of functional outcome later in development (Sigman & McGovern, 2005). Taken together, these data support that cognitive functioning is an important aspect of the clinical phenotype in individuals with ASD.

1.2.1 Sex-based IQ differences in preschoolers, school-age children, adolescents and adults with ASD.

Interestingly, this domain seems to be particularly vulnerable to sex differences. One of the most consistent findings in the literature pertaining to sex differences in ASD is that girls are more likely than boys to present with comorbid Intellectual Disability (ID); this pattern has been observed in both clinical (Lord et al., 1982; Volkmar et al., 1993; Pilowsky et al., 1998), and population based samples (CDC, 2012, CDC, 2007; Giarelli, 2010; Honda et al., 2005; Yeargin-Allsopp et al., 2003). A recent population surveillance study (CDC, 2012) suggested that 46% of school-age girls with ASD presented with comorbid MR, whereas only 37% of boys demonstrated this profile. An examination of findings from studies of school-age children and older revealed that females with ASD tended to score consistently lower on measures of cognitive ability compared to males (Lord et al., 1982; Volkmar et al., 1993). Furthermore, empirical evidence suggests that IQ moderates the differential prevalence of ASD in boys and girls, with the highest ratio of boys to girls in samples of children without cognitive impairment (Bryson, Clark, & Smith, 1988; Fombonne 2003; Volkmar et al., 1993; Wing, 1981). Fombonne (2003) found that the ratio of males to females with IQ in the Average range is 5.5:1, but drops to 1.95:1 in children with cognitive impairment, and Wing (1981) demonstrated a linear relationship between IQ and the sex dif-

ferences in prevalence across the ASD spectrum, with a higher proportion of boys as IQ increased. A recent article published by Dworzynski and colleagues (2012), suggested that IQ may impact the likelihood that a child is diagnosed with ASD. In their population sample, boys were more likely to meet diagnostic criteria for ASD when cognitive problems were not present.

1.2.2 *Sex-based IQ differences in toddlers with ASD.*

With the recent development of early ASD screening and diagnostic tools, it has become possible to explore potential sex differences in cognitive ability in very young children with ASD. It is of note that given the variability of language development in toddlers, a nonverbal measure of cognitive ability is typically used as an estimate of IQ in this population. The Visual Reception (VR) scale, on the Mullen Scales of Early Learning (MSEL; Mullen, 1995; a measure of early cognitive and developmental functioning), which taps nonverbal problem solving skills, is commonly used as an estimate of IQ in toddler studies. In contrast to the consistent finding that older females with ASD tend to be lower functioning than older males with ASD, findings from toddler studies have been far more variable. Both Hartley and Sikora (2009) and Zwaigenbaum and colleagues (2012) did not find sex differences in VR scores on the MSEL in their toddler samples, and Carter and colleagues (2007) suggested that girls performed better on the MSEL VR scale compared to boys in their sample of 90 toddlers with ASD, but only when language level was controlled. Interestingly, Zwaigenbaum and colleagues' (2012) study, which extended to high-risk siblings with and without ASD, revealed no sex differences in the non-ASD sample of toddlers as well. Overall, it appears that studies of sex differences in IQ in toddler samples do not replicate findings in older populations, which have suggested that females tend to be lower functioning.

1.2.3 Explanations of discrepant findings between toddlers and older individuals with ASD regarding IQ-based sex differences.

Clearly, findings and the consistency of findings pertaining to cognitive ability in ASD vary based on age. On one hand, considering that lower cognitive functioning is associated with earlier identification (Shattuck et al., 2009), it is possible there may be a higher proportion of lower functioning boys in toddler samples than in the general ASD population. On another hand, these differences also could be due to methodological inconsistencies in the way cognitive ability is measured across the lifespan. Studies in older children utilize traditional measures of intellectual functioning in order to classify high and low functioning individuals. These traditional measures incorporate nonverbal, verbal and motor functioning into the overall IQ metric. However, given that toddler studies commonly use nonverbal measures of IQ to distinguish high from low levels of functioning, perhaps language and/or motor skills drive the IQ sex differences observed in older samples. Despite these postulations, the equivocal nature of these findings combined with the potential influence of cognitive ability on the timing of identification, diagnosis and treatment outcomes, highlight the importance of further research examining cognitive related sex differences in toddlers.

1.3 Sex differences in developmental skills: Language and motor domains

Potential sex differences in other areas of general development, including language and motor functioning, also warrant further exploration given that individuals with ASD often demonstrate impairments in both areas (Bhat et al., 2012). In Lord and colleagues' (1982) investigation of general developmental differences in children with ASD, boys performed better in almost all areas including, motor skills and receptive language; however, when IQ was controlled, sex differences in these areas of development were no longer significant. This is consistent with recent findings from Hartley and Sikora's (2009) investigation of functioning in 199 toddlers with ASD; boys and girls evidenced similar developmental profiles based on the MSEL when cognitive ability was held constant. Despite this, other research sug-

gests that sex differences in developmental domains persist in toddlers when IQ is controlled. For example, in Carter and colleagues' (2007) sample of toddlers with ASD, sex was a significant predictor of language and motor skills, and boys performed significantly better than girls. In contrast, girl toddlers in Zwaigenbaum and colleagues' (2012) high-risk siblings with and without ASD, and typical controls, demonstrated stronger motor skills, despite ASD status.

In sum, it appears that in some studies, IQ accounts for differences in developmental skills, whereas in other studies, these differences exist above and beyond differences in cognitive abilities. Given the importance of language and motor milestones in screening, diagnosis and treatment of individuals with ASD, more research is needed to explore potential sex differences within these realms in order to improve our understanding of the sex specific presentation of ASD.

1.4 Including IQ as a covariate

Studies conducted within this body of research vary with regard to the control of IQ. However, given recent compelling arguments against controlling for IQ in studies of neurodevelopmental disorders, it is important to consider the justification for doing so in the context of the present study.

Volkmar and colleagues (1993) proposed that deciding whether it is appropriate to control for cognitive ability/intellectual functioning requires theoretical justification. The authors suggest that control for IQ is appropriate if IQ is conceptualized as the cause of behavioral sex differences in ASD, but would be inappropriate if it is an associated feature, or a result of sex differences. Controlling for IQ in the latter scenario might lead to, "the control of factors that are not confounding variables" (Volkmar et al., 1993, p. 581).

Research has supported that cognitive functioning is related to both general development (i.e., motor and language functioning), and ASD symptoms, such that lower cognitive functioning predicts more impairment in these areas. For example, Carter and colleagues (2007) demonstrated that nonverbal cognitive ability (MSEL VR score) significantly predicted overall language and motor abilities in tod-

dlers with ASD. With regard to ASD symptomology, findings from several studies suggest that cognitive ability is a significant predictor of core ASD symptoms (Carter et al., 2007; Hus et al., 2007; Bolte et al., 2011). Although these findings support a relationship between IQ and other behavioral outcomes, the question still remains whether IQ *causes* variation within these areas, or whether IQ is an outcome of the disorder itself.

Dennis and colleagues (2009) put forth strong arguments against the conceptualization of IQ as causal, and use of IQ as a covariate in the study of any neurodevelopmental condition. The authors have suggested that IQ does not predate, but “postdates the [neurodevelopmental] condition, charts the history of the condition, is always confounded with and/or by the condition, and can never be separated from the effects of the condition” (p. 2). The authors provide theoretical evidence against the idea that intelligence represents an innate construct that predicts aptitude and potential, but rather is a measure of achievement and performance that is shaped by the condition. According to Dennis and colleagues, this common misconception has led to the idea that IQ is a contributing factor to disorder, and should be controlled when other causal variable are of interest.

Despite these arguments, many researchers continue to engage in age matching and statistical control for cognitive functioning in studies of sex differences in ASD. Therefore, further research is needed to investigate how the statistical control of IQ in the study of ASD sex differences may influence findings, and the utility of these findings in clinical practice. For example, a question to consider is whether partialling out IQ will yield findings that can be used to improve screening tools such as the M-CHAT or diagnostic tools such as the ADOS, that do not account for cognitive ability in their scoring algorithms.

1.5 Addressing the gaps

Although there is great clinical value in understanding sex differences in ASD, little is known about the female phenotype, and findings from research examining sex differences are equivocal. Fur-

thermore, few studies have focused on toddlers, and there are no known studies that have comprehensively examined the sex specific behavioral presentation of all children who are at risk for ASD based on toddler screening in primary care settings to explore whether sex differences are specific to ASD or can be generalized to all children initially considered at risk. Research exploring potential differences in this group is important in order to gain an understanding about how these differences may differentially impact the effectiveness of early screening and diagnosis for boys and girls. Therefore, the present study expanded upon the current ASD sex differences literature, beyond children with a confirmed diagnosis, and it examined potential differences among a broader group of children referred for an ASD evaluation based on the M-CHAT.

Previous studies in toddlers are limited by homogenous samples that excluded children with genetic disorders and other health impairments (Carter et al., 2007). The present study attempted to improve upon this limitation by exploring sex differences in a representative sample of toddlers recruited through ASD screening procedures conducted in low-risk (population) samples, who were not excluded based on comorbid genetic or medical diagnoses. In addition, many of the previous studies examining sex differences in toddlers have only included children referred to early intervention programs due to previous parent or clinician concern (i.e., high-risk samples; Carter et al., 2007; Hartley & Sikora, 2009; Rivet & Matson, 2011). Although findings from these studies are important, studies of sex differences have yet to be conducted in children considered at risk based on level-one screening tools such as the M-CHAT. This is important to study given that level-one screeners are used in the general population of toddlers (i.e., low-risk samples), where concerns may not be endorsed. Therefore, children in a level-one screening sample may constitute children who are phenotypically different from these high-risk children that have previously been studied (Zwaigenbaum et al., 2009). Given this gap, the present study included only low-risk toddlers who screen positive on the M-CHAT, as this will contribute novel knowledge

about the manifestation of sex differences in this population that may be clinically useful in enhancing low-risk screening techniques.

Given the inconsistent control of IQ in studies of sex differences, and the theoretical debate over the appropriateness of this common practice in the study of neurodevelopmental disorders, more research is needed examining the effect of controlling for cognitive ability/intellectual functioning in studies exploring sex specific clinical profiles in ASD. In order to address this gap, current analyses exploring sex differences in the clinical phenotype were conducted both with and without the inclusion of a covariate, and qualitative observations were made regarding substantial differences in results across statistical method.

The primary aim of this study was to explore potential sex differences in the behavioral characteristics of toddlers who screen positive on a low-risk screening tool, the M-CHAT. Specifically, sex differences in ASD symptomology, intellectual functioning, and developmental skills were examined in toddlers with ASD, and in toddlers who were initially considered at risk for ASD based on the M-CHAT, but who were not diagnosed with ASD upon further evaluation. Findings will contribute knowledge about whether sex differences are specific to ASD or whether they can be generalized to all children considered at risk based on this level-one screening tool.

2 HYPOTHESES

2.1 Hypothesis 1: ASD symptoms

In order to explore potential sex differences in ASD symptoms, symptom profiles in males and females were compared.

2.1.1 Hypothesis 1a: Domain symptoms.

Consistent with recent findings (Carter et al., 2007; Hartley & Sikora, 2009), it was predicted that girls with ASD would demonstrate more symptomology within the social and communication domains than boys, whereas boys with ASD would demonstrate more symptomology within the repetitive and stereotyped patterns of behavior domain than girls. Analyses also examined whether this effect was specific to ASD or extended to all those initially considered at risk based on the M-CHAT. Based on evidence that boys in the general population tend to demonstrate more ASD symptoms than girls (Auyeung et al., 2008; Auyeung et al., 2009; Baron-Cohen et al., 2001; Baron Cohen et al., 2006; Constantino & Todd, 2003; Williams et al., 2008; Zwaigenbaum et al., 2012), it was predicted that sex differences would differ based on ASD status, such that non-ASD boys would demonstrate more ASD symptoms across domains, compared to girls.

2.1.2 Hypothesis 1b: Global symptom severity.

Exploratory analyses also examined whether there were sex differences in global symptom severity in toddlers with ASD, and whether this effect was specific to ASD or extend to all those initially considered at risk by the M-CHAT. It was predicted that boys, despite ASD status, would demonstrate more severe symptoms based on previous research exploring global symptom severity in toddlers with and without ASD (Zwaigenbaum, 2012).

2.1.3 Hypothesis 1c: Including IQ as a covariate.

Given recent debate regarding the control of IQ in examining differences between groups within the ASD phenotype (Dennis et al., 2009; Volkmar et al., 1993), all analyses within Hypotheses 1 were conducted with and without controlling for IQ in order to qualitatively explore how controlling for cognitive functioning may have changed these findings.

2.2 Hypothesis 2: Intellectual functioning (IQ)

In order to explore potential sex differences in cognitive functioning in this sample of toddlers, intellectual functioning (IQ) was explored. Although the literature on older individuals with ASD is clear that the ratio of males:females varies by cognitive ability, with greater differences in high functioning individuals, and smaller differences in low functioning individuals (Bryson, Clark, & Smith, 1988; Fombonne, 2003; Volkmar et al., 1993; Wing, 1981), this finding may not extend to toddlers with ASD (Carter, 2007; Hartley & Sikora, 2009; Zwaigenbaum et al, 2012). Therefore, it was predicted that the ratio of males to females with ASD would not differ based on level of cognitive functioning, and that there would be no sex differences in IQ. Analyses also explored whether these potential sex differences are specific to ASD, or extend to all children considered at risk. Based on recent evidence that cognitive ability does not differ between high-risk siblings with and without ASD (Zwaigenbaum et al., 2012), it was predicted that no sex-based differences would emerge in either diagnostic group.

2.3 Hypothesis 3: Developmental skills

In order to explore other areas of the ASD clinical profile that may be differentially affected in boys and girls that may have implications for early screening and diagnosis, sex differences in general developmental skills including language and motor functioning were examined.

2.3.1 Hypothesis 3a: Language and motor skills.

Recent findings from toddler studies are mixed. Some have suggested that girls with ASD present with lower language and motor skills than boys (Carter et al., 2007; Hartley & Sikora, 2009), whereas others have demonstrated the opposite pattern with regard to motor skills, specifically (Zwaigenbaum et al., 2012). Given that there is more research in toddlers supporting that girls present with weaker language and motor skills compared to boys, it was hypothesized that girls with ASD would demonstrate weaker levels of functioning within these areas compared to boys with ASD. Analyses also

examined whether these developmental sex differences are specific to the ASD phenotype, or can be generalized to all children at risk. Given recent findings that have suggested non-ASD high-risk sibling females demonstrate stronger developmental skills than their male counterparts (Zwaigenbaum, 2012), it was hypothesized that girls within the non-ASD group would demonstrate stronger language and motor skills than boys.

2.3.2 Hypothesis 3b: Including IQ as a covariate.

Additionally, all analyses were conducted with and without controlling for cognitive ability in order to explore how controlling for cognitive functioning (i.e., estimated IQ) may change these outcomes.

3 METHODS

The current study utilized data collected from validation studies of the Modified Checklist for Autism in Toddlers (M-CHAT), and a revised version of the original M-CHAT, the Modified Checklist for Autism in Toddlers-Revised (M-CHAT-R). These data are from a larger NICHD-funded study, and include the low-risk samples collected in metropolitan Atlanta and Connecticut. Participants consisted of toddlers screened at 18- and 24-month well-child visits at their pediatrician's office.

3.1 Participants

This study included children who (a) screened positive on the M-CHAT(-R) and Follow-Up Interview (FUI), (b) received a comprehensive diagnostic evaluation at either Georgia State University or the University of Connecticut as part of the M-CHAT(-R) studies, (c) were English speaking, and (d) for whom parental consent for participation was obtained. Although all children in the sample were administered the FUI, it is of note that some children's FUI data are physically missing. However, these children were included in the analyses under the assumption that all children who are evaluated based on an M-

CHAT(-R) screen positive status, must have also screened positive on the FUI in order to qualify for the evaluation.

This study excluded children (a) with missing sex, diagnostic status, age, or MSEL VR age equivalent score data as these were considered integral variables for the analyses, (b) children who were administered the ADOS Module 2 rather than Module 1 given the low frequency of children who were administered Module 2 in the screening sample and the differences that exist between the ADOS Modules, and (c) those with severely impaired sensory or motor functioning as this could influence the validity of the diagnostic measures used. There were no other exclusions for comorbid conditions.

The final sample included 356 toddlers between the ages of 16 and 43 months (mean age 25.3 months, $SD= 4.55$). The sample included 250 males ($n=139$ with ASD), and 106 females ($n=37$ with ASD). A between subjects ANOVA conducted with sex and ASD status as fixed factors, revealed no significant differences in age across groups ($ps>.05$). Children with ASD ($n=176$) were diagnosed with either Autistic Disorder ($n=74$) or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS; $n=102$). Children who were not diagnosed with ASD ($n=180$) were diagnosed with a Developmental Language Delay ($n=42$), Global Developmental Delay ($n=85$), another diagnosis ($n=5$), or were not given a diagnosis ($n=48$). A chi-square analysis conducted on the entire sample revealed no statistical differentiation from the predicted model in the number of males and females who were diagnosed with Autistic Disorder versus males and females diagnosed with PDD-NOS ($\chi^2 =.92, p=.338$). Please refer to Table 2 in Appendix B for extended demographic data.

3.2 Measures

3.2.1 ASD Screening Measures.

The Modified Checklist for Autism in Toddlers (M-CHAT) and Follow-Up Interview (FUI): The M-CHAT (Robins, Fein, & Barton, 1999 a & b; see Appendix) is an early screening tool used to detect chil-

dren who may be at risk for an ASD in low-risk samples. The M-CHAT is a 23 item, *yes/no* questionnaire filled out by parents, often during pediatric well-visits at 18 and 24 months. Items on the M-CHAT assess both typical childhood behavior (e.g., *looking toward an object a caregiver is looking at*) as well as behaviors that are commonly observed in toddlers with ASD (e.g., *unusual finger movements near the face*). Children who screen positive on the M-CHAT are considered at risk for ASD.

Children in this study were classified as a screen positive on the M-CHAT if their parents endorsed at least three items total or at least two out of six critical items (as determined by a Discriminate Function Analysis). These six best discriminators of a diagnosis are, *taking interest in other children, using a finger point to express interest, bringing objects to a parent to show, imitating caregivers, responding to a name call, looking toward object pointed out by parent*. Parents of children who screened positive were contacted for the FUI (see Appendix for an FUI example item) to confirm appropriate item interpretation. If a child continued to screen positive on the FUI, using the same thresholds of any three items or 2/6 critical items, they were invited for a full diagnostic evaluation.

Published Psychometrics: Internal consistency of the M-CHAT was reported as adequate for the entire checklist (Cronbach's alpha of .85), and for the subset of critical items (Cronbach's alpha of .83-.84; Kleinman et al., 2008; Robins, Fein, Barton, & Green, 2001). Positive predictive value (PPV; proportion of children who screen positive who have an ASD) of the M-CHAT is .11 in low-risk samples (Kleinman et al., 2008), but when the FUI is considered as part of the M-CHAT screening process, PPV increases to .57-.74 (Kleinman et al., 2008; Robins, 2008; Robins et al., 2001). Although psychometric properties such as absolute sensitivity (probability that those who have an ASD diagnosis will screen positive on the M-CHAT), specificity (probability that those who do not have an ASD diagnosis will pass the M-CHAT), and negative predictive value (NPV; proportion of children passing the M-CHAT who do not have an ASD diagnosis) cannot be calculated without confirming non-ASD status in screen negative cases, these values were estimated using a Discriminate Function Analysis. According to Robins and colleagues (2001),

the sensitivity of the M-CHAT was estimated to be .97, specificity, .95, and NPV, .99. When the FUI is considered as part of the M-CHAT screening process, sensitivity and NPV remained .97 and .99 respectively; however, specificity increased to .99 (Kleinman et al., 2008; Robins, 2008; Robins et al., 2001). The M-CHAT and FUI appropriately distinguish developmental delay from typical development, as 85-89% of children who screen positive on the M-CHAT are diagnosed with a DSM-IV disorder (e.g., language delay, global developmental delay) upon further evaluation (Chlebowski et al., 2013; Robins, 2008).

The Modified Checklist for Autism in Toddlers Revised (M-CHAT-R) and Follow-Up Interview

(M-CHAT-R FUI): The M-CHAT-R and M-CHAT-R FUI (Robins, Fein, & Barton, 2009 a & b) were developed to simplify wording of items, and improve the positive predictive value of the screening tool. The M-CHAT-R eliminated three items from the original M-CHAT that were least effective at predicting a diagnosis of ASD, improved wording, added examples to enhance item interpretation, reorganized item order to prevent parents from falling into a *yes* response pattern, and modified the critical score to include seven “best” items instead of the original critical six. These seven critical items contain five of six original critical items from the M-CHAT.

In May 2011, the Connecticut site changed FUI cutoff criteria to any two items, rather than any three or 2/7 “Best” items. In July 2011, the Atlanta site changed M-CHAT-R cutoff criteria to 14 “Best” items and changed the FUI cutoff criteria to any two items rather than any three or 2/6 “Best” items. These changes were made given preliminary data that suggested the sensitivity of the “best” items was low compared to the total score and that the score of 2 on the FUI indicated significant risk for ASD.

Published Psychometrics: The psychometric properties of the M-CHAT-R are limited as the authors are currently in the process of conducting a validation study. However, in a sample of 7,006 toddlers screened between the ages of 15 and 30 months, PPV for the M-CHAT-R was .11 and when the M-CHAT-R FUI was considered part of the screening process, PPV increased to .59 (Robins & Fein, IMFAR,

2011). Sensitivity, specificity and NVP of the M-CHAT-R were estimated to be .91, .95, and .99 respectively. When the FUI was considered, sensitivity decreased to .72, specificity increased to .99., and NVP remained at .99 (Robins & Fein, IMFAR, 2011).

3.2.2 Autism diagnostic measures.

Autism Diagnostic Observation Schedule (ADOS): The ADOS (Lord et al., 2000) is an observational measure of ASD behavior. The ADOS is administered as a play session structured to elicit opportunities for the individual to respond to social presses. There are four language- and developmental-level dependent modules, and each takes between 30 to 60 minutes to administer. Toddlers in the M-CHAT(-R) validation studies received either Module 1 (preverbal or single words). Behaviors within four domains are assessed (i.e., social, communication, stereotyped, repetitive behaviors, and play skills) and items are scored on a 2-4-point scale (scale depending on the item), with 0 indicating *no abnormality* and 3 indicating *moderate to severe abnormality*. Domain algorithm scores are computed based on an algorithm (only items shown to best predict a diagnosis are included, and scores of 3 are converted to 2), and a Total algorithm score is computed using Domain algorithm scores from the Social and Communication domains only. The ADOS classifies individuals into one of three categories; *autism*, *autism spectrum* or *non-spectrum*. In order to be classified with *autism* or *autism spectrum* on the ADOS, Domain algorithm scores on the social and communication domains, as well as the Total algorithm score (Communication Domain algorithm score + Social Domain algorithm score), must meet diagnostic cutoffs.

Recently, the ADOS2 was developed in order to reflect upcoming changes to the DSM criteria for ASD, and to clarify administration guidelines and codes in order to improve the accuracy and effectiveness of the tool (Lord et al., 2012). Although the ADOS2 protocols were not used in the present study, the administration and coding of the ADOS2 is “functionally identical” to the original ADOS, and scores from the original ADOS were used to compute new domain algorithms based on the ADOS2. ADOS2 algorithms have changed based on a recent validation study that improves sensitivity and specificity of the

tool. The ADOS2 has only two Domain algorithms including the Social Affect (SA) Domain, which collapses items across both the social and communication symptom domains, and the Restricted and Repetitive Behavior (RRB) Domain.

Although the original ADOS was developed solely as a diagnostic tool, Domain algorithm scores are commonly used as a stand-in measure of symptom severity. Because language and age affect scores, the ability to measure symptom severity based on scores is limited. The ADOS2 developed a standardized comparison score (Severity Score) that can be used as a measure of severity (Gotham, Pickles & Lord, 2009). The ADOS2 provides conversions to compute the comparison score based on the Total score for children 24 months of age and older; however given that many children included in this study were below this age threshold, the conversion formula for 24 month old children was used for all participants below 24 months.

Published Psychometrics: Inter-rater reliability of the ADOS2 Module 1 at the domain level is high with agreements of 97% for the SA Domain, 79% for the RRB Domain, and 97% for the overall Total. At the item level, all items had over 80% agreement across raters. Inter-rater agreement in diagnostic classification for *autism* versus *non-spectrum* was 95%. Test-retest reliability is 92% for the SA Domain, 68% for the RRB Domain and 87% for the overall Total. Items were not correlated more than .70. Fixed effect ANOVAs comparing autism and non-spectrum individuals were conducted; ASD specific items that did not yield a statistically significant difference were eliminated.

Childhood Autism Rating Scale (CARS): The CARS (Schopler et al., 1980) is a clinician-rating checklist of autistic behaviors used to assess ASD symptom severity. The CARS consists of 15 subscales of behavior within the areas of socialization, communication, cognitive functioning, and emotional and sensory responses. Each subscale is rated based on clinical impression of the child's behavior during the assessment, and parent report of behavior, on a seven-point Likert scale using .5 increments from 1 to 4. The CARS classifies children within one of three categories based on the total score; *Severe autism* (total

score of 37 or higher), *Mild-Moderate autism* (total score 30–36.5), and *non-autistic* (total score 15–30). The Childhood Autism Rating Scale, Second Edition (CARS2; Schopler et al., 2005) was developed to better assess autism severity in high functioning children over six years of age by including a new form tailored to this group. Some toddlers in these validation studies received the CARS2; however, there are no differences between the CARS and the CARS2 Standard Form, which is for all children under age six.

Published Psychometrics: Inter-rater reliability of the CARS is .71 (mean correlation coefficient) and internal consistency is .94 (alpha). Test-retest reliability was .88 (correlation coefficient) and mean total scores were not significantly different between testing at age two and testing at age three. Criterion-related validity of the CARS demonstrated that total scores on the CARS, and clinical ratings of behavior are in high agreement with a correlation of .80. Internal consistency of the CARS is .94 (alpha) and the verification sample using the CARS2 (Standard Form) was consistent (alpha of .93). The CARS reliably differentiates children with and without Autistic Disorder at two years old (Lord, 1995), and although the CARS does not provide diagnostic cutoffs for other disorders on the spectrum, a recent study suggested that a cutoff score of 25.5 reliably distinguishes children with and without and ASD at two years old (Chlebowski et al., 2010).

Within Sample Psychometrics: A series of Cronbach's Alphas were conducted in the current sample to explore internal consistency of the CARS(2) within this group of children. The first was conducted on the entire sample, and revealed a Cronbach's Alpha of .906. The sample was then split by sex; Cronbach's Alpha was .904 within males and .907 within females.

Diagnostic and Statistical Manual of Mental Disorders-IV Checklist (DSM-IV-CL): The DSM-IV-CL is a list of all the DSM-IV symptoms of autism. The checklist was used by clinicians in the M-CHAT(-R) validation studies to determine whether a child who screened positive on the M-CHAT(-R) had an ASD, based on all information gathered from the evaluation including ASD specific diagnostic measures, cognitive, developmental adaptive assessments, and parent interview of history. If a child endorsed a symp-

tom on the checklist, the symptom was checked off, and ASD diagnostic status was determined based on DSM-IV diagnostic criteria.

3.2.3 Measures of cognitive and developmental functioning.

Mullen Scales of Early Learning (MSEL): The MSEL (Mullen, 1995) is a direct assessment of cognitive, language and motor functioning in children from birth to 68 months. The MSEL provides T-Scores (mean of 50, standard deviation of 10) and age-equivalents for five scales including Visual Reception, Fine Motor, Gross Motor, Expressive Language and Receptive Language, as well as an Early Learning Composite (ELC) based on all scales except Gross Motor. Given variability in the onset of language in toddlers, nonverbal measures of cognitive ability such as the MSEL Visual Reception scale (VR) are often used as an estimate of IQ in toddlers (Carter, 2007; Hartley and Sikora, 2011, Lord et al., 1986). Furthermore, due to insensitivity of MSEL standard scores in low functioning individuals, age-equivalent scores are often used for analyses.

Published Psychometrics: Internal consistency ranging from .75 to .83 (alpha coefficients) within each scale, test-retest reliability from .75 to .96, and inter-rater reliability from .91 to .99. The ELC has shown adequate concurrent validity with the Bayley Scales of Infant Development Mental Development Index ($r=.70$; Bayley, 1969).

3.2.4 Measures not analyzed.

Two parent report measures were administered as part of the comprehensive diagnostic evaluation, but were not analyzed in the present study. However, psychometric data for these measures are included below given that all information gathered during the evaluation was used to make the final clinical diagnosis.

Autism Diagnostic Interview, Revised (ADI-R): The ADI-R (Lord et al., 1994) is a 93-item structured parent/informant interview of past and current behavior in children from age 18 months through

adulthood. Items assess behavior within three ASD symptom domains (i.e., social, communication, and restricted, repetitive and stereotyped behavior), plus a fourth domain assessing age of symptom onset. Behaviors are coded within each domain separately, and total scores within each domain are calculated based on an algorithm. Total scores are then compared to diagnostic cutoffs, which classify individuals into one of two categories; *autism* or *non-autism*. The ADI-R does not provide diagnostic cutoffs for any other disorders on the spectrum. Toddlers in the M-CHAT(-R) validation studies received either the original ADI-R, or another ASD parent interview: The ADI-R Toddler Version (contains the original algorithm items, but removed other items that applied only to older children and included additional items pertaining to this age group) the ADI-R Short Form (contains only the original algorithm items), or the Toddler ASD Symptom Interview (TASI; based on DSM-IV diagnostic criteria; Barton et al., 2012). Although all children evaluated received a parent interview, the variability in the type of ASD parent interview administered makes it difficult to integrate data from different measures. Therefore, scores from these measures were not used as outcome variables in the current study.

Published Psychometrics: Inter-rater kappas for the ADI-R range from .63 to .89 for each item and intraclass correlations ranged from .93 to .97 for subdomain and domain scores. Test-retest reliability is high with an intraclass correlation coefficients ranging from .93 to .97 for all algorithm items. When using DSM-IV/ICD-10 ASD classification, significant differences ($p < .05$) on communication and social interaction algorithm scores between an ASD and non-ASD group were found, demonstrating good validity. However several studies examining use of the ADI-R in toddlers, have demonstrated poor diagnostic agreement between the ADI-R and other diagnostic measures (Ventola, 2006; Wiggins, 2008). Psychometrics for the TASI (Barton et al., 2012) are not available given that this is a new tool that is currently being validated.

Vineland Adaptive Behavior Scales (VABS), Survey Interview Form: The VABS (Sparrow, Balla, & Cicchetti, 1984) is a parent interview scale used to assess adaptive skills within four domains of function-

ing including communication, motor, daily living, and socialization skills in children from birth through 18 years. Open-ended interview questions are used to determine whether an individual usually, sometimes/partially, or never engages in each behavior without scaffolding from a caregiver. Standard scores (mean of 100, standard deviation of 15), and age equivalents are provided for each domain, and V-scale scores (mean of 15, standard deviation of 3), and age-equivalents are provided for each subdomain. The VABS also provides an Adaptive Behavior Composite (ABC) based on scores from all four domains. Some toddlers in these validation studies received the Vineland Behavior Scales, Second Edition, Survey Interview Form (VABS-II; Sparrow, Balla, & Cicchetti, 2005).

Published Psychometrics of the VABS: Split-half reliability is satisfactory for the four domains (ranging from .83-.97), and excellent for the ABC (ranging from .94-.99). Intercorrelations between domains range from .39 to .55, indicating only a modest overlap among domains.

Published Psychometrics of the VABS-II: Correlations between each subdomain are moderate (75% of domain comparisons have a value of .75 or greater), but are higher for younger children (ages 0 to 6). Split-half reliability within each domain is between .91 and .95, and is .97 for the ABC. The average test-retest reliability was .85 for domains (not including the 14-21 age range) and .98 for the ABC. The average inter-rater reliability was .75 for the domains and .74 for the ABC. Clinical evidence supports the use of the VABS-II as a measure of adaptive functioning in different clinical populations, including those with Intellectual Disability (ID), Autism, and ADHD.

3.3 Procedure

Pediatricians practicing in northeastern US states (primarily in Connecticut), and within 60 miles of Georgia State University in metropolitan Atlanta, Georgia were invited to offer the M-CHAT(-R) to their patients as part of a research study. After meeting with the principal investigator or a graduate student in which study procedures were described, all willing pediatricians signed a form agreeing to participate. Participating sites received an enrollment packet, which contained a description of M-

CHAT(-R) study procedures, M-CHAT(-R) forms with attached consent forms, and copies of consent forms to give to participating parents, as well as envelopes for returning screening forms to the research lab. Participating physicians were asked to invite parents at 18-month and 24-month well-child visits to participate in a research study of child development. It was requested that when introducing the study the pediatricians refrain from mentioning “autism.” This was done so as to not worry parents and to help reduce biased responses. The consent form clearly states the purpose of the study, background information, procedures, risks, benefits and compensation, voluntary participation and withdrawal, confidentiality, and contact information. Participating parents received a copy of the consent form to take home.

Parents who gave consent complete the M-CHAT(-R); completed forms and consents were sent to research staff at the local university (GSU or the University of Connecticut), where they were scored. Parents with screen positive results on the M-CHAT(-R) (indicating risk of ASD) were then called for the M-CHAT(-R) FUI. The same parent who completed the M-CHAT was administered the phone interview. In this study, a screen positive on the M-CHAT administration entailed indicating risk for ASD on both the M-CHAT(-R) questionnaire *and* the FUI. Those who screen positive were invited to visit Georgia State University or the University of Connecticut for a free diagnostic evaluation in which valuable information about the child’s development will be gathered and provided for the parents in the form of a comprehensive psychological report as well as oral feedback. Parents were informed of the benefits of receiving such an evaluation for free (typically can be very costly in private clinics); they were also assured that they will be with their child during the entire evaluation. Interested parents are scheduled for the evaluation.

Upon arrival for the evaluation, the families were greeted by a trained research staff member who explains the procedures of testing, including voluntary participation and limits of confidentiality, asked the parent to sign a consent form that allows video recording (for research and training purposes),

and answered any questions. Once video consent had been obtained, testing began. During the evaluation, a research staff member or graduate student clinician, as well as a supervising licensed psychologist completed the ADOS, ADI-R or TASI, VABS(-II), Mullen, and the CARS(2). All testing was collected in one room, with parent interviews and child measures administered simultaneously. Research staff was trained to be reliable on all measures. After testing, the clinicians took a short break to score the measures, review DSM-IV criteria, and discuss any applicable diagnoses and recommendations. Based on the results of the autism diagnostic measures, child observations, and DSM-IV criteria, the clinicians classified children in one of the following non-overlapping groups: Autistic Disorder, PDD-NOS, Language Disorder, Global Developmental Delay, other diagnosis, no diagnosis (one or more scores outside typical range, but no applicable diagnosis), or typical development. The licensed psychologist and student clinician then met with the parents for a short oral feedback session in which the child's strengths and weaknesses, diagnosis (including specific symptoms), and recommendations for treatment are described. Any questions parents may have had were also answered. Within four to six weeks, the parents received a written report describing test results, recommendations, and community outreach and treatment resources in greater detail, which was written by the lead graduate student clinician. Overall, the appointment typically lasted 3-4 hours with breaks taken as needed.

3.4 Data analyses

Data were entered into FileMaker Pro. To reduce human error, measures were double-scored and all data are double-entered by two independent persons into FileMaker Pro. Variables of interest were exported into excel and converted to SPSS where all analyses were conducted. Descriptive statistics were conducted, and correlations between variables and violations of statistical assumptions were assessed.

Standard scores on measures such as the MSEL tend to be insensitive to variability in lower functioning toddler populations due to floor effects (Carter et al., 2007). For example, toddlers often

bottom-out on the MSEL with a T-score of 20. In order to capture variability in item performance in this lower functioning cohort, previous studies have utilized age-equivalent (AE) scores instead of standard scores in analyses (Carter et al., 2007; Hartley & Sikora, 2009). Therefore, AE scores for the MSEL (Carter et al., 2007; Hartley & Sikora, 2009) were used as outcome variables in the proposed analyses.

Given the variability in the onset of language in toddlers, a nonverbal reasoning measure, such as the MSEL Visual Reception scale score, is often used as an estimate of IQ in this age group (Carter et al., 2007; Hartley & Sikora, 2009). Therefore, a nonverbal Intelligence Quotient (NVIQ) score was computed for each subject by dividing the MSEL Visual Reception AE, by chronological age, and multiplying this number by 100. This NVIQ score was used as an estimate of IQ for all pertinent analyses.

4 RESULTS

3.5 Missing data

The percent of data missing from each of the dependent variables of interest ranged from 0.3-0.8%. Since less than 5% of data is missing from each variable of interest, a listwise deletion of cases was determined to be the best technique to handle missing data as this method is suggested when less than 5% of data is missing (Tabachnick & Fidell, 2001).

3.6 Overall Results

The following sections will detail results from all primary and follow-up analyses conducted. However, it is important to note that minimal sex differences were discovered. Across all outcomes, there was a significant effect of diagnostic status such that children in the ASD group demonstrated higher scores on ASD diagnostic tools (i.e., ADOS2 and CARS(2)), and weaker scores on developmental measures (i.e., MSEL scales); however no significant main effects of sex, or interactions between sex and diagnostic status were found. The one finding contrary to this pattern emerged in the area of fine motor

skills. Results showed a significant interaction between sex and ASD status, and a main effect of ASD status, such that the effect of ASD status was greater in the non-ASD group for females, compared to males. Although significant, this finding should be interpreted with caution, and the clinical utility considered, given the small effect size.

3.7 Results for hypothesis 1: ASD symptoms

Analyses were conducted to explore the prediction that girls with ASD would demonstrate increased symptomology within the social and communication domains as measured by the ADOS2 Social Affect (SA) domain score, compared to boys, and that boys with ASD would demonstrate greater symptomology within the repetitive and stereotyped patterns of behavior domain, as measured by the ADOS2 Repetitive & Restricted Interests (RRB) domain score, compared to girls. Exploratory analyses were also conducted in order to examine whether *global* ASD symptom severity as measured by the ADOS2 Severity Score (SS) and the Childhood Autism Rating Scale (CARS) Total score would also differ based on sex. It was predicted that boys, despite ASD status, would demonstrate more severe global symptoms based on previous research exploring global symptom severity in toddlers with and without ASD (Zwaigenbaum, 2012). Exploratory analyses were also conducted to examine whether sex differences in ASD symptoms are specific to ASD, or extend to children initially considered at risk based on the M-CHAT(-R), but not diagnosed with an ASD upon further evaluation. It was postulated that within the non-ASD group, males would demonstrate greater symptoms than females across all measures given previous work supporting that boys in the general population demonstrate more ASD symptoms than girls (Auyeung et al., 2008; Auyeung et al., 2009; Baron-Cohen et al., 2001; Baron Cohen et al., 2006; Constantino & Todd, 2003; Williams et al., 2008, Zwaigenbaum et al., 2012).

In order to examine these hypotheses, a series of between subjects ANOVAs and ANCOVAs were conducted with sex and ASD status as fixed factors, and one of the aforementioned ASD measures as the outcome in each analysis. All sample sizes, means, standard deviations for each measure within group

can be found in Table 3 in Appendix B and means and standard errors are depicted in Figures 1-4 in Appendix C. Given that age was correlated with the ADOS2 RRB domain score ($r=.18, p=.001$), and CARS Total ($r=.15, p=.004$), and was independent of both fixed factors (Fields, 2009), it was justifiably included as a covariate in these analyses. In contrast, given that age was not correlated with ADOS2 SA domain score or the ADOS2 SS ($ps>.05$), it was not included as a covariate for subsequent analyses involving these variables. Exploration of outliers revealed no extreme outliers that were greater than three times the interquartile range. Levene's tests for the ADOS2 SA revealed that variances were equal across groups ($ps>.05$); however Levene's test was significant ($p<.001$) for analyses exploring ADOS2 RRB, ADOS2 SS, CARS Total indicating unequal variances across groups on these outcomes. According to Field (2009), the Levene's test is stringent, and not always the best way to judge whether variances are unequal enough to cause statistical problems. Instead, Hartley's F (also called the variance ratio) is recommended as a valid way to check (Field, 2009). The squares of the largest variance and the smallest variance were compared to Hartley's critical values for the ADOS2 RRB, ADOS2 SS and CARS Total analyses, which revealed that the differences in the variance of these variables were acceptable for conducting AN(C)OVAs with these variables.

3.7.1 Overall Results for hypothesis 1.

No interactions between sex and diagnostic status emerged across any ASD specific variables. In addition, no main effects of sex emerged; however, as expected, toddlers with ASD demonstrated higher scores than non-ASD toddlers, across all measures of ASD symptomology. Please see Table 4 in Appendix B for summarized effect sizes for all major analyses within this hypothesis.

3.7.2 Results for Hypothesis 1a: Domain Symptoms.

The ADOS2 Social Affect (SA) Domain: A two-way between subjects ANOVA exploring the effect of sex and ASD status on ADOS2 SA domain score revealed no significant interaction between sex and

ASD status (partial $\eta^2=.002$), or a main effect of sex (partial $\eta^2=.004$); however, as expected, there was a significant effect of ASD status on ADOS2 SA, such that toddlers with ASD demonstrated higher scores than toddlers without ASD, $F(1,350)=536.9$, $p<.001$, partial $\eta^2=.605$.

In order to explore the size of the effect of ASD status on ADOS2 SA score within sex, two post-hoc one-way ANOVAs were conducted; one within males only and another within females only. Both ANOVAs suggested that ASD status significantly impacts ADOS2 SA such that individuals with ASD demonstrate higher SA scores, and the effect size is comparable in males and females (males: $F(1,248)=455.2$, $p<.001$, partial $\eta^2=.648$; females: $F(1,104)=194.9$, $p<.001$, partial $\eta^2=.654$).

Although the main effect of sex was non-significant, two additional post-hoc one-way ANOVAs were conducted, one within the ASD group only and one within the non-ASD group only, in order to explore whether the size of the effect of sex differs based on ASD status. Both ANOVAs were non-significant and effect sizes were comparable across ASD status (Non-ASD: $F(1,178)=.123$, $p=.73$, partial $\eta^2=.000$; ASD: $F(1,174)=1.65$, $p=.2$, partial $\eta^2=.009$).

The ADOS2 Restricted and Repetitive Behavior (RRB) Domain: A two-way ANCOVA exploring the effect of sex and ASD status on ADOS2 RRB scores affirmed that the covariate of age accounted for a significant amount of variance in the model, $F(1,349)=8.86$, $p=.003$, partial $\eta^2=.003$. Results also revealed no significant interaction between sex and ASD status (partial $\eta^2=.001$), or a main effect of sex (partial $\eta^2=.000$); however, as expected, there was a significant effect of ASD status on ADOS2 RRB, such that toddlers with ASD demonstrated higher scores, $F(1,349)=119.36.5$, $p<.001$, partial $\eta^2=.255$.

In order to explore the size of the effect of ASD status on ADOS2 RRB score within sex, two post-hoc one-way ANCOVAs were conducted; one within males only and another within females only. Both ANCOVAs suggested that ASD status significantly impacts ADOS2 RRB such that individuals with ASD demonstrate higher RRB scores, and the effect size is comparable in males and females (males: $F(1,246)=115.6$, $p<.001$, partial $\eta^2=.320$; females $F(1,246)=40.5$, $p<.001$, partial $\eta^2=.284$).

Although the main effect of sex was non-significant, two additional post-hoc one-way ANCOVAs were conducted, one within the ASD group only and one within the non-ASD group only, in order to explore whether the size of the effect of sex differs based on ASD status. Both ANOVAs were non-significant and effect sizes were comparable across ASD status (Non-ASD: $F(1,176)=.45$, $p=.502$, partial $\eta^2=.003$; ASD: $F(1,172)=.182$, $p=.670$, partial $\eta^2=.001$).

3.7.3 Results for hypothesis 1b: Global symptom severity.

The ADOS2 Severity Score (SS): A two-way ANOVA exploring the effect of sex and ASD status on ADOS2 SS revealed no significant interaction between sex and ASD status (partial $\eta^2=.000$), or a main effect of sex (partial $\eta^2=.003$); however, as expected, there was a significant effect of ASD status on ADOS2 SS, such that individuals with ASD demonstrated higher SS, $F(1,350)=515.2$, $p<.001$, partial $\eta^2=.595$.

In order to explore the size of the effect of ASD status on SS within sex, two post-hoc one-way ANOVAs were conducted, one within males and another within females. Both ANOVAs suggested that ASD status significantly impacts ADOS2 SS such that individuals with ASD demonstrate higher SS scores, and the effect size is comparable in males and females (males: $F(1,248)=430.8$, $p<.001$, partial $\eta^2=.636$; females: $F(1,104)=207.2$, $p<.001$, partial $\eta^2=.668$). Although the main effect of sex was non-significant, two additional post-hoc, one-way ANOVAs were conducted, one within the ASD group only and one within the non-ASD group only, in order to explore whether the size of the effect of sex differs based on ASD status. Both ANOVAs were non-significant and effect sizes were comparable across ASD status (Non-ASD: $F(1,178)=.48$, $p=.491$, partial $\eta^2=.002$; ASD: $F(1,174)=.446$, $p=.51$, partial $\eta^2=.002$).

The CARS Total: A two-way ANCOVA exploring the effect of sex and ASD status on CARS Total score revealed that age significantly accounted for variance in the model ($F(348)=7.6$, $p=.006$, partial $\eta^2=.021$). The analysis also revealed no significant interaction between sex and ASD status (partial $\eta^2=.001$), or a main effect of sex ($\eta^2=.000$); however, as expected, there was a significant effect of ASD

status on CARS Total, such that individuals with ASD demonstrated higher scores, $F(1,348)=432.7$, $p<.001$, partial $\eta^2=.554$.

In order to explore the size of the effect of ASD status on CARS Total within sex, two post-hoc one-way ANCOVAs were conducted, one within males and another within females. Both ANCOVAs suggested that ASD status significantly impacts CARS Total score such that individuals with ASD demonstrate higher CARS scores, and the effect size is comparable in males and females (males: $F(1,245)=362.0$, $p<.001$, partial $\eta^2=.596$; females: $F(1,102)=161.1$, $p<.001$, partial $\eta^2=.612$).

Although the main effect of sex was non-significant, two additional post-hoc one-way ANCOVAs were conducted, one within the ASD group only and one within the non-ASD group only, in order to explore whether the size of the effect of sex differs based on ASD status. Both ANCOVAs were non-significant and effect sizes were comparable across ASD status (Non-ASD: $F(1,175)=.072$, $p=.798$, partial $\eta^2=.000$; ASD: $F(1,172)=.141$, $p=.708$, partial $\eta^2=.001$).

3.7.4 Results for hypothesis 1c: Including IQ as a covariate.

Given the debate over including IQ as a covariate in studies of neurodevelopmental disorders (Dennis et al, 2009; Volkmar et al., 1993), the aforementioned analyses within this hypothesis were also conducted with nonverbal IQ (NVIQ) included as a covariate. It is of note that when including a covariate, it is important to insure that the covariate satisfies the assumption of independence of the covariate and the treatment effect, which means that there should be no significant difference in the covariate between groups (Field, 2009). As explored in Hypothesis 2, there is a significant difference in NVIQ between the ASD and non-ASD groups and the inclusion of a covariate is not statistically justified. However, these analyses will be conducted in order to compare to previous studies, which have included IQ as a covariate. Results indicated were no changes in significance, or changes in effect sizes across all analyses within Hypothesis 1 when NVIQ was included as a covariate. Results from these analyses are summarized in Tables 5-8 in Appendix B.

3.8 Results for hypothesis 2: Intellectual functioning (IQ)

In contrast to previous studies in older children, recent findings have suggested cognitive ability does not differ based on sex in toddlers with or without ASD (Carter, 2007; Hartley & Sikora, 2009; Zwaigenbaum et al., 2012). Therefore, analyses were conducted to explore the prediction that the ASD sex ratio would be consistent regardless of level of cognitive functioning (i.e., high versus low), and that average IQ would not differ based on sex for toddlers with or without ASD.

3.8.1 Overall results for hypothesis 2.

No sex differences in the prevalence of boys and girls who were high versus low functioning across the entire sample, or within the ASD and non-ASD groups emerged. In addition, there was no interaction between sex and diagnostic status on mean NVIQ, nor a main effect of sex. However, toddlers within the ASD group were lower functioning compared to their non-ASD peers.

3.8.2 IQ-based prevalence.

In order to explore IQ-based prevalence, toddlers were categorized based on level of NVIQ (i.e., high or low). An IQ score below 70 (two standard deviations below the mean) typically classifies an individual with Intellectual Disability (called Global Developmental Delay in toddlers), and was used to classify individuals as low or high functioning for the purposes of this study; individuals with NVIQ scores below 70 were considered low functioning and individuals with NVIQ scores of 70 or above were considered high functioning.

A chi-square analysis conducted on the entire sample revealed no statistical differentiation from the predicted model in the number of males and females who were low functioning, relative to males and females who were high functioning ($\chi^2 = 1.9, p = .982$; See Table 9 in Appendix B for the number of participants falling in each quadrant for all chi-square analyses). Of all low functioning toddlers, 73.5% were males. Consistently, of all high functioning toddlers, 67.5% were males.

Examining only the ASD sample, a chi-square analyses revealed no differentiation from the predicted model in the number of males and females who were low functioning, verses males and females who were high functioning ($\chi^2 = .002, p = .962$). Of all low functioning toddlers, 79.1% were males. Consistently, of all high functioning toddlers, 78.8% were males. This suggests that the ratio of males to females does not differ based on level of cognitive functioning in the ASD sample.

An additional chi-square conducted in the non-ASD sample revealed that the number of males and females who were low functioning, verses males and females who were high functioning did not differ from the expected model ($\chi^2 = .001, p = .982$). Of all low functioning toddlers, 61.5% were males. Consistently, of all high functioning toddlers, 61.7% were males. This suggests that the ratio of males to females does not differ based on level of cognitive functioning in the non-ASD sample.

3.8.3 Mean nonverbal IQ (NVIQ).

In order to explore differences in average NVIQ, a two-way ANCOVA was conducted with sex and ASD status as fixed factors, age as a covariate, and NVIQ as the outcome. Age was significantly correlated with NVIQ ($r = -.21, p < .001$), and was independent of both fixed-factors (Fields, 2009); age was justifiably included as a covariate. No outliers were identified that were greater than three times the interquartile range, and a Levene's test revealed that variances between groups were equal ($p > .05$). Results determined that the covariate of age accounted for a significant amount of variance in the model, $F(1,351) = 14.3, p < .001$, partial $\eta^2 = .039$. Sample size, means, standard deviations for nonverbal IQ (NVIQ) within each group can be found in Table 3 in Appendix B and means and standard errors are depicted in Figure 5 in Appendix C.

There was no significant interaction between sex and ASD status (partial $\eta^2 = .007$), or a main effect of sex (partial $\eta^2 = .000$); however, there was a significant effect of ASD status on NVIQ such that toddlers with ASD demonstrated lower NVIQ scores than toddlers not diagnosed with ASD,

$F(1,351)=49.9, p<.001, \text{partial } \eta^2=.124$. Please see Table 10 in Appendix B for effect sizes for the effect of sex, effect of ASD status, and interaction between these factors on NVIQ.

In order to explore the size of the effect of ASD status on NVIQ within sex, two post-hoc one-way ANCOVAs were conducted, one within the male group only and another within female group only. Both ANCOVAs suggested that ASD status significantly impacts NVIQ such that individuals with ASD demonstrate lower IQ scores; however this effect was medium in males, and large in females (males: $F(1,247)=26.5, p<.001, \text{partial } \eta^2=.097$; females: $F(1,103)=25.7, p<.001, \text{partial } \eta^2=.200$).

Although the main effect of sex was non-significant two additional post-hoc, one-way ANCOVAs were conducted, one within the ASD group only and one within the non-ASD group only, in order to explore whether the size of the effect of sex differs based on ASD status. Both ANCOVAs were non-significant; however the effect size in the ASD group was negligible, and the effect size in the non-ASD group was considered small (Non-ASD: $F(1,177)=.79, p=.377, \text{partial } \eta^2=.004$; ASD: $F(1,173)=1.8, p=.18, \text{partial } \eta^2=.010$).

3.9 Results for hypothesis 3: Developmental skills

Analyses were conducted to explore the prediction that girls with ASD would demonstrate more impaired language and motor skills given recent findings from toddler studies that have suggested this pattern (Carter et al., 2007; Hartley & Sikora, 2009). In addition, exploratory analyses were completed in order to examine whether this pattern is specific to children with ASD, or extends to all children who are at risk for an ASD based on the M-CHAT(-R). Given recent findings that have suggested non-ASD high-risk sibling females demonstrate stronger developmental skills than their male counterparts (Zwaigenbaum, 2012), it was hypothesized that girls within the non-ASD group would demonstrate stronger language and motor skills than boys.

In order to explore these hypotheses, a series of two-way ANCOVAs were conducted with sex and ASD status as fixed factors and one of the following MSEL age-equivalent (AE) scale scores as the

dependent variable in each analysis: Receptive Language, Expressive Language, and Fine Motor. Age was correlated with MSEL Receptive Language ($r=.23, p<.001$), Expressive Language ($r=.27, p<.001$), and Fine Motor ($r=.30, p<.001$) AE scores, was independent of each fixed-factor (Fields, 2009), and therefore was justifiably included as a covariate in all analyses within this hypothesis. Exploration of outliers revealed no extreme outliers greater than three times the interquartile range, and Levene's tests revealed that variances were equal across groups for all outcome variables ($ps>.05$). Sample size, means, standard deviations for the three developmental outcomes explored within each group can be found in Table 3 in Appendix B, and means and standard deviations are depicted in Figures 6-8 in Appendix C.

3.9.1 Overall results for hypothesis 3.

With regard to language skills, there was no significant interaction between sex and ASD status on receptive or expressive skills, nor a main effect of sex. However, toddlers with ASD demonstrated weaker language skills compared to their non-ASD peers. With regard to fine motor skills, a significant interaction between sex and ASD status emerged, such that the effect of ASD status on fine motor functioning was greater in females compared to males; however, this result should be interpreted with caution and clinical utility considered given the small effect size and numerous AN(C)OVAs conducted. Please see Table 10 in Appendix B for summarized effect sizes for all major analyses within this hypothesis.

3.9.2 Results for hypothesis 3a: Language and motor skills.

Receptive language skills: An ANCOVA exploring the effect of sex and ASD status on Receptive Language AE affirmed that age was significantly related to AEs, $F(1,349)=35.2, p<.001$, partial $\eta^2=.092$. There was no significant interaction between sex and ASD status (partial $\eta^2=.005$), or a main effect of sex (partial $\eta^2=.001$); however, there was a significant main effect of ASD status on Receptive Language AE

after controlling for age, such that toddlers with ASD demonstrated weaker receptive skills, $F(1, 349)=103.0, p<.001$, partial $\eta^2=.228$.

In order to explore effect sizes of the effect of ASD status on Receptive Language AEs within sex, two post-hoc one-way ANCOVAs were conducted, one within males and another within females. Both ANCOVAs suggested that ASD status significantly impacts receptive language such that individuals with ASD demonstrate lower language scores, and the effect size is comparable in males and females (males: $F(1,245)=73.0, p<.000$, partial $\eta^2=.230$; females: $F(1,103)=40.9, p<.000$, partial $\eta^2=.284$).

Although the main effect of sex was non-significant two additional post-hoc one-way ANCOVAs were conducted, one within the ASD group only and one within the non-ASD group only, in order to explore whether the size of the effect of sex differs based on ASD status. Both ANCOVAs were not significant and effect sizes were comparable across ASD status (Non-ASD: $F(1,177)=.307, p=.580$, partial $\eta^2=.002$; ASD: $F(1,171)=1.5, p=.215$, partial $\eta^2=.009$).

Expressive language skills: An ANCOVA exploring the effect of sex and ASD status on Expressive Language AEs affirmed that age significantly contributed to variance in AEs, $F(1,351)=37.6, p<.001$, $\eta^2=.092$. There was no significant interaction between sex and ASD status (partial $\eta^2=.008$), or a main effect of sex ($\eta^2=.001$); however, there was a significant main effect of ASD status on Expressive Language AE after controlling for age, such that toddlers with ASD demonstrated weaker expressive skills, $F(1,351)=49.8, p<.001$, partial $\eta^2=.124$.

In order to explore the size of the effect of ASD status on Expressive Language AEs within sex, two post-hoc one-way ANCOVAs were conducted; one with males and another within females. Both ANCOVAs suggested that ASD status significantly impacts expressive language such that individuals with ASD demonstrate lower language scores, and the effect size is two-fold greater in females than in males (males: $F(1,247)=26.5, p<.001$, partial $\eta^2=.097$; females: $F(1,103)=24.3, p<.001$, partial $\eta^2=.191$).

Although the main effect of sex was non-significant two additional post-hoc one-way ANCOVAs were conducted, one within the ASD group only and one within the non-ASD group only, in order to explore whether the size of the effect of sex differs based on ASD status. Both ANCOVAs were non-significant; however the effect size was negligible in the non-ASD group and was considered small in the ASD group (Non-ASD: $F(1,177)=.61$, $p=.44$, partial $\eta^2=.003$; ASD: $F(1,173)=2.98$, $p=.086$, partial $\eta^2=.017$).

Fine motor skills: An ANCOVA exploring the effect of sex and ASD status on Fine Motor scale AEs affirmed that age contributed to a significant amount of variance in fine motor skills, $F(1,350)=42.4$, $p<.001$, partial $\eta^2=.108$. Results also revealed a significant interaction between sex and ASD status after controlling for age, such that females in the non-ASD group demonstrated stronger fine motor skills than boys without ASD, but this effect was reversed in the ASD group such that boys demonstrated stronger motor skills compared to females, $F(1,350)=4.0$, $p=.047$, partial $\eta^2=.011$. There was no significant main effect of sex ($\eta^2=.001$); however, there was a significant main effect of ASD, $F(1,350)=30.4$, $p<.001$, partial $\eta^2=.080$, which must be interpreted in the context of the interaction, such that the effect of diagnostic status was greater in the non-ASD group for females compared to males.

3.9.3 Results for hypothesis 3b: Including IQ as a covariate.

In order to explore whether including IQ as a covariate changes outcomes, the aforementioned analyses within this hypothesis were also conducted with NVIQ as a covariate. As previously stated, when including a covariate, it is important to insure that the covariate satisfies the assumption of independence of the covariate and the treatment effect, which means that there should be no significant difference in the covariate between groups (Fields, 2009). Hypothesis 2 revealed a significant difference in NVIQ between diagnostic groups; therefore, the inclusion of a covariate is not statistically justified. However, these analyses will be conducted in order to compare to previous studies, which have included IQ as a covariate. Results indicated that when NVIQ was added to the follow-up analyses exploring the size of the effect of ASD status on MSEL Expressive Language within sex, the effect size in males

went from a large effect ($\eta^2=.097$) to a small effect ($\eta^2=.022$), and the effect size in females changed from a large effect ($\eta^2=.191$) to a medium effect ($\eta^2=.060$). No changes in significance, or other changes in effect sizes were observed across all analyses within Hypothesis 3. Results are summarized in Tables 11-13 in Appendix B.

5 DISCUSSION

One of the most consistent findings in the ASD literature is that the prevalence is about four times greater in males than in females. Despite this, few studies have explored sex differences in the clinical phenotype of children with ASD, and findings from existing studies are inconsistent. Even fewer studies of sex differences have been conducted in very young children with ASD, and only one study (to the author's knowledge) has extended the study of sex differences to children at risk for ASD, with and without the disorder (Zwaigenbaum et al, 2012). It is important to study sex differences in toddlers as findings may have implications for early screening, diagnosis and intervention; this study addressed whether there are sex differences in the clinical profiles of toddlers considered at risk for ASD based on the M-CHAT(-R), and whether these differences are specific to ASD, or can be generalized to all those initially at risk for an ASD based on the M-CHAT(-R). These questions were explored within three aspects of the ASD clinical profile including ASD symptoms, IQ, and developmental skills. Analyses were conducted both with and without the inclusion of IQ as a covariate given inconsistent use of this covariate in studies of sex differences in ASD despite recent compelling arguments against using IQ as a covariate in studies of neurodevelopment (Dennis et al., 2009).

3.10 Discussion of findings: ASD symptoms

With regard to ASD symptoms, sex differences did not emerge in either diagnostic group; however, as expected, children diagnosed with ASD scored higher than children with non-ASD on all diagnos-

tic measures. Results were similar when nonverbal IQ (NVIQ) was controlled. Compared to results from recent studies conducted in toddlers, these findings are consistent with results from Carter et al. (2007), but differ from Hartley and Sikora (2009), which replicated findings from previous work in older individuals with ASD and found that males demonstrated more repetitive and restricted interests than females. The current data are also inconsistent with Zwaigenbaum and colleagues' (2012) study of sex differences in high-risk toddler siblings with and without ASD, and typical toddler controls, as in their sample, boys endorsed more severe symptoms on the ADOS2, and higher scores on the social and communication domains of the ADI-R than girls. This effect was not specific to ASD, but generalized to all children with and without ASD. Interestingly, data from other studies of sex differences in ASD symptomology in non-ASD and non-developmentally delayed samples, have suggested that boys in these groups generally demonstrate more ASD traits than girls (Allison et al., 2008; Auyeung et al., 2008; Baron-Cohen et al., 2001; Baron Cohen et al., 2006; Ludwig et al., IMFAR, 2012; Park et al., 2012; Zwaigenbaum et al., 2012); however, these sex differences have not emerged in developmentally delayed children (Rivet & Matson; 2011; Sipes et al., 2011). Given that 74% of the non-ASD toddlers in the current sample were developmentally delayed, the lack of sex differences in ASD symptoms for the non-ASD sample is consistent with Rivet & Matson's (2011), and Sipes and colleagues' (2011) findings. Zwaigenbaum and colleague's sample of non-ASD high-risk siblings and typical controls were generally not delayed and were substantially higher functioning (MSEL Early Learning Composite Standard Scores of 109.6 and 120.4 for the non-ASD high-risk siblings and typical controls, respectively) than the current non-ASD sample (MSEL Early Learning Composite = 73.2) and may explain why current results within the non-ASD sample are inconsistent with their work. Overall, the current study supports that there are no sex differences in ASD domain symptoms or overall global in toddlers identified by the M-CHAT(-R), despite ASD status.

It is important to note that this sample is very young and lower functioning than older samples, and therefore these findings cannot be generalized to older and higher functioning individuals with ASD.

Given that studies exploring older samples of individuals with ASD have demonstrated sex differences in some areas of the behavioral presentation, future longitudinal studies should explore the developmental trajectory of the emergence of potential sex differences in the clinical phenotype of ASD, which will hopefully shed light on the equivocal findings within this body of literature.

3.10.1 Methodological differences that may have contributed to inconsistent findings in ASD

symptoms across toddler studies

Consistency between current findings and previous studies of sex differences in ASD symptoms within toddler samples is mixed, and it is important to explore how inconsistent methods used to quantify symptoms may affect these differences.

Different versions of the ADOS: The ADOS2 was not developed until after Carter et al. (2007) and Hartley and Sikora (2009) were published, and therefore both studies utilized the original ADOS Social, Communication, and Restricted, Repetitive and Stereotyped Behaviors Domain algorithm scores to compare ASD symptoms across sex. Each Domain algorithm score on the ADOS is comprised of a specific set of algorithm items within that domain that best predict an autism diagnosis. The ADOS2 only has two domain scores: one includes a new set of algorithm items collapsed over the social and communication symptom domains (Social Affect; SA), and the other domain includes algorithm items that fall within repetitive and restricted patterns of behavior domain (Repetitive and Restricted Behaviors; RRB). In addition, the specific algorithm items used to compute these domain total scores differ based on whether the child uses words or not. The changes in algorithm items, in addition to the collapsing of the social and communication domains, make it difficult to compare the current findings with previous studies in toddlers that have used the original ADOS (i.e., Carter et al., 2007 and Hartley & Sikora, 2009). Despite these changes, results pertaining to social and communication symptoms from both Carter and colleagues (2007) and Hartley and Sikora (2009), are consistent with current results; no sex differences in social and communication symptoms emerged. In contrast, changes in item content within the RRB do-

main between versions of the ADOS may contribute to discrepant findings across toddler studies observed within this domain; Hartley and Sikora (2009) found that boys demonstrated more repetitive and restricted interests, whereas the current study and Carter and colleagues (2007) found no sex differences. It is important to note that consistent with the current study, Zwaigenbaum and colleagues (2009) used the ADOS2, yet differences in findings across studies persist. This suggests that there may be other factors contributing to inconsistent findings pertaining to ASD symptoms across studies.

Various recruitment methods: Another reason results may be inconsistent with previous toddler work pertains to the different methods by which toddlers were recruited across studies. Carter and colleagues (2007) recruited children who had already received an autism diagnosis, and Hartely and Sikora (2009) recruited toddlers who were referred to an interdisciplinary autism clinic by their primary care physician due to concerns about an ASD. In contrast to these recruitment methods, the M-CHAT(-R) is able to detect ASD prior to parent and/or doctor concerns in some children (Robins, 2008), and therefore may identify children who are phenotypically different from clinic referred samples. For example, children who screen positive on the M-CHAT may demonstrate subtle deficits that are not easily identified by pediatricians, but are detected by the measure. Because of this screening strategy, the current sample is younger (mean age of 25.3 months) than both Carter et al. (2007; mean age of 28 months) and Hartley and Sikora (2009; mean age of 35.7 months), which also may contribute to different sex based clinical profiles compared to older samples, especially if the emergence of sex specific differences in the ASD phenotype is age dependent. For instance, some of the clinical symptoms do not pertain to very young toddlers, such as conversational skills. If one sex demonstrates greater impairment in this skill area, sex differences would not be observed until later in toddlerhood when these symptoms typically emerge. In order to explore whether a different pattern of sex differences may emerge with development, future work will examine sex differences in this sample as they approach the preschool years (42-48 months).

Type of risk: High-risk siblings verses low-risk population screening: Risk status of participants in the sample is also important to consider when comparing findings across studies. Zwaigenbaum and colleagues' (2012) design is most similar to the current study because it explores sex differences in children both with and without ASD who were initially at risk; however children in their study were at risk due to ASD sibling status, a sample that may be phenotypically different from children at risk based on M-CHAT(-R) screen positive status. For example, the increased genetic liability in sibling samples could influence the clinical presentation of ASD (Zwaigenbaum et al., 2007), and potentially the sex specific presentation of ASD, compared to children who are at risk based on the M-CHAT(-R), most of whom do not have an older sibling with ASD. These two groups of children may have a number of differences that have not yet been investigated or able to be detected by current clinical tools (Zwaigenbaum et al., 2007), which should be considered when integrating results across studies. Given that little research has compared high-risk siblings and risk status based on early screening, future research is warranted that will explore potential differences that may characterize the ASD phenotype based on mechanisms by which children are identified.

The characteristics of the non-ASD samples in both Zwaigenbaum et al. (2012) and the current study should also be considered when comparing findings across studies, specifically within the non-ASD groups. Zwaigenbaum and colleagues explored sex differences in a non-ASD high risk sibling sample and also a typical control sample. Both of Zwaigenbaum et al.'s (2012) non-ASD sample differ greatly from the current M-CHAT(-R) screen positive, non-ASD sample, given that children who screen positive on the M-CHAT(-R) and are not diagnosed with an ASD are typically diagnosed with some other developmental delay (74% in the current sample), whereas toddlers in Zwaigenbaum and colleagues (2012) non-ASD samples were generally not developmentally delayed. Inspection of MSEL Early Learning Composite scores revealed that non-ASD high-risk siblings demonstrated a mean score in the Average range ($SS=109.6$), and the typical controls demonstrated a mean score in the High Average range ($SS=120.4$),

compared to a Borderline mean score in the current non-ASD sample ($SS=73.2$). These cognitive differences should be taken into consideration when comparing findings across these study samples.

3.11 Discussion of findings: IQ

With regard to IQ, the current data supports that toddlers with ASD who are identified with a Level 1 (low-risk) screening tool may be lower functioning than children ascertained in other ways, suggesting that the children detected through primary care screening are a subset of the whole ASD population. In the current sample, 63% of boys and 62% of girls presented with comorbid intellectual disability (ID; $NVIQ < 70$); however a recent surveillance study conducted by the CDC (2012) found much lower rates of comorbid ID in a sample of eight-year-olds on the autism spectrum, such that only 37% of males and 46% of females presented with comorbid ID. Findings from the current study also indicated that toddlers with ASD were lower functioning than those in the non-ASD group, but no sex differences in the number of children who were high versus low functioning, nor in nonverbal IQ scores emerged across diagnostic groups. The lack of sex differences in IQ within the ASD group are consistent with findings from the three most recent toddler studies (Carter et al., 2007; Hartley & Sikora, 2009; Zwaigenbaum et al., 2012), but are inconsistent with previous literature conducted in older populations of individuals with ASD, which have suggested that females tend to be lower functioning than males. It is important to note that the MSEL Visual Reception (VR) scale was used to compute a nonverbal IQ score, a common practice in toddler samples given the variability in language development at this age. However, using the MSEL VR scale is not a perfect way to estimate nonverbal IQ because this scale has been shown to rely on receptive language skills, specifically in toddlers evaluated for ASD based on M-CHAT(-R) screen positive status. For example, Anderson, Robins and Adamson (2013) found that the MSEL Receptive Language scale score accounted for a significant amount of variance in a subsample of children included in the current sample. Therefore, it is important to consider the effect of language variability on NVIQ in this study; it is possible that the communication delays that are hallmark impairments in ASD, may have

contributed to lower NVIQ scores in this study, and in previous studies exploring sex differences in toddler samples.

3.11.1 Discussion about the lack of sex differences in IQ in toddler samples.

If girls truly are lower functioning in older samples, one possible explanation for this disparate finding across age cohorts is that toddler samples represent a higher proportion of girls than the proportion of girls in the entire ASD population because lower functioning individuals tend to be identified earlier (Eaves & Ho, 2004; Shattuck et al., 2009). This idea also fits with research that has suggested an increased male to female ratio in higher functioning individuals and a decreased ratio in lower functioning individuals (Fombonne, 2003; Wing, 1981). If a higher proportion of females are identified in toddlerhood compared to the proportion of females with ASD across the lifespan, it would suggest that sex differences in IQ would not emerge until later in life as increased numbers of higher functioning males are diagnosed. Boys may initially be missed due to the higher likelihood of compensatory strategies in boys compared to females that keep them from meeting the clinical threshold at an early age. Another notion is that boys and girls are not missed differentially in toddlerhood, rather, higher functioning females who are initially missed possess even greater compensatory mechanisms than boys, such that symptoms never reach clinical significance even into adulthood. Support for this postulation includes recent evidence that girls require a greater ASD genetic load to manifest the same amount of deficits as to boys (Robinson et al., 2013), thereby suggesting some kind of compensatory mechanism that reduces the clinical manifestation of impairment. Overall, findings from studies of sex-based differences in IQ in toddlers with ASD may not necessarily be representative of all individuals with ASD across the lifespan.

Future research will examine whether there is a greater proportion of high functioning boys to girls whom were missed by the M-CHAT(-R), that is, children who had initially screened negative in toddlerhood, but were later diagnosed with an ASD, as this may facilitate understanding about the nuanced role of the effect of sex on IQ across different age cohorts. In addition, future research is warranted to

investigate sex differences in the prevalence and presentation of Broader Autism Phenotype (Piven et al., 1997), which characterizes individuals with subthreshold ASD symptoms, will help elucidate whether there are differences in compensatory strategies between males and females, or whether higher functioning females are just more difficult to detect than high functioning boys, given that our current diagnostic criteria are based primarily on the male presentation of the disorder.

3.12 Discussion of findings: Developmental skills

3.12.1 Language skills.

With regard to language development, there were no sex differences in either diagnostic group; however, the ASD group demonstrated weaker receptive and expressive skills than the non-ASD group. Findings pertaining to language skills in the ASD group are consistent with results in Hartley and Sikora (2009), but Carter and colleagues (2007) found that language abilities were higher in males compared to females with ASD. In addition, current results are contrary to Zwaigenbaum and colleagues' (2013) finding that girls demonstrated better language skills than boys despite sibling risk status and ASD status.

3.12.2 Motor skills.

In contrast to language skills, sex differences emerged in the area of motor skills; results suggest that there is a significant interaction between sex and ASD status on fine motor skills, such that the effect of ASD status was greater in females compared to males. However, given the small effect size of the current finding, and large number of ANOVAs conducted, it is important to consider that the current significant interaction could be spurious, and even if this finding represents a true difference, the clinical utility should be considered given the small effect size. Despite this, it is important to note that previous findings published on sex differences in toddlers with ASD also have revealed small effect sizes. For example, current results are contrary to Zwaigenbaum and colleagues' (2013) finding that girls demon-

strated better motor skills than boys, despite risk status and ASD status; however the effect size of this findings was also small (partial $\eta^2 = .03$).

Although readers are encouraged not to place great emphasis on this finding, future work in the area of ASD sex differences in the development of fine motor skills is warranted given that some researchers suggest that very early motor impairments are more common in children with ASD than low-risk children, and are predictive of later communication impairments (Bhat, Galloway & Landa, 2012). This suggests that early motor issues specific to ASD may a very early behavioral sign of the disorder, and identification of these motor impairments may enhance early screening practices. Interestingly, recent preliminary findings from our lab conducted in a sample of 18,742 children screened on the M-CHAT screening tool suggest that there may be sex differences in the efficacy of motor items. These preliminary findings suggest that the gross motor items are among the best predictors of an ASD diagnosis in females, but not males (Ludwig et al., in preparation). Unfortunately, the present study is limited because we were unable to explore differences in gross motor skills given that nearly half of the children in the sample were missing MSEL Gross Motor scores. This is because the Gross Motor scale is optional to compute the overall Early Learning Composite (ELC) and in the current sample, this scale was typically administered only when time permitted and when child compliance was adequate for valid administration. Although this preliminary finding (Ludwig et al., in preparation) pertains to gross motor skills (the M-CHAT does not include fine motor items) these data combined with findings from the present study suggest that sex differences in motor abilities may be an important area for future work, especially in the context of early ASD screening.

3.12.3 *Interpretation of scores with large standard errors.*

Upon examination of the standard errors of the means, it appears that errors are relatively large for several variables. This suggests that scores are highly variable, and means may not be a good representative of central tendencies. It is of note that standard errors for the ADOS2 RRB domain scores were

especially large. This may be due to the high variability in the presentation of repetitive and restricted exhibited in during the 45-minute ADOS session, or perhaps the heterogeneity in symptom presentation across the ASD spectrum. However, compared to other studies of sex differences in toddlers (Carter et al., 2007, Hartely & Sikora, 2009, Zwaigenbaum et al., 2012), standard error sizes are comparable. Therefore, it may be that highly variable ASD symptoms and developmental skills are characteristic of this population of children, and future research should explore how high standard errors may affect the utility of means and the accuracy various statistical analyses when used in young children with ASD.

3.13 Including an IQ as a covariate: Does it change anything?

There has been debate in the literature as to whether inclusion of IQ as a covariate is justified in studies of neurodevelopmental disorders (Dennis et al., 2009; Volkmar et al., 1993). Volkmar and colleagues (1993) argue that a covariate must be conceptualized as a cause rather than an outcome, and Dennis and colleagues (2009) make a compelling case that IQ is not a cause, but rather an outcome of neurodevelopmental disorders such as autism, and to partial out IQ would be partialling out the effect of the disorder. The current authors concur with Dennis and colleagues' (2009) that IQ is not a cause of sex differences in IQ, rather ASD may manifest differentially in boys and girls, thereby contributing to differences in performance on tasks we use to measure IQ (in this case MSEL Visual Reception). The current authors agree with Dennis and colleagues (2009) not only due to the theoretical evidence discouraging the use of IQ as a covariate, but also to increase the clinical utility of results of studies of sex differences. For example, screening tools like the M-CHAT(-R) do not account for differences in IQ in their clinical scoring systems. Therefore, providing results of sex differences that may exist when IQ is controlled may not provide information that can be easily integrated into the improvement of these tools without requiring some measure of IQ be included into the diagnostic algorithm, which would decrease the feasibility of widespread dissemination of these tools as use would require some measure of IQ. In contrast, providing data about sex differences that exist without the control of IQ provides information

that can be used to directly improve early screening and diagnostic tools. Despite these views, in order to compare current findings to previous investigations of sex differences in toddler profiles, the current study conducted analyses both with and without IQ included as a covariate. This was done even though the assumption of independence of the covariate and the treatment effect was not met, given that mean NVIQ scores were significantly different across ASD status (Fields, 2009). Qualitative comparison of both sets of analyses revealed minimal differences in the analyses conducted. This may suggest that the inclusion of IQ as a covariate does not change statistical results in studies of sex differences; however, it is important to remember that the inclusion of this covariate violated one of the statistical assumptions necessary for the inclusion of IQ as a covariate, and results should be interpreted with caution. Future research should continue to explore the justification of IQ as a covariate in studies of sex differences in ASD and the way in which including IQ may impact findings.

3.14 Conclusion

This study contributes to the literature because it is the first to explore sex differences in a very young sample of children considered at risk for an ASD based on a widely used early screening tool. In summary, findings from the current study revealed minimal sex differences across multiple measures of the ASD symptoms, nonverbal IQ, and developmental skills in children who are considered at risk for ASD based on the M-CHAT(-R). The only sex difference observed was that the effect of ASD status on fine motor skills was reduced in boys compared to girls. As expected, children with ASD demonstrated greater symptoms across all ASD symptom measures compared to children who are initially at risk based on the M-CHAT(-R), but who are not diagnosed upon further evaluation; however they also demonstrated weaker intellectual abilities and language skills. These data suggest that there may not be true sex differences in ASD symptoms, intellectual abilities and developmental skills this early on in the progression of ASD that can be detected by our current direct observation tools. Given the limitations of our

current observational tools, future work should explore potential genetic and endophenotypic sex differences that may be useful in improving early identification and interventions for both boys and girls.

REFERENCES

- American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.). Washington, DC: Author.
- Allison, C., Baron-Cohen, S., Wheelwright, S., Charman, T., Richler, J., Pasco, G. et al. (2008). The QCHAT (quantitative checklist for autism in toddlers): A normally distributed quantitative measure of autistic traits at 18–24-months of age: Preliminary report. *Journal of Autism and Developmental Disorders*, 38(8), 1414–1425.
- Anderson, S, Robins, D. L., Adamson, L. (2013). Inter-related Visual Reception and Receptive Language Scales of the Mullen Scales of Early Learning. Poster presented at the International Neuropsychology Society Annual Meeting. Hawaii, USA.
- Auyeung, B., Baron-Cohen, S., Wheelwright, S., Allison, C. (2007). *The Autism Spectrum Quotient: Children's Version (AQ-Child)*. *Journal of Autism and Developmental Disorders*, 38(7), 1230-1240.
- Auyeung, B., Wheelwright, S., Allison, C., Atkinson, M., Samarawickrema, N., Baron-Cohen, S. (2008). *The Children's Empathy Quotient: Sex differences in developmental and in autism spectrum disorders*. *Journal for Autism and Developmental Disorders*, 39, 1509-1521.
- Baron-Cohen, S., Hoekstra, R., Knickmeyer, R., Wheelwright, S. (2006). The Autism-Spectrum Quotient (AQ) Adolescent version. *Journal of Autism and Developmental Disorders*, 36, 343–350.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31, 5–17.
- Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E. (2001). The Autism Spectrum Quotient (AQ): Evidence from Asperger syndrome/high functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*, 31, 5–17.
- Barton, M., Boorstein, H., Dumont-Mathieu, T., Herlihy, L., Fein, D. (2012). Toddler ASD Symptom Inter-

view. Self-published.

- Bhat, A. N., Galloway, J. C., & Landa, R. J. (2012). Relation between early motor delay and later communication delay in infants at risk for autism. *Infant Behavior and Development, 35*(4), 838-846.
- Bolte, S., Duketis, E., Poustka, F., Holtman, M., (2011). Sex differences in cognitive domains and their clinical correlates in higher-functioning autism spectrum disorders. *Autism, 15*, 497-511.
- Bryson, S. E., Clark, B. S., Smith, I. M. (1988). First report of a Canadian epidemiological study of autistic syndromes. *Journal of Child Psychology and Psychiatry, 4*, 433-445.
- Carter, A. S., Black, D. O., Tewani, S., Connolly, C. E., Kadlec, M. B., Tager-Flusberg, H. (2007). Sex differences in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*, 86-97.
- Centers for Disease Control and Prevention. (2007). Prevalence of autism spectrum disorder: Autism and developmental disability monitoring network, 14 sites, United States, 2002. *Morbidity and Mortality Report, 56*.
- Centers for Disease Control and Prevention. (2012). Prevalence of autism spectrum disorders: Autism and developmental disability monitoring network, 14 sites, United States, 2008. *Morbidity and Mortality Report, 61* (SS03), 1-19.
- Chlebowski, C., Green, Barton, M., Fein, D. (2010). Using the Childhood Autism Rating Scale to diagnose autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*, 787-799.
- Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-Scale Use of the Modified Checklist for Autism in Low-Risk Toddlers. *Pediatrics, 131*(4), e1121-e1127.
- Constantino, J. N., Todd, R. D. (2003). Autistic traits in the general population: A twin study. *Archives of General Psychiatry, 60*(5), 524-530.
- Dattalo, P. (2008). *Determining Sample Size: Balancing Power, Precision, and Practicality*. Oxford: Oxford University Press.

- Dennis, M., Francis, D.J., Cirino, P.T., Schachar, R., Barnes, M.A., Fletcher, J.M. (2009). Why IQ is not a covariate in cognitive studies of neurodevelopmental disorders. *Journal of the International Neuropsychological Society*, 15, 331–343.
- Dworzynski, K., Ronald, A., Bolton, P., & Happé, F. (2012). How Different Are Girls and Boys Above and Below the Diagnostic Threshold for Autism Spectrum Disorders? *Journal of the American Academy of Child & Adolescent Psychiatry*.
- Eaves, L.C. & Ho, H.H. (2004). The very early identification of autism: Outcome to age 41/2–5. *Journal of Autism and Developmental Disorders*, 34, 367–378.
- Ehlers, F., Gillberg, C., Wing, L. (1999) A screening questionnaire for Asperger Syndrome and other high functioning autism spectrum disorders in school age children. *Journal of Autism and Developmental Disorders*, 29, 129-142.
- Eikeseth, S., Klintwall, L., Jahr, E., Karlsson, P. (2012). Outcome for children with autism receiving early and intensive behavioral intervention in mainstream preschool and kindergarten settings. *Research in Autism Spectrum Disorders*, 6(2), 829-835.
- Eldevik, S., Hastings, R.P., Hughes, J.C., Jahr, E., Eikeseth, S., Cross, S. (2009). meta-analysis of early intensive behavioral intervention for children with Autism. *Journal of Clinical Child & Adolescent Psychology*, 38(3), 439-450.
- Faul, F., Erdfelder, E., Lang, A.G., Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191.
- Field, A. (2009). *Discovering statistics using SPSS*. Sage Publications Limited.
- Fombonne, E. (2003). Epidemiological surveys of autism and other pervasive developmental disorders: An update. *Journal of Autism and Developmental Disorders*, 33, 365–382.

- Giarelli, E., Wiggins, L.D., Rice, C.E., Levy, S.E., Kirby, R.S., Pinto-Martin J., Mandell, D. (2010). Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and Health Journal*, 3(2), 107-16.
- 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, 693-705.
- Hartley, S. L., & Sikora, D. M. (2009). Sex differences in autism spectrum disorder: An examination of developmental functioning, autistic symptoms, and coexisting behavior problems in toddlers. *Journal of Autism and Developmental Disorders*, 39(12), 1715-1722.
- Holtmann, M., Bolte, S., Poustka, F. (2007). Autism spectrum disorders: Sex differences in autistic behavior in autistic behavior domains and coexisting psychopathology. *Developmental Medicine and Child Neurology*, 49, 361–366.
- Holtmann, M., Bolte, S., Poustka, F. (2007). Autism spectrum disorders: Sex differences in autistic behavior in autistic behavior domains and coexisting psychopathology. *Developmental Medicine and Child Neurology*, 49, 361–366.
- Honda, H., Shimizu, Y., Imai, M., Nitto, Y. (2005). Cumulative incidence of childhood autism: a total population study of better accuracy and precision. *Developmental Medicine and Child Neurology*, 47, 10–18.
- Horovitz, M., Matson, J., Turygin, N., Beighley, J.S. (2012). The relationship between gender and age of first concern in toddlers with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 6, 466-471.
- Hus, V., Pickles, A., Cook, 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, 438-448.
- Krug, D. A., Arick, J. R., Almond, P. G. (1980). Behavior checklist for identifying severely handicapped in-

- individuals with high levels of autistic behavior. *Journal of Child Psychology and Psychiatry*, 21, 221-229.
- Lai, M., Lombardo, M.V., Pasco, G., Ruigrok, A.N., Wheelwright, S.J. et al. (2011). A behavioral comparison of male and female adults with high functioning autism spectrum conditions. *PLOS One*, 6(6).
- Le Couteur, A., Rutter, M., Lord, C., Rios, P. (1989). *Autism Diagnostic Interview: A standardized investigator-based instrument. Journal of Autism and Developmental Disorders*, 19, 363–387.
- Lord C, Rutter, M., Le Couteur A. (1994). *Autism Diagnostic Interview – Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorder. Journal of Autism and Developmental Disorders*, 24, 659–685.
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C. et al. (2000). *The Autism Diagnostic Observation Schedule-Generic: A standard measure of social and communication deficits associated with the spectrum of autism. Journal of Autism and Developmental Disorders*, 30, 205–223.
- Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule: ADOS-2*. Western Psychological.
- Lord, C., Schopler, E., Revicki, D. (1982). Sex differences in autism. *Journal of Autism and Developmental Disorders*, 12(4), 317–330.
- Ludwig, N., Robins, D.L., Fein, D.A. (May, 2011). Oral Presentation: Sex Differences in the early screening of autism spectrum disorders. International Meeting for Autism Research (IMFAR), San Diego, California.
- Ludwig, N., Robins, D.L. Fein, D.A. (May, 2012). Oral Presentation: Examining differences in item endorsement on the Modified Checklist for Autism in Toddlers (M-CHAT). International Meeting for Autism Research (IMFAR), Toronto, Canada.

- Mandy, W., Chilvers, R., Chowdhury, U., Salter, G., Seigal, A., & Skuse, D. (2012). Sex differences in autism spectrum disorder: evidence from a large sample of children and adolescents. *Journal of autism and developmental disorders*, 42(7), 1304-1313.
- Matson, J. L., Wilkins, J., Sevin, J. A., Knight, C., Boisjoli, J. A., Sharp, B. (2009). Reliability and item content of the baby and infant screen for children with autism traits (BISCUIT): Parts 1–3. *Research in Autism Spectrum Disorders*, 3(2), 336–344.
- McLennan, J. D., Lord, C., Schopler, E. (1993). Sex differences in higher functioning people with autism. *Journal of Autism and Developmental Disabilities*, 23, 217–227.
- Mullen, E. (1995). *Mullen Scales of Early Learning*. AGS Edition. Circle Pines, MN: American Guidance Service, Inc.
- Newborg, J. (2005). *Battelle Developmental Inventory* (2nd ed.). Itasca, NY: Riverside.
- Park, S., Cho, S., Cho, I.H., Kim, B., Kim, J., Shin, M. et al. (2012) Sex differences in children with autism spectrum disorders compared with their unaffected siblings and typically developing children. *Research in Autism Spectrum Disorders*, 6, 861-870.
- Pilowsky, T., Yirmiya, N., Sulman, C., Dover, R. (1998). The *Autism Diagnostic Interview-Revised* and the *Childhood Autism Rating Scale*: Differences between diagnostic systems and comparison between gender. *Journal of Autism and Developmental Disorders*, 28, 143–151.
- Piven, J., Palmer, P., Jacobi, D., Childress, D., & Arndt, S. (1997). Broader autism phenotype: evidence from a family history study of multiple-incidence autism families. *American Journal of Psychiatry*, 154(2), 185-190.
- Reichow, B. (2012). Overview of meta-analyses on early intensive behavioral intervention for young children with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 42, 512-520.
- Rivet, T.T., Matson, J.M. (2011). Gender differences in core symptomology in autism spectrum disorders

- across the lifespan. *Journal of Developmental and Physical Disabilities*, 23, 399-420.
- Robins , D.L., Fein, D., Barton, M. (1999 a). *Modified Checklist for Autism in Toddlers (M-CHAT)*. Self-published.
- Robins , D.L., Fein, D., Barton, M. (1999 b). *Modified Checklist for Autism in Toddlers (M-CHAT) Follow-Up Interview*. Self-published.
- Robins, D. L. (2008). Screening for autism spectrum disorders in primary care settings. *Autism*, 12(5), 537-556.
- Robins , D.L., Fein, D., Barton, M. (2009 a). *Modified Checklist for Autism in Toddlers, Revised (M-CHAT-R)*. Self-published.
- Robins , D.L., Fein, D., Barton, M. (2009 b). *Modified Checklist for Autism in Toddlers (M-CHAT-R) Follow-Up Interview*. Self-published.
- Robins, D.L., Fein, D. (May, 2011) Preliminary findings from the *Modified Checklist for Autism in Toddlers, Revised (M-CHAT-R)*. International Meeting for Autism Research (IMFAR), San Diego, California.
- Robins, D.L., Pandey, J., Chlebowski, C., Carr, K., Zaj, J., Arroyo, M., Barton, M., Green, J., Fein, D. (May, 2010). *M-CHAT Best7: A New Scoring Algorithm Improves Positive Predictive Power of the M-CHAT*. International Meeting for Autism Research (IMFAR), Philadelphia, PA.
- Robinson, E. B., Lichtenstein, P., Anckarsäter, H., Happé, F., & Ronald, A. (2013). Examining and interpreting the female protective effect against autistic behavior. *Proceedings of the National Academy of Sciences*.
- Schopler, E., Reichler, R. (1979). *Psychoeducational Profile: Individualized assessment and treatment for autistic and developmentally delayed children*. Baltimore: University Park Press.
- Schopler, E., Reichler, R., DeVellis, R., Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders*, 10, 91-103.

- Schopler, E., Van Bourgondien, M.E., Wellman, G.J., Love, S.R. (2005). *The Childhood Autism Rating Scale, Second Edition*. Los Angeles: Western Psychological Services.
- Shattuck, P. T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D., Wiggins, L., et al., (2008). Timing of identification among children with an autism spectrum disorder: Findings from a population-based surveillance study. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48 (5), 474-483.
- Sigman, M., McGovern, C.W. (2005). Improvement in cognitive language skills from preschool to adolescence in autism. *Journal of Autism and Developmental Disorders*, 35(1), 15-23.
- Sipes, M., Matson, J., Worley, J.A., Kozlowski, A.M. (2011). Gender differences in symptoms of Autism Spectrum Disorders in toddlers. *Research in Autism Spectrum Disorders*, 5, 1465-1470.
- Sparrow, S. S., Balla, D. A., Cicchetti, D. V. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service, Inc.
- Sparrow, S. S., Cicchetti, V. D., Balla, A. D. (2005). *Vineland adaptive behavior scales (2nd ed.)*. Circle Pines, MN: American Guidance Service, Inc.
- Szatmari, P., Liu, X. Q., Goldberg, J., Zwaigenbaum, L., Paterson, A.D., Woodbury-Smith, M. et al. (2011). Sex differences in repetitive stereotyped behaviors in autism: Implications for genetic liability. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics*, 159B(1), 5–12.
- Tabachnick B.G., Fidell, L.S., (2001). *Using Multivariate Statistics (4th ed.)*
Boston, MA: Allyn & Bacon.
- Ventola, P.E., Kleinman, J., Pandey, J., Barton, M., Allen, S., Green, J et al. (2006). Agreement among four diagnostic instruments for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 36, 839-847.
- Volkmar, F.R., Szatmari, P., Sparrow, S.S. (1993). Sex differences in pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 23(4), 579–591.

- Wiggins, L.D., Robins, D.L. (2008). Brief Report: Excluding the *ADI-R* behavioral domain improves diagnostic agreement in toddlers. *Journal of Autism and Developmental Disorders*, 38, 972-976.
- Williams, J.G., Allison, C., Scott, F.J., Bolton, P.F., Baron-Cohen, S., & Matthews, F.E. et al. (2008). The Childhood Autism Spectrum Test (CAST): sex differences. *Journal of Autism and Developmental Disorders*, 38, 1731-9.
- Wing, L. (1981). Sex ratios in early childhood autism and related conditions. *Psychiatry Research* 5, 129–137.
- World Health Organization. (1990). International classification of diseases (10th rev.): Criteria for research. Geneva, Switzerland: Author.
- Zwaigenbaum, L., Thurm, A., Stone, W., Baranek, G., Bryson, S., Iverson, J., ... & Sigman, M. (2007). Studying the emergence of autism spectrum disorders in high-risk infants: methodological and practical issues. *Journal of autism and developmental disorders*, 37(3), 466-480.
- Zwaigenbaum, L., Bryson, S. E., Szatmari, P., Brian, J., Smith, I. M., Roberts, W., ... & Roncadin, C. (2012). Sex Differences in Children with Autism Spectrum Disorder Identified Within a High-Risk Infant Cohort. *Journal of Autism and Developmental Disorders*, 1-12.

APPENDICES

Appendix A

M-CHAT

Please fill out the following about how your child usually is. Please try to answer every question. If the behavior is rare (e.g., you've seen it once or twice), please answer as if the child does not do it.

- | | | |
|--------------------------------------------------------------------------------------------------------------------------|-----|----|
| 1. Does your child enjoy being swung, bounced on your knee, etc.? | Yes | No |
| 2. Does your child take an interest in other children? | Yes | No |
| 3. Does your child like climbing on things, such as up stairs? | Yes | No |
| 4. Does your child enjoy playing peek-a-boo/hide-and-seek? | Yes | No |
| 5. Does your child ever pretend, for example, to talk on the phone or take care of a doll or pretend other things? | Yes | No |
| 6. Does your child ever use his/her index finger to point, to ask for something? | Yes | No |
| 7. Does your child ever use his/her index finger to point, to indicate interest in something? | Yes | No |
| 8. Can your child play properly with small toys (e.g. cars or blocks) without just mouthing, fiddling, or dropping them? | Yes | No |
| 9. Does your child ever bring objects over to you (parent) to show you something? | Yes | No |
| 10. Does your child look you in the eye for more than a second or two? | Yes | No |
| 11. Does your child ever seem oversensitive to noise? (e.g., plugging ears) | Yes | No |
| 12. Does your child smile in response to your face or your smile? | Yes | No |
| 13. Does your child imitate you? (e.g., you make a face-will your child imitate it?) | Yes | No |
| 14. Does your child respond to his/her name when you call? | Yes | No |
| 15. If you point at a toy across the room, does your child look at it? | Yes | No |
| 16. Does your child walk? | Yes | No |
| 17. Does your child look at things you are looking at? | Yes | No |
| 18. Does your child make unusual finger movements near his/her face? | Yes | No |
| 19. Does your child try to attract your attention to his/her own activity? | Yes | No |
| 20. Have you ever wondered if your child is deaf? | Yes | No |
| 21. Does your child understand what people say? | Yes | No |
| 22. Does your child sometimes stare at nothing or wander with no purpose? | Yes | No |
| 23. Does your child look at your face to check your reaction when faced with something unfamiliar? | Yes | No |

Appendix B

Table 1

Methodological characteristics of studies of behavioral sex differences in ASD

Author(s)	Sample*	Age** Range (mean)	ASDs Included	Additional Inclusion (In)/ Exclusion (Ex)	Control for IQ	IQ Esti- mate	Direct Observation	Parent Re- port	Self Report
Lord et al., 1982	384 M 91 F	3-8 y (5.3)	Autism based on CARS	NA	Both with and without a co- variate	Variety (nonver- bal)	PEP, PPVT, CARS	VABS, Social Maturity Scale	NA
McLennan et al, 1993	24 M 24 F	6.2-36 y	Autism based on DSM-III	<u>In:</u> Nonverbal IQ of 60 or above <u>Ex:</u> Complicated medical condi- tions, severe sensory impair- ments	NA	Raven's Std. Prog. Matrices	NA	ADI-R	NA
Volkmar et al., 1993	346 M (214 A,132 D) 142 F (59 A, 83 D)	(~10 y)	Autism or PDD/atypi- cal PDD based on both DSM- III and DSM III-R	NA	Both with and without a co- variate	Variety (some only nonverbal)	ICD-10 symp- tom score	VABS, ABC, ICD-10 symptom score	NA
Pilowsky et al.,1998	18 M 18 F	20 m-34 y	"Autism"	NA	Matche d	Variety	CARS	ADI-R, CARS	NA
Holtman et al., 2007	23 M 23 F	5-20 y (11.75)	ASD based on ICD-10	NA	Matche d	Variety (verbal and non- verbal)	ADOS	ADI-R, CBCL	NA
Carter et al., 2007	86 M 22 F	18-33 m (28.1)	AD or PDD based on clinical impression	<u>In:</u> ADOS (at least ASD cutoffs), ADI-R (at least ASD cutoffs) <u>Ex:</u> Genetic dx, phys handi- capped, neurological dx	Covari- ate	MSEL (Vis- ual Recep- tion Scale)	ADOS, MSEL	ADI-R, VABS, ITSEA	NA

Hartley & Sikora, 2009	157 M 42 F	18-47 m (~35.7)	AD or PDD based on DSM-IV-TR	<u>In</u> : ADOS (at least ASD cutoffs on soc and soc+com)	Covariate	MSEL (Visual Reception Scale)	ADOS, MSEL	CBCL	NA
Bolte et al., 2011	48 M (25 A, 23 S) 56 F (21 A, 35 S)	(~14 y)	ASD based on ICD-10	<u>In</u> : IQ of 70 or above, ADOS (AD on all cutoffs for AD and Asp; at least ASD on all cutoffs for Asp), ADI-R (social cutoff plus one other cutoff) <u>Ex</u> : Lack of functional language, severe comorbid health problems	Covariate	Variety (nonverbal)	ADOS, an executive functioning battery	ADI-R	CBCL/ YABCL
Lai et al., 2011	33 M 29 F	18-45 y	AD or Asp based on DSM-IV or ICD-10	<u>In</u> : IQ of 70 or above <u>Ex</u> : Psychotic dxs, substance use, medical conditions associated with ASD, epilepsy, hyperkinetic dx, and Tourette's syndrome	Matched (some analyses including covariate)	WASI (Full scale)	ADOS, Eyes Test	ADI-R	AQ, EQ, SQ, BAI, BDI, OCI-R
Rivet & Matson, 2011	132 M (66 A, 66 NA) 132 F (66A, 66 NA)	17-36 m	AD or PDD based on clinical impression	<u>Ex</u> : Identified sex chromosome dx	Matched	BDI-2 DQ	NA	M-CHAT, BDI-2, BIS-CUIT	NA
Sipes et al., 2011	294 M 96 F	17-36 m	AD or PDD based on clinical impression	NA	Analyses split (high and low)	BDI-2 DQ	NA	M-CHAT, BDI-2, BIS-CUIT	NA
Park et al., 2012	164 M (91 A, 51 S, 26 T) 97 F (20 A, 51 S, 26 T)	4-15 y (8.49)	ASD based on DSM-IV-TR	<u>In</u> : IQ of 70 or above <u>Ex</u> : Neurological dxs, serious medical conditions, or known chromosomal abnormalities	Matched	Leiter (Performance scale)	NA	ADI-R, SCQ, ASDS, CBCL	NA
Zwaigenbaum et al, 2012	233 M (55 A, 115 S, 63 NA) 202 F (25 A, 114 S,	36-42 m	AD or PDD based on DSM-IV-TR	<u>In</u> : For A sample, must have a sibling with ASD; for low-risk sample, no 3 rd degree or less relatives with ASD <u>Ex</u> : Genetic and neurological	Covariate	MSEL (Early Learning Composite)	ADOS2, MSEL		

	63 NA)			disorders, birth prior to 36 weeks, birth weight less than 2,500 grams					
--	--------	--	--	------------------------------------------------------------------------	--	--	--	--	--

Note. M=Males; F=Females; A=ASD; NA=non-ASD; D=Developmental disabled; S=Unaffected siblings; T=Typically developing; y=years; m=months; AD=Autistic Disorder; PDD=Pervasive Developmental Disorder-Not Otherwise Specified; dx=diagnosis/diagnoses; ABC=Autism Behavioral Checklist; ADI-R=Autism Diagnostic Interview, Revised; ADOS=Autism and Diagnostic Observation Schedule; ASDS=Asperger Syndrome Diagnostic Scale; AQ=Autism Quotient; BAI=Beck Anxiety Inventory; BDI=Beck Depression Inventory; BDI-2 DQ=Battelle Developmental Inventory Developmental Quotient; BISCUIT=Baby and Infant Screen for Children with Autism; CARS=Childhood Autism Rating Scale; CBCL=Child Behavior Checklist; DSM=Diagnostic and Statistical Manual of Mental Disorders; EQ=Empathy Quotient; ICD-10=International Classification of Diseases, Tenth Edition; ITSEA=Infant Toddler Social Emotional Assessment; M-CHAT=Modified Checklist for Autism in Toddlers; MSEL= Mullen Scales of Early Learning; OCI-R=Obsessive Compulsive Inventory, Revised; PEP=Psychoeducational Profile; PPVT=Peabody Picture Vocabulary Test; SQ=Systemizing Quotient; VABS=Vineland Adaptive Behavior Scales; WASI=Wechsler Abbreviated Scales of Intelligence; Approximate (~) total sample means were computed for studies that included the mean age for the subgroups only.

Table 2

Subject characteristics of males and females with and without ASD in the current sample

Demographic	Total	Males (%)	Females (%)
Total N	356	250 (70)	106 (30)
Age (months)			
Mean	25.3	25.3	25.1
Range	16-43	16-39	16-43
SD	4.55	4.4	4.9
Diagnostic Status			
ASD	176	139 (79)	37 (21)
Autistic Disorder	74	61 (82)	13 (18)
PDD-NOS	102	78 (76)	24 (24)
Non-ASD	180	111 (62)	69 (38)
Language Disorder	42	21 (50)	21 (50)
Global Delay	85	58 (68)	27 (32)
Other	5	3 (60)	2 (40)
No Diagnosis	32	21 (65)	11 (35)
Typical	16	8 (50)	8 (50)
Race			
Caucasian	219	159 (72)	60 (27)
Black	79	46 (58)	33 (42)
Asian	15	11 (73)	4 (27)
Biracial	25	19 (76)	6 (24)
Other	4	4 (100)	0 (0)
Unknown	14	11 (79)	3 (21)
Maternal Education (years)			
Mean	14.6	14.5	14.6
Range	8-20	8-20	9-20
SD	2.64	2.7	2.4

Table 3

Mean scores across variables of interest pertaining to ASD symptoms, intellectual functioning and developmental skills for the entire sample, and split by diagnostic status and sex

Measures	All		ASD				Non-ASD			
	N	M(SD)	N	Males M(SD)	Females M(SD)	N	Males M(SD)	Females M(SD)		
MSEL										
Visual Reception IQ (NVIQ)	356	71.8(24.7)	139	63.8(21.3)	37	58.6(22.9)	111	79.4(25.0)	69	82.7(23.0)
RL AE	354	14.3(7.2)	137	11.2(5.8)	37	9.8(6.5)	111	17.4(6.8)	69	17.9(6.9)
EL AE	356	13.5(6.3)	139	11.9(5.4)	37	10.2(5.1)	111	15.2(6.4)	69	15.9(6.6)
FM AE	355	18.5(4.9)	138	17.6(4.5)	37	16.2(4.2)	111	19.3(4.8)	69	20.0(5.3)
ADOS2										
SA	354	9.4(6.1)	139	14.1(3.5)	36	15.0(3.7)	110	4.4(3.7)	69	4.6(3.6)
RRB	354	1.7(1.9)	139	2.8(1.9)	36	2.6(2.1)	110	0.6(0.9)	69	0.7(1.0)
SS	354	3.8(2.7)	139	5.9(1.9)	36	6.1(1.9)	110	1.6(1.0)	69	1.7(1.2)
CARS(2)	353	26.6(6.9)	138	31.9(4.9)	37	32.2(5.3)	110	21.4(3.4)	68	21.3(3.5)

Note. MSEL= Mullen Scales of Early Learning; RL= MSEL Receptive Language Scale; EL=MSEL Expressive Language Scale; FM=MSEL Fine Motor Scale; AE=Age equivalent; ADOS2=Autism Diagnostic Observation Schedule, Second Edition; SA=ADOS2 Social Affect Domain; RRB= ADOS2 Restricted and Repetitive Behavior Domain; SS=ADOS2 Severity Score; CARS(2)= Childhood Autism Rating Scale (Second Edition); NVIQ=(MSEL Visual Reception Age Equivalent/Chronological Age) x 100.

Table 4

Summarized effect sizes (partial eta-squared) of the effect of sex, effect of ASD status, and interaction for all ASD symptom analyses when IQ was not included as a covariate

Measure	Interaction	Sex	ASD Status
ADOS2 SA	$\rho\eta^2 = .002$	$\rho\eta^2 = .004$	$\rho\eta^2 = .605^*$
ADOS2 RRI	$\rho\eta^2 = .010$	$\rho\eta^2 = .003$	$\rho\eta^2 = .255^*$
ADOS2 Severity Score	$\rho\eta^2 = .000$	$\rho\eta^2 = .003$	$\rho\eta^2 = .595^*$
CARS Total Score	$\rho\eta^2 = .001$	$\rho\eta^2 = .000$	$\rho\eta^2 = .554^*$

Note. P = partial; * $p < .001$

Table 5

Results from analysis exploring the effect of sex and ASD status on ADOS2 Social Affect (SA) Domain algorithm scores with NVIQ included as a covariate in the entire sample, sample split by sex, and sample split by ASD status

Sample	Effect	F	df	p	η^2
Whole Sample	IQ	45.0	1,349	<.001	.114
	ASD Status	420.3	1,349	<.001	.546
	Sex	1.2	1,350	.267	.004
	Interaction	.043	1,350	.836	.000
Males	IQ	45.6	1,246	<.001	.153
	ASD Status	391.2	1,249	<.001	.614
Females	IQ	4.7	1,102	.03	.044
	ASD Status	135.3	1,102	<.001	.570
ASD	IQ	24.2	1,172	<.001	.123
	Sex	0.69	1,172	.408	.004
Non-ASD	IQ	21.1	1,176	<.001	.107
	Sex	.453	1,176	.502	.003

Table 6

Results from analysis exploring the effect of sex and ASD status on ADOS2 Repetitive Restricted Behaviors (RRB) Domain algorithm scores with NVIQ included as a covariate in the entire sample, sample split by sex, and sample split by ASD status

Sample	Effect	F	df	p	η^2
Whole Sample	IQ	23.3	1,348	<.001	.063
	ASD Status	77.6	1,348	<.001	.182
	Sex	.095	1,348	.759	.000
	Interaction	1.36	1,348	.244	.004
Males	IQ	14.1	1,245	<.001	.055
	ASD Status	87.0	1,245	<.001	.262
Females	IQ	9.5	1,101	.003	.086
	ASD Status	20.7	1,101	<.001	.166
ASD	IQ	17.2	1,171	<.001	.092
	Sex	0.9	1,171	.365	.005
Non-ASD	IQ	8.54	1,175	.004	.047
	Sex	.774	1,175	.380	.004

Table 7

Results from analysis exploring the effect of sex and ASD status on ADOS2 Severity Scores (SS) with NVIQ included as a covariate in the entire sample, sample split by sex, and sample split by ASD status

Sample	Effect	F	df	p	η^2
Whole Sample	IQ	51.8	1,349	<.001	.129
	ASD Status	402.0	1,349	<.001	.535
	Sex	.70	1,349	.403	.002
	Interaction	.111	1,349	.739	.002
Males	IQ	45.6	1,246	<.001	.156
	ASD Status	368.5	1,246	<.001	.600
Females	IQ	7.9	1,102	.006	.071
	ASD Status	142.2	1,102	<.001	.582
ASD	IQ	33.8	1,172	<.001	.164
	Sex	.007	1,172	.931	.000
Non-ASD	IQ	21.7	176	<.001	.110
	Sex	1.1	176	.301	.006

Table 8

Results from analysis exploring the effect of sex and ASD status on CARS(2) Total score with NVIQ included as a covariate in the entire sample, sample split by sex, and sample split by ASD status

Sample	Effect	F	df	p	η^2
Whole Sample	IQ	86.2	1,347	<.001	.199
	ASD Status	337.1	1,347	<.001	.493
	Sex	.004	1,347	.948	.000
	Interaction	.073	1,347	.788	.000
Males	IQ	60.1	1,244	<.001	.198
	ASD Status	310.9	1,244	<.001	.560
Females	IQ	25.6	1,244	<.001	.202
	ASD Status	106.9	1,244	<.001	.514
ASD	IQ	21.9	1,171	<.001	.114
	Sex	.004	1,171	.951	.000
Non-ASD	IQ	103.3	1,174	<.001	.373
	Sex	.145	1,174	.703	.001

Table 9

Results from chi-square analyses conducted to explore the IQ-based sex prevalence (hypothesis 2) in the entire sample, within the ASD sample only, and within the non-ASD sample only

	Cognitive Ability		Total N
	Low	High	
Whole Sample			
Males			
Count	119	131	250
% Within Sex	47.6	52.4	
% Within Cognitive Ability	73.5	67.5	
Females			
Count	43	63	106
% Within Sex	40.6	59.4	
% Within Cognitive Ability	26.5	32.5	
Total N	162	194	356
ASD Sample			
Males			
Count	87	52	139
% Within Sex	62.6	37.4	
% Within Cognitive Ability	79.1	78.8	
Females			
Count	23	14	37
% Within Sex	62.2	37.8	
% Within Cognitive Ability	20.9	32.5	
Total N	110	66	176
Non-ASD Sample			
Males			
Count	32	79	111
% Within Sex	28.8	71.2	
% Within Cognitive Ability	61.5	61.7	
Females			
Count	20	49	69
% Within Sex	29.0	71.0	
% Within Cognitive Ability	38.5	38.3	
Total N	162	194	180

Table 10

Summarized effect sizes (partial eta-squared) of the effect of sex, effect of ASD status, and interaction for IQ and developmental skills when IQ was not included as a covariate

Measure	Interaction	Sex	ASD Status
Mullen Visual Reception IQ	$\rho\eta^2 = .007$	$\rho\eta^2 = .000$	$\rho\eta^2 = .124^*$
Receptive Language	$\rho\eta^2 = .005$	$\rho\eta^2 = .001$	$\rho\eta^2 = .228^*$
Expressive Language	$\rho\eta^2 = .008$	$\rho\eta^2 = .001$	$\rho\eta^2 = .124^*$
Fine Motor	$\rho\eta^2 = .011^{**}$	$\rho\eta^2 = .001$	$\rho\eta^2 = .080^*$

Note. P = partial; * $p < .001$; ** $p < .05$

Table 11

Results from analysis exploring the effect of sex and ASD status on MSEL Receptive Language Age Equivalent (AE) scores with NVIQ included as a covariate in the entire sample, sample split by sex, and sample split by ASD status

Sample	Effect	F	df	p	η^2
Whole Sample	IQ	257.1	1,348	<.001	.425
	ASD Status	47.7	1,348	<.001	.121
	Sex	.108	1,348	.734	.000
	Interaction	47.7	1,348	<.001	.000
Males	IQ	152.1	1,244	<.001	.384
	ASD Status	43.1	1,244	<.001	.150
Females	IQ	112.4	1,102	<.001	.524
	ASD Status	12.4	1,102	.001	.108
ASD	IQ	63.6	1,170	<.001	.272
	Sex	.348	1,170	.556	.002
Non-ASD	IQ	225.7	1,176	<.001	.562
	Sex	.027	1,176	.869	.000

Table 12

Results from analysis exploring the effect of sex and ASD status on MSEL Expressive Language Age Equivalent (AE) scores with NVIQ included as a covariate in the entire sample, sample split by sex, and sample split by ASD status

Sample	Effect	F	df	p	η^2
Whole Sample	IQ	173.2	1,350	<.001	.33
	ASD Status	11.8	1,360	.001	.033
	Sex	.306	1,350	.580	.001
	Interaction	1.0	1,350	.311	.003
Males	IQ	156.5	1,246	<.001	.389
	ASD Status	5.5	1,246	.019	.022
Females	IQ	29.3	1,102	<.001	.223
	ASD Status	6.5	1,102	.012	.060
ASD	IQ	50.6	1,172	<.001	.227
	Sex	1.5	1,172	.223	.009
Non-ASD	IQ	126.1	1,176	<.001	.417
	Sex	.072	1,176	.789	.000

Table 13

Results from analysis exploring the effect of sex and ASD status on MSEL Fine Motor Age Equivalent (AE) scores with NVIQ included as a covariate in the entire sample

Sample	Effect	F	df	p	η^2
Whole Sample	IQ	316.4	1,349	<.001	.476
	ASD Status	.776	1,349	.379	.002
	Sex	.23	1,349	.632	.001
	Interaction	1.5	1,349	.218	.004

Appendix C

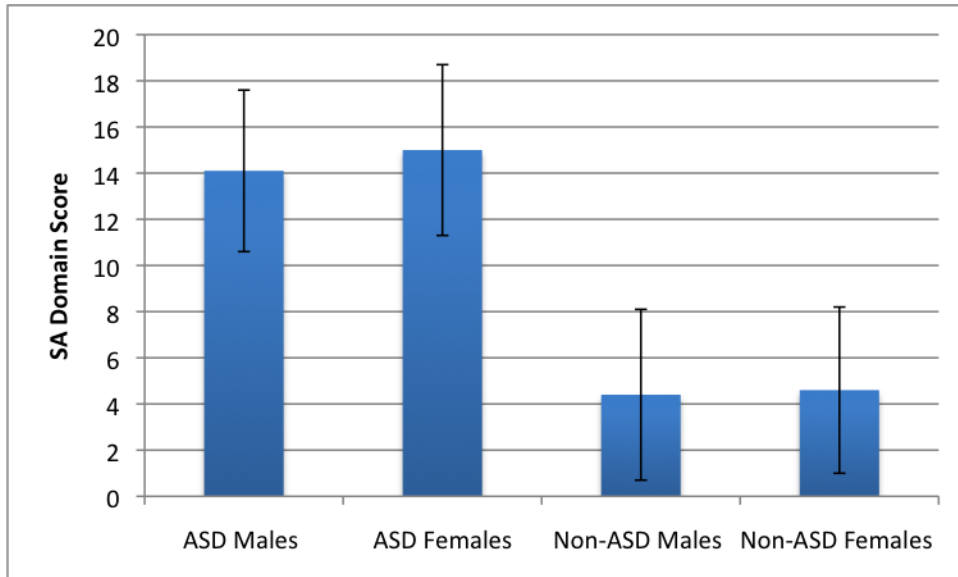


Figure 1. ADOS2 Social Affect (SA) Domain score means across sex and diagnostic groups. Error bars represent standard errors.

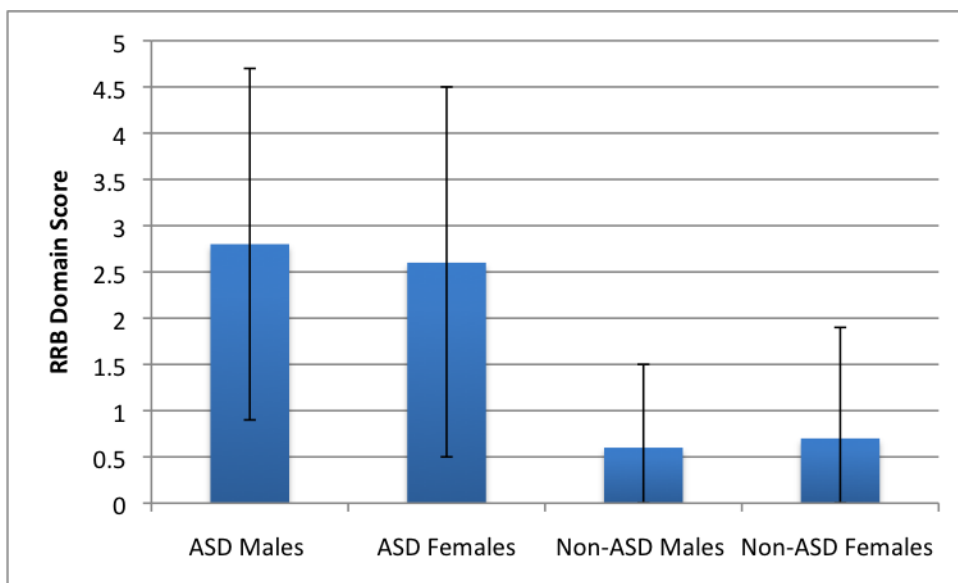


Figure 2. ADOS2 Repetitive and Restricted Behaviors (RRB) Domain score means across sex and diagnostic groups. Error bars represent standard errors.

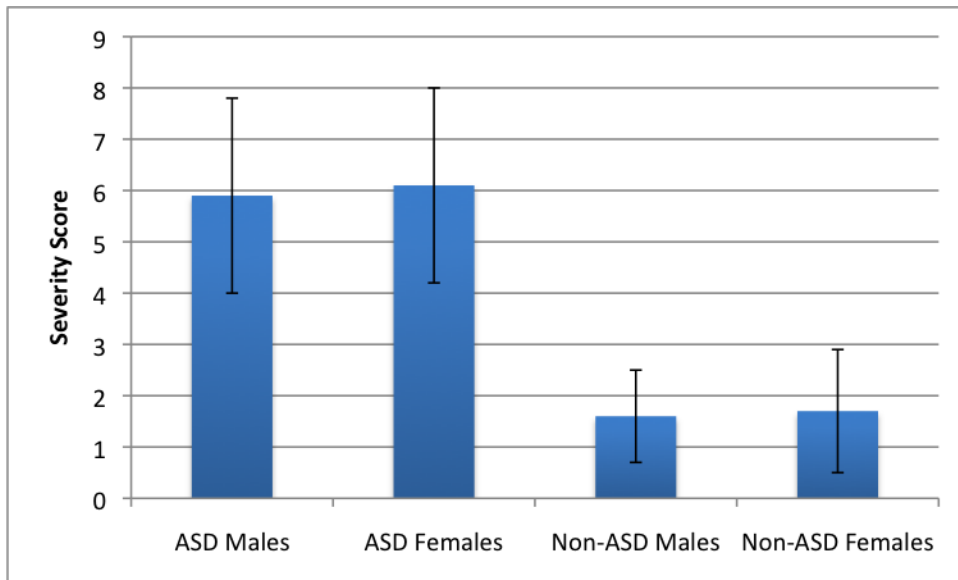


Figure 3. ADOS2 Comparison Score (Severity Score) means across sex and diagnostic groups. Error bars represent standard errors.

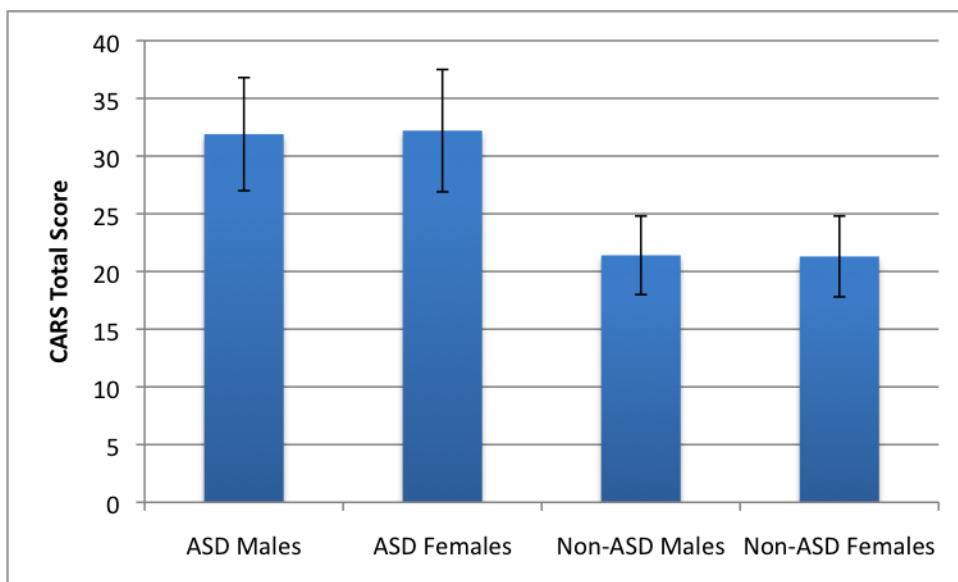


Figure 4. Childhood Autism Rating Scale (CARS) Total score means across sex and diagnostic groups. Error bars represent standard errors.

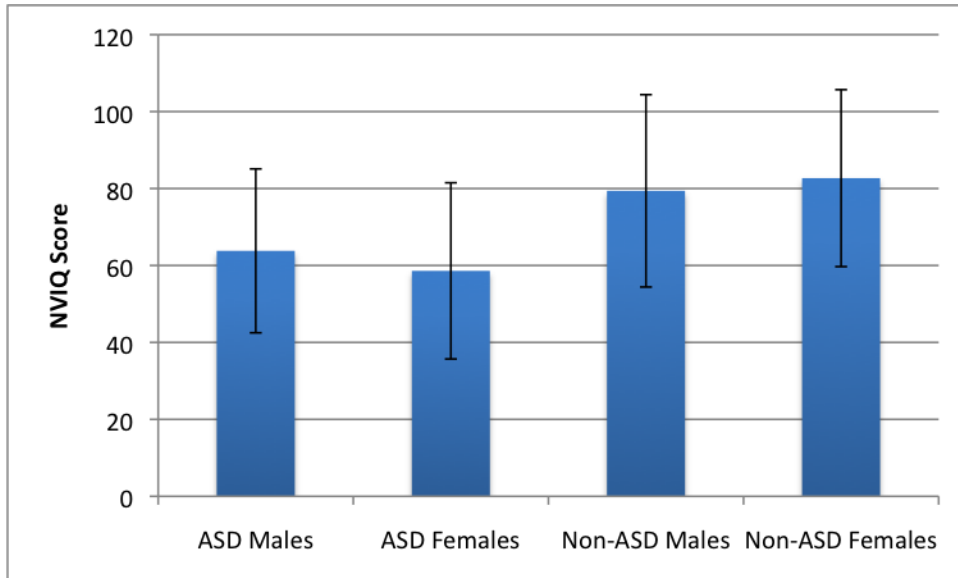


Figure 5. Nonverbal IQ (NVIQ) score means across sex and diagnostic groups. The NVIQ is computed from the Mullen Scales of Early Learning (MSEL) Visual Reception Scale age-equivalent score. Error bars represent standard errors.

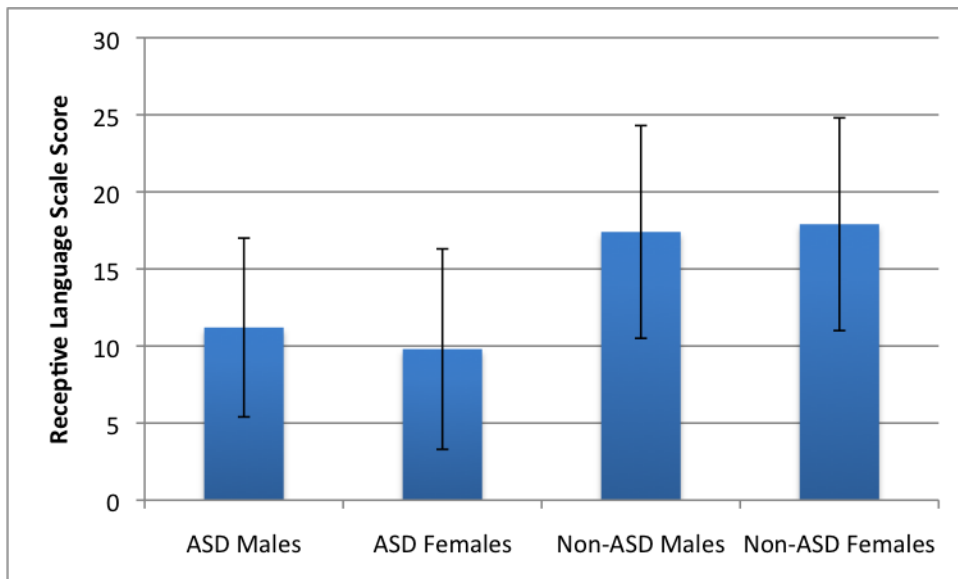


Figure 6. Mullen Scales of Early Learning (MSEL) Receptive Language Scale age-equivalent score means across sex and diagnostic groups. Error bars represent standard errors.

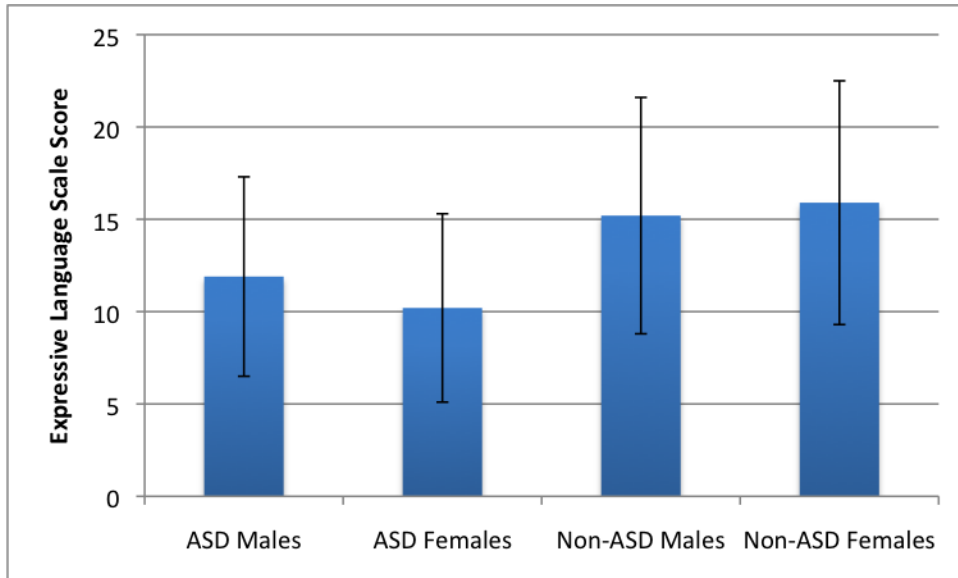


Figure 7. Mullen Scales of Early Learning (MSEL) Expressive Language Scale age-equivalent score means across sex and diagnostic groups. Error bars represent standard errors.

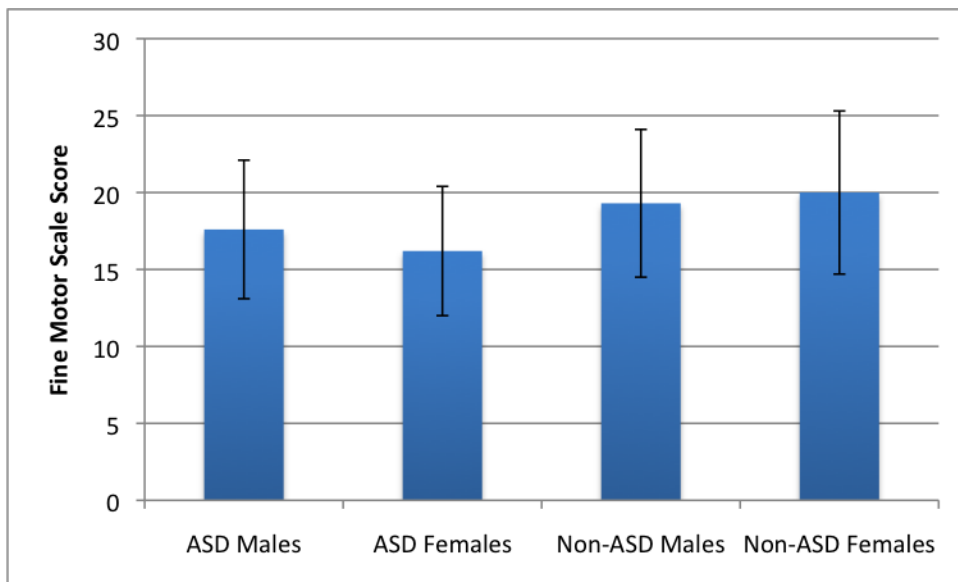


Figure 8. Mullen Scales of Early Learning (MSEL) Fine Motor Scale age-equivalent score means across sex and diagnostic groups. Error bars represent standard errors.