EXAMINING THE CROSS-CULTURAL VALIDITY OF THREE EARLY AUTISM SCREENING INSTRUMENTS

by

JAIMI DONALD MCDONOUGH

(Under the Direction of Jonathan Campbell and Amy Reschly)

ABSTRACT

Despite evidence to support improved outcomes with early intervention, Autism Spectrum Disorder (ASD) is still being diagnosed, on average, much later than the time at which it can be reliably diagnosed. Research suggests that differences exist between minority racial and SES groups with regard to age at first diagnosis of ASD. Population-wide screening for ASD has been recommended to aid in early identification as well as to reduce disparities in timing of diagnosis, particularly for underserved groups. Although research related to the psychometric properties of ASD assessment instruments is established, little information is available regarding use of ASD assessment instruments with culturally diverse populations. The purpose of the present study was to examine the validity of three early ASD screening instruments across cultural groups to inform clinical use of the instruments with diverse populations. Parent ratings from the *Social Communication Questionnaire, Current Version* (SCQ-Current), the *Modified Checklist for Autism in Toddler* (M-CHAT), and the *Pervasive Developmental Disorders Screening Test-II* (PDDST-II) were collected for 121 children (*M* age = 3.74 years, *SD* = 1.14 years) at risk for ASD who were participating in a comprehensive diagnostic evaluation. No significant differences in age at evaluation were found across race or

maternal education level. In addition, no significant differences in screener accuracy were observed for minority race groups or for raters without a high school diploma. In contrast, the M-CHAT and M-CHAT Critical Item Total Score differentiated ASD and non-ASD participants more effectively than the SCQ in the total sample. None of the screeners demonstrated acceptable diagnostic accuracy for both sensitivity and specificity within a referred sample. Findings and implications for clinical practice are discussed.

INDEX WORDS: Autism Spectrum Disorder, ASD, Assessment, Cross-Cultural, Screening

EXAMINING THE CROSS-CULTURAL VALIDITY OF THREE EARLY AUTISM SCREENING INSTRUMENTS

by

JAIMI DONALD MCDONOUGH

B.A., Clemson University, 2008

M.A., University of Georgia, 2012

A Dissertation Draft Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2015

EXAMINING THE CROSS-CULTURAL VALIDITY OF THREE EARLY AUTISM SCREENING INSTRUMENTS

by

JAIMI DONALD MCDONOUGH

Major Professors: Jonathan Campbell
Amy Reschly

Committee: Scott Ardoin
Stacey Neuharth-Pritchett

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2015

ACKNOWLEDGMENTS

I would like to extend my gratitude to my family and friends who have provided me tremendous support. Specifically, I would like to thank my husband, Zach McDonough, for embarking on this graduate school journey with me and providing constant encouragement. I would like to thank my parents, Sam and Joelyn Donald, for instilling in me a love for learning. I would also like to thank my major professor, Dr. Jonathan Campbell, for his support and academic guidance throughout this process and Drs. Amy Reschly, Scott Ardoin, and Stacey Neuharth-Pritchett for their valuable feedback. Additionally, I would like to express my appreciation to the staff and patients of the University of Tennessee Health Science Center Boling Center For Developmental Disabilities. Specifically, I would like to thank Drs. Laura Murphy and Lauren Benner for the clinical and research support, without which this study would not have been possible. I would also like to thank the Organization for Autism Research for providing grant funding for data collection through the BCDD. Finally, I would like to thank my school psychology cohort for their friendship and guidance.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

According to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision* (DSM-IV-TR) (2000), pervasive developmental disorders (PDDs) are a group of disorders characterized by deficits in communication and social behavior as well as the presence of restricted and repetitive patterns of behaviors and/or interests. As defined in the DSM-IV TR, PDDs include Autistic Disorder, Asperger's Disorder (AspD), Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS), Rett's disorder and childhood disintegrative disorder (CDD) (American Psychiatric Association [APA], 2000). Each of the five PDDs share the aforementioned triad of qualitative impairments but differ in terms of the severity and intensity of symptoms and the developmental course of the disorder (APA, 2000; Volkmar, Lord, Bailey, Schultz, & Klin, 2004). Given the heterogeneity observed among PDDs, researchers have begun to conceptualize this group of disorders as occurring on a spectrum rather than discrete diagnoses.

Within the most recent edition of *Diagnostic and Statistical Manual of Mental Disorders* (i.e., DSM-5), individual diagnoses are subsumed under the umbrella diagnosis of Autism Spectrum Disorder (ASD), with specifiers to indicate symptom severity (APA, 2013). Several studies have investigated how the changes in diagnostic criteria between the DSM-IV-TR and DSM-5 affect ASD prevalence rates. For example, McPartland, Reichow and Volkmar (2012) found that use of DSM-5 criteria improves specificity in ASD diagnosis; however, the authors reported low sensitivity for AspD and PDD-NOS as well as for those participants with cognitive

scores greater than 70. Young and Rodi (2014) found similar results when applying DSM-IV-TR and DSM-5 criteria concurrently during diagnostic assessment. Kulage, Smaldone, and Cohn (2014) conducted a meta-analysis to compare published studies examining changes in ASD diagnosis when employing both the DSM-IV TR and DSM-5 criteria in the same sample. Authors examined 14 studies and found that sensitivity in ASD diagnosis is reduced when utilizing DSM-5 criteria. Specifically, authors concluded that a decreased number of individuals were diagnosed with ASD with the utilization of DSM-5 ASD criteria as opposed to DSM-IV-TR criteria (Kulage et al., 2014). Kulage et al. noted that diagnostic differences between DSM-5 and DSM-IV-TR are more likely to be observed for individuals meeting DSM-IV-TR diagnostic criteria for PDD-NOS when compared to Autistic Disorder or AspD. Authors also reported reduced DSM-5 ASD diagnoses for those individuals meeting DSM-IV-TR diagnostic criteria for AspD; however, the difference was not significant. Likewise, only four studies included in the meta-analysis examined the impact of DSM-5 diagnostic criteria on AspD. Results revealed significant variation by age in the reduction of individuals identified with ASD between DSM-IV-TR and DSM-5 criteria, with the greatest reduction of diagnoses observed for children. Kulage et al. cautioned that, due to the small sample of included studies examining children, additional research is necessary to determine if using DSM-5 criteria disproportionately reduces ASD diagnosis in children when compared to DSM-IV-TR criteria.

The most recent report from the Centers for Disease Control's Autism and Developmental Disabilities Monitoring (ADDM) Network indicates that 1 in 68 children in the United States has been diagnosed with an ASD by age 8, according to data collected from 11 nationwide sites in 2010 (Centers for Disease Control [CDC], 2014). The current estimate suggests a 123% increase in ASD prevalence during the period 2002 to 2010 (CDC, 2014). Of

note, the most recent CDC prevalence estimates are based on DSM-IV-TR diagnostic criteria (APA, 2000). Although it is unclear whether the increase in prevalence rates is due to an actual increase in ASD, improved assessment methods, or increased awareness of ASD symptoms, the importance of early identification of ASD has received significant research attention.

Early identification of young children with ASD has been shown to foster improvement in social, communicative, and cognitive skills, as early intervention is often contingent upon identifying the disorder in early childhood (Osterling, Dawson, & Munson, 2002). Dawson (2008) described a "sensitive period" for social development in individuals with autism during which skills can be taught. The author also posited that early intervention can actually modify the developmental trajectory of young children with ASD in favor of more typical brain and behavior development. Dawson noted that deficits in facial processing can be improved or corrected during the sensitive period, adding that the sensitive period may extend to other areas of dysfunction associated with autism (Dawson, 2008).

Similarly, children with ASD who are identified at a young age and enrolled in early intervention programming demonstrate improved long-term outcomes when compared to children with ASD who are identified later (Johnson & Meyers, 2007; Rogers & Vismara, 2008). Children typically cannot access ASD-specific early intervention services without a diagnosis. Likewise, a formal ASD diagnosis increases the likelihood of identification by ASD surveillance systems, which provide information regarding ASD prevalence. Accuracy in prevalence rates of ASD assists with policy decisions as well as educational and treatment planning (Wiggins, Baio, & Rice, 2006). As such, early diagnosis is critical for improved prognosis and quality of life for individuals with ASD.

One of the earliest studies to investigate outcomes for children with autism found that age of entry into an intervention program was strongly related to positive treatment outcomes (Fenske, Zalenski, Krantz, & McClannahan, 1985). Authors followed nine children diagnosed with autism and enrolled in an intensive day treatment program. Authors cited improved life outcomes and long-term prognoses for those participants who received early intervention (i.e., entering a treatment program prior to 60 months of age), including maintaining residence with family of origin and public school attendance. Likewise, authors emphasized the societal advantages of early identification and intervention with individuals diagnosed with autism (Fenske et al., 1985).

Jacobson, Mulick, and Green (1998) conducted a cost-benefit analysis of early, intensive behavioral intervention for children with autism based on a range of treatment outcomes reported in the literature. Authors assumed a range of effects including normal functioning, partial effects, and minimal effects and assumed participation in treatment for three years between the age of 2 years and school entry. After subtracting the estimated cost of early intensive behavioral intervention and depending on rate of effectiveness, the model estimate suggested that savings ranged from $187,000 to $203,000 per child between ages 3 - 22 years and between $656,000 to nearly $1.1 million per child for ages 3 – 55 years. As such, this model would suggest that the upfront cost of early intervention greatly outweighs the societal cost of providing long-term supports, even when accounting for varied treatment responses (Jacobson et al., 1998).

Even in light of evidence to suggest that early identification and intervention are associated with improved outcomes, the most recent published CDC surveillance study indicated that the median age of first known diagnosis of ASD is 53 months (CDC, 2014). Another study in the metropolitan Atlanta area found that the mean age of first ASD evaluation was 48 months,

while the mean age of first ASD diagnosis was 61 months, suggesting that, on average, there is a 13-month delay between initial ASD evaluation and first ASD diagnosis (Wiggins, Baio, & Rice, 2006). The CDC estimates that the median age of first ASD diagnosis has not changed significantly over the period from 2000 to 2010, despite an increase in research, community outreach, and awareness of ASD symptoms (CDC). Research suggests that ASD can be diagnosed reliably at 24 months (Chawarska, Klin, Paul, & Volkmar, 2009). Furthermore, when an ASD diagnosis is rendered by a skillful clinician at 18 to 24 months, the stability of the diagnosis is 80-90% (Chawarska et al., 2009).

In most cases, parents along with primary care providers are the first to recognize developmental delays. A recent study examining professional knowledge of ASD, particularly with respect to DSM-5 diagnostic criteria, in Western New York included professionals such as hospital and community medical providers, related service providers, and educators (Hartley-McAndrew, Doody, & Mertz, 2014). Authors reported that professionals demonstrated accurate knowledge of ASD prevalence; however, knowledge gaps were observed with respect to professionals' knowledge of specific diagnostic criteria. The majority of practitioners reported that they would benefit from additional training in the area of ASD (Hartley-McAndrew et al., 2014). A study including parents of children referred for an autism evaluation found that the most common first concerns of parents included speech and language delays, atypical social-emotional response, and medical problems (De Giacomo & Fombonne, 1998). A more recent study examined maternal symptom reports across racial-ethnic groups. Authors reported that Anglo-American mothers reported significantly more developmental concerns and ASD symptoms when compared with Latino-American mothers, despite the finding of more severe symptomatology in the Latino-American group (Blacher, Cohen, & Azad, 2014).

Despite improvements in early detection of ASD, growing evidence suggests that differences exist between racial and SES groups with regard to age at first diagnosis of ASD. Children from ethnic minority or low SES families, as well as those from families residing in rural areas are, on average, diagnosed later than their majority group counterparts; however, according to three large prevalence studies, there is no known difference in the epidemiology of ASD by race, ethnicity, or SES (Bertrand et al., 2001; Fombonne, 2003; Yeargin-Allsopp et al., 2003). Mandell, Novak, and Zubritsky (2005) found that near-poor children were diagnosed with an ASD, on average, 11 months later than children from families with incomes greater than 100% above the poverty line. Likewise, children in suburban and rural areas typically receive a diagnosis later than children residing in metropolitan areas (Chuan-Yu, Liu, Su, Huang, & Lin, 2008; Mandell et al., 2005).

Mandell, Listerud, Levy, and Pinto-Martin (2002) examined cases of children with autism receiving Medicaid services in Philadelphia (N = 406) and found that Caucasian children, on average, received an ASD diagnosis over one year earlier than African-American children and over two years earlier than Latino children. Likewise, authors noted that African-American children required three times as many visits to be diagnosed with ASD (Mandell et al., 2002). The most recent CDC surveillance study reported a significantly higher prevalence of ASD diagnoses among non-Hispanic Caucasian children than that for Hispanic or African-American children (CDC, 2014). In contrast, Mandell et al. (2005) and Shattuck et al. (2009) found no significant differences in the age at first ASD diagnosis based on race group membership.

The findings of no correlation between ethnicity and later ASD identification may provide evidence that disparities are improving. On the other hand, these findings may also suggest that ethnicity and SES are correlated and that SES has a stronger association with age at ASD diagnosis than ethnicity (Mandell et al., 2005).

Several possibilities exist to explain why low SES and minority ethnic group status appear to be related to later ASD identification and diagnosis. First, differences in symptom interpretation may exist among caregivers. Second, traditionally underserved populations may have limited access to specialty care professionals who diagnose ASD. Third, clinicians may be biased in their identification of ASD among children from underserved populations.

As evidence accumulates to support a disparity in timing of ASD identification among traditionally underserved populations, the importance of early identification of ASD has become more apparent. Using structured ratings such as standardized screening instruments may decrease ethnic or SES bias in ASD diagnosis. Correspondingly, routine screening for ASD at the population level has been shown to improve early recognition in traditionally underserved groups, such as those from racial/ethnic minority and low SES groups (Liptak et al., 2008).

Desirable ASD screening instruments are those with the strongest psychometric properties, particularly with respect to standardization, reliability and validity. Several early autism screeners have been developed for detection of ASD in young children. Screening instruments can be evaluated by sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). Sensitivity is the most salient concern for early autism screeners; however, an adequate balance between sensitivity and specificity is ideal (Barton, Dumont-Mathieu, & Fein, 2012).

A major limitation of the ASD screening literature involves limited information related to potential differences across racial and SES groups with respect to the psychometric properties of the screeners. In their discussion of evidence-based assessment, Hunsley and Mash (2007) emphasized the importance of developing assessment instruments and methods that take into consideration the diverse populations on which they will be used as part of an evidence-based approach to assessment.

Although validation studies have been conducted on most ASD screening instruments, evidence for cross-cultural validity is rarely reported. For example, in one of the initial validation studies of the *Modified Checklist for Autism in Toddlers* (M-CHAT), Kleinman et al. (2008) indicated that racial and ethnic information was not collected from participants. The authors reported that there were children from minority and low socioeconomic groups in the sample, but specific demographics were not discussed. These limitations were noted in the published report of the M-CHAT validation study, but the authors still concluded that the M-CHAT is appropriate for use in the general population (Kleinman et al., 2008). To date, there are no published studies comparing the predictive validity of early autism screening instruments across various race or SES groups.

Purpose of the Study

The purpose of the present study is to examine the cross-cultural validity of three early ASD screening instruments for preschool-aged children to inform the validity of use of the instruments with diverse populations. First, I reviewed the literature regarding understanding ASD within a cultural context, ASD assessment practices and diagnosis, early identification of ASD, and the importance of developing assessment instruments which consider the diverse populations for which they will be used.

I argued that understanding ASD within a cultural context is necessary for clinicians working with diverse cultural groups. Second, I argued that evidence of cross-cultural validity of measures purported to detect ASD in population samples is essential for valid use of such instruments. Finally, I tested the validity of three screeners within a diverse sample. The proposed study has the potential to guide selection of ASD screening instruments for preschool-age children of varied racial and SES backgrounds.

CHAPTER 2

REVIEW OF THE LITERATURE

Autism Spectrum Disorders Within a Cultural Context

Cultural influences shape how individuals perceive and experience disorders (Mandell & Novak, 2005). The role of sociocultural variables in autism prevalence has been discussed since Kanner's (1943) account of early infantile autism. Historically, autism was thought to occur predominantly in children of parents with high levels of education and socioeconomic status (SES) (Kanner, 1943). Despite the early contention that high SES is linked to autism, surprisingly little is known about ASD within a cultural context (Daley, 2002). The relative lack of cross-cultural research on ASD could be explained by the supposed neurobiological etiology of the disorder which makes environmental or non-biological aspects of the disorder less salient (Cuccaro, Wright, Rownd, & Abramson, 1996). According to three large prevalence studies, there is no known difference in the epidemiology of ASD by race, ethnicity, or SES (Bertrand et al., 2001; Fombonne, 2003; Yeargin-Allsopp et al., 2003). However, the presentation of ASD symptomatology may be susceptible to cultural influences (Daley, 2002).

Cross-cultural Clinical Presentation of ASD

Despite nearly equal rates of ASD internationally, the degree to which clinical presentation of ASD varies across cultures is largely unknown (Daley, 2002). There is a possibility that cultural differences affect the presentation of ASD due to either environmental or genetic factors (Mandell & Novak, 2005). For example, given the heterogeneous symptom presentation of ASD, certain behavioral or symptom presentations may be associated with a

particular cultural background; however, at the present time, no published studies have examined the possibility of cross-cultural differences in ASD presentation (Mandell & Novak, 2005). Chung et al. (2012) investigated differences in challenging behaviors for children diagnosed with ASD across four countries using the *Autism Spectrum Disorders-Diagnostic for Children* (ASD-DC). Authors found consistent behavioral topographies and intensities for children in the United States (n = 156), Israel (n = 48) and South Korea (n = 54) based on parent ratings on the *Autism Spectrum Disorders – Problem Behavior Checklist for Children (*ASD-PBC) (Chung et al., 2012). In contrast, significant differences with respect to presence and severity of behaviors were found between children from the United States and the United Kingdom (n = 27). When differences were observed between behavioral presentations, parent ratings for children from the United Kingdom consistently showed higher endorsements in presence and severity of challenging behaviors (Chung et al., 2012). Authors cautioned that observed differences may be the result of varied sampling procedures, as participants in the US sample were drawn from various sources while those in the UK sample were drawn exclusively from schools. However, previous research has indicated that children in the UK present with more severe parent-reported symptoms of ASD when compared to children in the US, Israel, and South Korea; therefore, true differences in the presence and severity of challenging behaviors may exist between the US and UK (Matson et al., 2011; Chung et al., 2012). Results suggest the possibility of different interpretations or conceptualizations of challenging behaviors across countries; however, results should be interpreted with caution due to small sample size and unequal number of participants across groups.

Similarly, Matson et al. (2012) compared cross-cultural differences between parent-reported social skills in children diagnosed with Autistic Disorder in the United States (n = 156) and South Korea (n = 54) using the *Matson Evaluation of Social Skills with Youngsters, Second Edition* (MESSY-II). Authors found that children from the United States were rated by parents as having significantly more inappropriate social skills as well as more adaptive social skills compared to parent ratings of children from South Korea. However, mean social skills ratings fell in the same category of impairment across participants, suggesting that similar social skills impairments exist across cultures (Matson et al., 2012).

Cuccaro et al. (2007) found that African-American children with ASD were more likely to experience language delays when compared to Caucasian children. Likewise, based on clinical practice, one group of researchers reported that African-American children diagnosed with ASD have fewer deficits in nonverbal communicative behaviors, including eye contact, than Caucasian children with ASD (Sell, Giarelli, Blum, Hanlon, & Levy, 2012). According to Cuccaro et al. (2007), no differences in social and repetitive behaviors were noted between these groups; however, in a recent population-based study, Sell et al. (2012) found that Caucasian children with ASD demonstrated more symptoms of DSM-IV-TR Restricted Interests and Repetitive/Stereotyped Behaviors than African-American children with ASD. Similarly, Sell et al. found that Caucasian children with ASD had higher rates of delayed motor development and odd responses to sensory stimuli than African-American children with ASD. Authors of a study of 8-year-old children diagnosed with ASD found that aggression and defiance were observed more often in African-American children with ASD than Caucasian children with ASD (Giarelli et al., 2010). In contrast, Sell et al. found no difference in symptoms of hyperactivity, aggression, or oppositional behavior across racial groups.

Autism Spectrum Disorders and Socioeconomic Status

Leo Kanner first described children diagnosed with autism as coming from affluent, European-American backgrounds (Kanner, 1943). Kanner's research began a trend of including European-American participants nearly exclusively in studies of autism (Kanner, 1943). Although the historical clustering of ASD among high SES families was likely due to a social class bias in access to specialty psychological and medical services, a recent study demonstrated that disparities still exist in ASD diagnosis and treatment among underserved populations (Cuccaro et al., 1996).

Research suggests that ASD may be underdiagnosed in children from low SES families. Specifically, a Centers for Disease Control and Prevention study found that higher family income and maternal education were associated with ASD when an autism diagnosis was not comorbid with intellectual disability. The authors suggest that this disparity is likely due to more access to specialty services among high SES families (Bhasin & Schendel, 2007). Another study found that school-aged children from poor and non-poor families experience similar overall rates of ASD; however, significant differences exist when comparing rates of ASD among preschool-aged children, indicating a discrepancy across SES groups in early diagnosis of ASD (Liptak et al., 2008). Likewise, authors of an international survey study of parents raising a child diagnosed with ASD found that higher income and parent education levels were associated with a younger age at first ASD diagnosis (Goin-Kochel, Mackintosh, & Myers, 2006). A recent cross-sectional study utilizing nationwide surveillance data from the ADDM Network found that prevalence of ASD increased with SES in a dose-response fashion. Furthermore, a stronger SES gradient in ASD prevalence was observed for children with a documented ASD diagnosis versus those without a pre-existing diagnosis (Durkin et al., 2010).

Through survey data collected in Pennsylvania, Mandell and colleagues (2005) found that near-poor children, on average, received an ASD diagnosis 11 months later than children from families with incomes greater than 100% above the poverty line. Mandell and Daniels (2013) found in their study of Medicaid-eligible preschool-aged children with Autistic Disorder that children living in counties with higher median household incomes were diagnosed significantly earlier. Additionally, a telephone survey in Taiwan found that the lowest parent-reported rate of ASD for children younger than age 6 was among poor children (Chuan-Yu et al., 2008). Further evidence for a disparity in ASD identification among children from low SES families comes from a study in which healthcare professionals made clinical judgments about children through vignettes with symptom information suggestive of ASD. The vignettes sent to professionals only differed with respect to ethnicity of the child and SES of the family. Results of the study suggest that higher SES is associated with a greater likelihood of a professional judgment of ASD (Cuccaro et al., 1996).

Autism Spectrum Disorders and Race/Ethnicity

Travers, Tincani, and Krezmien (2011) pointed out that the majority of research on disproportionate representation of ethnic minorities in disability categories has focused on overrepresentation. In contrast, disparities in ASD identification have centered on underrepresentation of individuals in minority race groups within this category. Ethnic differences in ASD identification may be due to true group differences in prevalence or diagnostic disparities (Mandell et al., 2009). In general, African Americans are underrepresented in published autism research (Hilton et al., 2010).

Numerous studies have demonstrated differences in the age at first ASD diagnosis between majority and minority ethnic groups. For example, Mandell et al. (2002) reported that,

on average, African-American children are 2.6 times less likely to receive an ASD diagnosis on their first visit to a specialty care clinician. In Mandell et al.'s study, 72% of the Caucasian sample versus 57% of the African-American sample received ASD diagnoses on the first specialty care visit. Additionally, Caucasian children, on average, received their first ASD diagnosis at 6.3 years, while African-American and Latino children received their first ASD diagnosis at 7.9 and 8.8 years, respectively (Mandell et al., 2002). A recent study examining Medicaid-eligible preschool-aged children found that Hispanic children were diagnosed significantly later than other racial and ethnic groups (Daniels & Mandell, 2013). Similarly, a separate study found differences in parent-reported ASD severity, with greater severity for African-American and Latino children reported. Such findings may suggest that autism has been underdiagnosed in children from racial minority groups who experience less severe symptoms (Liptak et al., 2008).

The CDC reports that ASDs are almost five times more prevalent in males than females, with 1 in 42 males diagnosed with ASD compared to 1 in 189 females (CDC, 2014). Additionally, the CDC report indicates ASD prevalence differences across racial groups. Hispanic children are diagnosed with ASD at a rate of 10.8 in 1000 children. For non-Hispanic African-American children, the ASD prevalence is 12.3 in 1000. Non- Hispanic Caucasian children are diagnosed with ASD at a rate of 15.8 in 1000 children, suggesting that Non-Hispanic Caucasian children were approximately 30% more likely to be identified with ASD than non-Hispanic African-American children and almost 50% more likely to be identified with ASD than Hispanic children (CDC, 2014).

One correlate of delayed identification is that African-American children are more likely than Caucasian children to be identified with an ASD by the public school system as opposed to being diagnosed prior to beginning school (Bhasin & Schendel, 2007; Yeargin-Allsop et al., 2003). Furthermore, another study found that African-American children are three times more likely to receive another diagnosis before ultimately receiving an ASD diagnosis (Mandell, Ittenback, Levy, & Pinto-Martin, 2007). Specifically, Attention Deficit Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder (ODD) are commonly diagnosed before ASD in African-American children, even though the DSM-IV-TR requires that ASD be ruled out before these disorders can be diagnosed (APA, 2000; Mandell et al., 2007). It is unclear whether the new diagnostic criteria in the fifth edition of the DSM will remedy this issue. Possible explanations for African-American children receiving another diagnosis before ASD are differences in parental descriptions of symptoms, differences in clinician interpretations or expectations, or differences in symptom presentation (Mandell et al., 2007).

Many published studies suggest racial disparities in ASD diagnosis; however, evidence also exists suggesting no difference in the age at first diagnosis of ASD based on race. For example, Mandell et al. (2005) found no difference among ethnic groups in the age at first diagnosis of ASD. However, later ASD identification was significantly correlated with low family income and rural residence (Mandell et al., 2005). Likewise, a population-based CDC autism surveillance study found no significant differences in the age at first ASD diagnosis among different racial groups. However, in the same study, African-American and Hispanic children were less likely than Caucasian children to have a documented ASD diagnosis in their educational or medical records (Shattuck et al., 2009). Goin-Kochel et al. (2006) found no racial differences in age of ASD diagnosis in an international survey study of families with a child

diagnosed with ASD. Sell et al. (2012) found no difference in age of diagnosis among Caucasian and African-American children with ASD in a population sample. In contrast to many other publications, a large (N = 28,722) study of Medicaid-eligible children newly diagnosed with ASD between 2001 and 2004 found that children from ethnic minority groups were diagnosed with Autistic Disorder at an earlier age than children from majority racial groups (Mandell et al., 2010). Authors reported that, on average, Asian-American children were diagnosed at 60.9 months, while Latino, African-American, and Caucasian children were diagnosed at 61.6, 62.0, and 63.0 months, respectively. Results of the study suggest that, in Medicaid-eligible children, age at first diagnosis of ASD is earlier for minority groups (Mandell et al., 2010). The findings of no correlation between ethnicity and later ASD identification may provide evidence that the disparity is improving. On the other hand, these findings may also suggest that ethnicity and SES are correlated and that SES has a stronger association with age at ASD diagnosis than ethnicity (Mandell et al., 2005).

Explanations for Disparities in Diagnosis

If disparities among groups do exist, several possibilities may explain why low SES and minority ethnic group status appear to be related to later ASD identification and diagnosis. First, differences in symptom interpretation may exist among caregivers. Second, traditionally underserved populations may have limited access to specialty care professionals who diagnose ASD. Third, clinicians may be biased in their identification of ASD among children from underserved populations.

Previous research indicates that there is a substantial gap between when ASD can be diagnosed and when ASDs are actually diagnosed. The most recent published surveillance study indicated that the current mean age for diagnosis of ASD is 4 years, 5 months (CDC, 2014).

Likewise, the average age of first parental concerns regarding their child's development is between 17 and 18 months, with 80-90% of parents reporting concerns about their child's development prior to age 2 years (Chawarska et al., 2007). The gap between first parental concern and actual diagnosis is concerning given the importance of early intervention. Possible explanations for the delay in diagnosis include lack of knowledge of early signs of ASD by professionals and limited access to specialty care professionals with expertise in ASD assessment (Saulnier & Ventola, 2012).

Symptom interpretation. In general, Coonrod and Stone (2004) found that parents were more likely to notice general developmental delays or regression in language skills than social or communicative deficits, regardless of cultural background. Evidence exists to suggest that caregivers from diverse ethnic and SES groups interpret ASD symptoms in different ways. Mandell and Novak (2005) outlined this possibility by suggesting that different racial/ethnic groups may differentially emphasize the importance of language or social delays during development. Likewise, caregivers from varying cultures may not be as aware of the typical timing of language, social, and motor developmental milestones when compared to caregivers from majority groups (Mandell & Novak, 2005). For example, in a study with Indian families of children with ASD, only 45% initially noticed social difficulties and only 32% noticed a delay in speech in their children (Daley, 2002). Similarly, Reijnevald, Harland, Brugman, Verhulst, and Verloove-Vanhorick (2005) found a weaker relationship between parent report and clinician identification of psychosocial problems in immigrant children compared to majority group

children. Cho, Singer, and Brenner (2003) found that Asian families were less likely than families from Western cultures to attribute their child's atypical behavior to a disorder. Likewise, Mandell et al. (2009) hypothesized that disparity in ASD diagnoses across SES, as measured by maternal education, may be due to less knowledge of developmental milestones or reduced ability to advocate for proper diagnosis, such as responding affirmatively to screening items.

Even among individuals who recognize symptoms of atypical development, conceptualization of a disability may differ (Welterlin & LaRue, 2007). For example, in many Native American groups, individuals with disabilities are perceived as spiritual leaders of their community with special gifts (Cho et al., 2003). Likewise, differences in perceived symptom severity have been reported across ethnic minority groups and those in poverty. Liptak et al. (2008) found that parent-reported ASD symptom severity was greater for African-Americans, Latinos, and those in poverty. In addition to differences in cultural beliefs, Zuckerman, Mattox, Sinche, Blaschke, and Bethell (2014) noted that historical mistreatment of minorities within health care and educational institutions may lead to minority parents attributing less value to providers and, therefore, seeking their assistance less often.

Access to care. Recent emphasis on early intervention services for children with ASD has led to increased importance being placed on early, accurate diagnosis of ASD (Osterling et al., 2002). Inaccurate assessments and diagnoses have significant consequences for treatment and educational planning as well as long-term prognosis for children with ASD. One study examined assessment practices for ASD evaluations within public schools, a hospital-based mental health clinic, and a community-based program for developmental disability evaluations (Williams, Atkins, & Soles, 2009). Authors found similar rates of ASD classification across

sites; however, the rate of agreement for diagnosis of individual children across clinicians from different sites was only 45%. Likewise, authors found that the majority of clinicians did not follow best practice guidelines in their ASD assessments (e.g., use of a standardized ASD measure, use of a standardized adaptive behavior measure, observation of the child in more than one setting, assessment of the child in their primary language). Notably, the highest rate of adherence to best practice in autism assessment was found for the hospital-based clinic. The hospital clinic employed an interdisciplinary diagnostic approach involving specialists (Williams et al., 2009).

Limited access to specialty care professionals among traditionally underserved populations could also lead to disparities in identification of ASD. For example, Mandell et al. (2005) found that children referred to a specialist received an ASD diagnosis, on average, four months earlier than children seen only by a primary care physician. General medical practitioners are typically not trained to recognize symptoms of ASD and may associate such symptoms with other conditions, such as Attention Deficit/Hyperactivity Disorder (ADHD) or intellectual disability. Through a study of parents of children under age 11 diagnosed with an ASD, Thomas, Ellis, McLaurin, Daniels, and Morrissey (2007) demonstrated that ethnic minority families, those with low levels of education, and those residing in non-metropolitan areas had limited access to specialty care. Authors found that racial minority families had one-fourth the odds of families from a racial majority group of accessing services from a psychologist or developmental pediatrician (Thomas et al., 2007). Likewise, Liptak et al. (2008) found that parent-rated ASD severity was inversely related to perceived access to specialty care, indicating that children most in need of specialist care experienced the most difficulty accessing such care.

In two prevalence studies, African-American children were more likely than Caucasian children to have been identified with an ASD based on school records rather than clinical evaluation, suggesting that African-American children may have less access to specialized clinical evaluation services than Caucasian children (Bhasin & Schendel, 2007; Yeargin-Allsopp et al., 2003). Similarly, Mandell et al. (2009) found that African-American, Hispanic, and children of other non-Caucasian ethnicities were less likely than Caucasian children to have a documented ASD in their records. For African-American children, this disparity existed even after controlling for IQ (Mandell et al., 2009). Furthermore, a research database study of more than 10,000 children with ASD found race differences in type of ASD diagnosed. Authors reported that Caucasian children were significantly more likely to be diagnosed with PDD-NOS or Asperger's Disorder (AspD) than children from other races. Likewise, children diagnosed with AspD were significantly less likely to be Latino than those with other ASD diagnoses (Rosenberg, Daniels, Law, Law, & Kaufmann, 2009). A CDC surveillance study also found that African-American children diagnosed with an ASD were more likely to be diagnosed with Autistic Disorder (AD) rather than AspD or PDD-NOS (Wiggins et al., 2006). Internationally, ethnic minorities have been found to be underrepresented among children referred to autism institutions compared to the known community prevalence, based on a study of case records in the Netherlands (Begeer, El bouk, Boussaid, Terwogt, & Koot, 2009).

Liptak et al. (2008) posited several explanations for discrepancies in access to care for racial and SES minority groups. Specifically, authors hypothesized that differences may stem from lack of understanding of the importance of symptoms, transportation, referrals to specialists, and a usual source of primary care, as well as differences in English-language proficiency. Another possibility for differences in access to care among ethnic minority and low

SES groups involves cultural conceptualizations of the importance of health care. For example, Liptak et al. (2008) theorized that members of minority cultural groups may be resistant to use health care services in an effort to "keep problems within the family" (p. 157).

Clinician bias. Several published studies have demonstrated clinician bias in ASD identification and diagnosis. According to Mandell et al. (2002), clinicians screen less often for ASD in children from minority ethnic groups. This disparity may be due to practitioners having different expectations regarding service needs by race or ethnicity. For example, practitioners may be more dismissive of concerns presented by African-American parents compared to those of Caucasian parents regarding developmental delays. Similarly, practitioners may be more likely to involve parents of majority versus minority racial groups in medical visits (Mandell et al., 2002). Likewise, results of a multisite study suggest that some practitioners may be less likely to further assess ethnic minority children for ASD when cognitive impairment is observed (Mandell et al., 2009). In Cuccaro et al.'s (1996) vignette study, clinician perceptions of ASD symptomatology were not influenced by the ethnicity of the child. However, another study employing vignettes found pediatricians' spontaneous clinical judgments based on vignettes to indicate ASD for European children more often than for non-European children (Begeer et al., 2009).

A recent study utilizing a web-based questionnaire collected data from pediatricians across six states regarding compliance with AAP screening guidelines. Results indicated that 59.8% of physicians screened for ASD in children age 18 months and 50.2% screened for ASD at the 24-month visit (Arunyanart et al., 2012). Interestingly, pediatricians with practices in which 10-30% of patients were Medicaid-insured were significantly less likely to screen for developmental delay when compared to pediatricians with more than 50% of patients insured by

Medicaid (Arunyanart et al., 2012). In contrast, pediatricians with 10-30% of patients insured by Medicaid were more likely to screen patients for autism than pediatricians with more than 30% of Medicaid-insured patients. This finding suggests that children of lower SES groups, for which Medicaid insurance serves as a proxy, are significantly more likely to be screened for developmental delay and significantly less likely to be screened for autism when compared to their higher SES counterparts (Arunyanart et al., 2012).

As evidence accumulates to support a disparity in ASD identification among traditionally underserved populations, the importance of early identification of ASD has become more apparent. A growing body of literature suggests that early identification of ASD leads to improvement in social, communicative, and cognitive skills (e.g., Osterling et al., 2002). Largely due to the consensus that early identification leads to improved outcomes, several early autism screeners have been developed. In one study in which clinician judgment of ASD differed by ethnicity, the disparity disappeared when explicit diagnostic criteria were used (Begeer et al., 2009). Using structured ratings, such as standardized screening instruments, may decrease ethnic or SES bias in ASD diagnosis. Correspondingly, routine screening for ASD at the population level has been shown to improve early recognition in traditionally underserved groups (Liptak et al., 2008).

Autism Spectrum Disorder Assessment

Autism spectrum disorders present a unique challenge to clinicians in terms of diagnosis given the heterogeneity in symptomatology and varying patterns of onset. The importance of evidence-based practice in professional psychology has garnered support in recent years. Intervention practices have received most attention within the movement toward evidence-based practice; however, the interplay between assessment and intervention should not be minimized

(Hunsley & Mash, 2007). Utilizing sound assessment practices increases the likelihood of choosing an effective evidence-based intervention. As such, clinical assessments can play a role in the treatment progress made by those being assessed. Evidence-based assessment is a guiding philosophy to assessment practice that uses research and theory to guide all aspects of an evaluation including selection of constructs to be measured, instruments used in the assessment, and interpretation of evaluation results (Hunsley & Mash, 2007). The use of psychometrically sound measures is an important consideration in assessment; however, the decision-making process for testing hypotheses within an assessment should also be informed by science.

Early Identification

Largely due to the consensus among clinicians and researchers that early identification and intervention lead to improved outcomes, increasing prevalence rates of ASD, and a lack of reliable biological markers for ASD, several early autism screeners have been developed for detection of ASD in young children (Johnson & Meyers, 2007; Osterling et al., 2002; Rogers & Vismara, 2008; Volkmar et al., 2004). Well check-up appointments serve as the primary means of developmental screening and health care in the United States.

The American Academy of Pediatrics (AAP) published guidelines indicating the organization endorsed ASD-specific screening in the general population beginning at age 18 months (AAP, 2006). The AAP recommends that general developmental screening occur at 9, 18 and 24, or 30 months or when parent concerns are noted. The AAP also recommends that primary care professionals conduct ASD-specific screening at 18 and 24 months (AAP, 2006). A recent study in a primary care setting found that ASD symptoms could be identified reliably at age 12 months using the *Communication and Symbolic Behavior Scales Developmental Profile Infant-Toddler Checklist* (CSBS-DP-IT-Checklist), a parent report screener (Pierce et al., 2011).

Pinto-Martin, Dunkle, Earls, Fliedner, and Landes (2005) found that repeated screening through primary care physicians decreases the age at which children begin receiving ASD-specific intervention. Additionally, using standardized screening instruments decreases differences in identification across racial groups (Pinto-Martin et al., 2005). A recent study of the association between compliance with AAP guidelines for well-child care check-ups and age at first diagnosis of Autistic Disorder (AD) in Medicaid-enrolled preschool-aged children found that overall compliance from birth to first AD diagnosis was 55%. Authors noted that the mean age at first AD diagnosis was 37.4 months and children whose well-child check-up compliance ranged from 76 to 100% were diagnosed earliest, suggesting that greater compliance with AAP guidelines may result in earlier diagnosis (Daniels & Mandell, 2013). Daniels and Mandell (2013) further reported that children in full compliance with AAP guidelines for check-ups were diagnosed, on average, 1.61 months earlier than children without any well-child visits.

In general, screening can be conceptualized as occurring at two levels. Level one screening is conducted with individuals in the general population, rather than with those known to be at-risk for a disorder (Barton et al., 2012). Level-one screeners are used to screen individuals in populations at low risk for the disorder to differentiate those children at risk for ASD from the general population. Such instruments may assess broad areas of development or assess disorder-specific symptoms but should be brief and require little expertise to complete or administer (Barton et al., 2012). Level two screening is undertaken for children who are identified through level one screening as at-risk for a specific disorder. As such, level two screeners should be utilized with selected populations known to be at-risk for a specific disorder (Barton et al., 2012). Level two screeners take more time to administer and more expertise to

score and interpret. Thus, level two screeners are not designed for use in primary care settings but are used in specialty-care or diagnostic clinics (Barton et al., 2012).

Two possible strategies for ASD screening have been posited. The first involves using a general screening instrument with all children then following up with an ASD-specific screener for those children identified as at-risk by the former measure. The second strategy involves utilizing an ASD-specific screener with all children as part of general screening in a primary care setting. Pinto-Martin et al. (2008) directly compared these strategies by evaluating the number of children screening positive for ASD risk on a general developmental screening instrument versus positive screens on an ASD-specific measure. Authors reported that, of those children identified as at-risk on the *Parents' Evaluation of Developmental Status* (PEDS), 16% screened positive for ASD on the *Modified Checklist for Autism in Toddlers* (M-CHAT; Pinto-Martin et al., 2008). Likewise, 14% of children who did not screen positive on the PEDS were identified as at-risk by the M-CHAT. Authors posited that general developmental screeners and ASD-specific screeners identify different areas of developmental concern (Pinto-Martin et al., 2008). Specifically, general developmental screeners may not reliably differentiate those children with language or cognitive delay from those at-risk for ASD. As such, results indicated the need for using an ASD-specific screener with all children (Pinto-Martin et al., 2008).

Early ASD screening instruments are designed to detect symptoms that signal the presence of ASDs, such as limited responsivity to name, eye contact, joint attention, or imitation skills as well as lack of social smile (Saulnier & Ventola, 2012). Desirable ASD screening instruments are those with the strongest psychometric properties (Sattler, 2008). Although the importance of the psychometric properties of screening instruments should not be understated, practicality of use in non-research settings is another important consideration. Specifically,

screeners should be brief and require no specialized training to administer and score. The majority of ASD screeners available for use in clinical settings rely on parent report to provide information regarding symptoms possibly indicative of ASD. Parent report is important given that parents' observations of children's behavior in naturalistic settings may be more accurate compared with behaviors observed in pediatric offices (Dumont-Mathieu & Fein, 2005). All ASD screeners measure observable behavior in an attempt to identify those children at risk for the disorders.

Several early ASD screening instruments with varying psychometric properties have been developed. As delineated by Sattler (2008), desirable screening instruments are those that are standardized, reliable, and valid. Screening instruments can be judged by criterion-related validity, or the degree to which a test correlates with an outcome measure, for which sensitivity and specificity are a measure. The most important property of screening instruments is that of predictive validity, specifically the ability to identify individuals with ASD (sensitivity) and exclude those without ASD (specificity) (Sattler, 2008). Ideal values for sensitivity and specificity are subjective and measure-dependent, as values are dependent upon the consequences of inaccurate diagnosis; however, such values are interdependent in that as one increases, the other decreases (Sattler, 2008). Positive predictive validity (PPV), or the proportion of clients screening positive for ASD who actually have the disorder, is typically low among ASD screeners because ASD is considered a low incidence disorder (Oosterling et al., 2009). Additionally, screening instruments should be brief and require minimal expertise to complete.

Screening Measures for Autism Spectrum Disorder

The *Modified Checklist for Autism in Toddlers* (M-CHAT) is a 23-item parent-report version of the *Checklist for Autism in Toddlers* (CHAT; Baron-Cohen, Allen, & Gillberg, 1992). The CHAT was designed to identify children exhibiting symptoms of Autistic Disorder (AD). The aim of the M-CHAT is to identify children at risk for any ASD, as opposed to specifically AD. The M-CHAT is intended for use with children ages 18-48 months; however, one study found the M-CHAT to be useful in detecting ASD in children as young as 16 months (Kleinman et al., 2008). The initial standardization study of the M-CHAT included 1,293 children at age 24 months. The M-CHAT has been adapted and validated across several languages including the Sinhala language spoken in Sri Lanka as well as Portuguese (Perera, Wijewardena, & Aluthwelage, 2009; Losapio & Ponde, 2008). Chinese and Japanese versions of the M-CHAT have also been developed (Inada, Kamio, & Koyama, 2010; Wong et al., 2004). A major advantage of the M-CHAT is that the measure is accessible with no associated cost.

The *Pervasive Developmental Disorders Screening Test – II* (PDDST-II) is a set of three screening instruments, consisting of three forms: a Primary Care Screener, a Developmental Clinic Screener (DCS), and an Autism Clinic Severity Screener (Siegel, 2004). The PDDST-II was developed for use with children between the ages of 12 and 48 months. The PDDST-II was designed as a parent-report rating scale and requires 10-20 minutes to complete (Siegel, 2004). The DCS is a parent-report measure consisting of 14 items and has shown moderate sensitivity (.73) and low specificity (.49) within an autism diagnostic service setting (Siegel, 2004). Although reportedly widely used as a screener (Wiggins, Bakeman, Adamson, & Robins, 2007), few studies independently evaluating the validity of the PDDST-II exist in the literature. In an unpublished thesis manuscript, Garland-Daniels (2002) examined the validity of the PDDST-II

in a sample of children referred for a developmental evaluation (*N* = 292). The author reported that, at a cut-off score of 10, the PDDST-II demonstrated a sensitivity of .63 and a specificity of .78. The author also investigated validity with a cut-off score of 8; however, nearly all participants with general developmental delay were identified at a cut-off score of 8, suggesting low specificity. The author posited that a cut-off score of 10 on the PDDST-II may be more appropriate for differentiating children with ASD from those with general developmental delay (Garland-Daniels, 2002). It should be noted that the PDDST-II manual recommends using a cut-off score of 5 for both the Stage 1-PCS and the Stage 2-DCS forms.(Siegel, 2004).

Similar to the M-CHAT in that both screeners were developed from a longer measure, the *Social Communication Questionnaire* (SCQ; Rutter, Bailey, & Lord, 2003) is a 40-item parent-report instrument based on the Autism Diagnostic Interview – Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994). The SCQ was designed for use with children ages 48 months and older; however, research has examined its extension for children as young as 17 months (Wiggins et al., 2007). Authors of the SCQ intended for the measure to detect risk for ASD as opposed to autism alone, which is similar to the objective of the M-CHAT. The SCQ has demonstrated high sensitivity (Oosterling et al., 2009). Eaves, Wingert, and Ho (2006) reported a specificity of 68% for the SCQ based on a study of children ages 2-6 years.

Snow and Lecavalier (2008) utilized two screening instruments, the SCQ and Modified-Checklist for Autism in Toddlers (M-CHAT), with a small sample of 39 preschoolers. Snow and Lecavalier found that the SCQ and M-CHAT were adequate screeners; however, the authors did not compare test accuracy for the scales and did not examine the potential improvement in prediction by combining screening results. Oosterling et al. (2009) compared three screening instruments, the M-CHAT, SCQ, and Infant-Toddler Checklist (ITC) in a sample of 238 young

children but, similar to Snow and LeCavalier (2008), did not contrast test accuracy statistically nor examine the potential of combining results from screening indications. Oosterling et al. (2009) found that the SCQ was not an adequate second-level screener. Schanding et al. (2012) recently examined the utility of the SCQ-Current and the *Social Responsiveness Scale* (SRS) as teacher-report screening instruments. Authors reported that the SCQ-Current and SRS produced lower sensitivity and specificity values than would be desirable; however, the SRS demonstrated slightly better predictive validity than the SCQ-Current as a teacher-report screener. Authors did not examine the predictive validity of the instruments when combined (Schanding et al., 2012). Overall, the literature offers little guidance for selecting ASD screening instruments for young children from among the current group of screeners.

Limitations of Screening

In spite of the importance of early detection of ASD for intervention programming, Barton et al. (2012) outlined several limitations for utilizing ASD-specific screening instruments with young children. One major limitation of screening for ASD in young children is the production of false positives or identification of children as at-risk for ASD who do not actually meet diagnostic criteria for the disorders. Higher false positive rates may also be observed for children diagnosed with PDD-NOS or "atypical autism" (Charwarska et al., 2007). Likewise, screening for ASD at early ages may also result in false negatives, or not identify children as at-risk for ASD who actually meet criteria for the disorder. In addition, exclusively screening for

ASD at young ages is particularly problematic for those children who experience regression or later emergence of ASD symptoms. Another issue regarding population-wide screening at well-baby-check-ups is that caregivers may not adhere to these typical check-ups and only visit the doctor when children are sick, which may decrease the likelihood of children being identified early as at-risk for ASD (Schanding et al., 2012).

Cultural Issues in ASD Screening

A major limitation of the ASD screening literature involves limited information related to potential differences across racial, ethnic, and socioeconomic status (SES) groups with respect to the psychometric properties of screeners. Harris, Barton, and Albert (2014) pointed out that, with culturally and linguistically diverse groups being the fastest growing segment of the U.S. population, greater emphasis should be placed on ensuring that ASD assessment tools are adequate for use with these populations. Growing evidence suggests that differences exist between racial/ethnic and SES groups with regard to age at first diagnosis of ASD. Harris et al. noted the possibility that the assessment methods practitioners currently use may not be appropriate for culturally and linguistically diverse populations, given the disparities in ASD identification in such populations.

Many studies have been conducted examining the psychometric properties of ASD-specific screeners translated into non-English languages; however, little evidence exists regarding the psychometric properties or predictive validity of such scales across various racial and SES groups for English-speaking populations. Within the ASD screening literature for young children, the researcher found no studies that examined differences in test accuracy or screening test scores between racial/ethnic or SES groups.

Although several ASD-specific screening instruments exist, there are general limitations of the ASD screening literature. Most important, few published studies have compared the predictive validity of more than one screening instrument on the same population. Another limitation of the ASD screening literature is a dearth of information regarding the potential differences across ethnic and SES groups with respect to the psychometric properties of ASD screening instruments. Validation studies have been conducted on most ASD screening instruments; however, evidence for cross-cultural validity is rarely reported. For example, in a validation study of the M-CHAT, racial and ethnic information was not collected from participants. The authors reported that there were children from minority and low socioeconomic groups in the sample, but specific demographics were not discussed. These limitations were noted in the published report of the M-CHAT validation study, but the authors still conclude that the M-CHAT is appropriate for use in the general population (Kleinman et al., 2008).

It is imperative that screening instruments demonstrate strong psychometric properties; few ASD-specific screening instruments have been validated across ethnic, racial, or SES groups (Daley, 2002). The issues related to the validity of using measures on a population not included in the original validation study of the measure were rarely mentioned. Similarly, ASD instruments are occasionally translated into other languages without examining the loss in psychometric properties (Williams et al., 2009). The dearth of ASD instruments adequately translated into languages other than English, and validated with diverse groups, signals the need for more cross-cultural ASD research.

The importance of understanding culture in the assessment of psychological disorders should not be understated. As such, the American Psychological Association (APA) developed

guidelines for practitioners serving ethnically, linguistically, or culturally diverse populations. These guidelines encourage providers of psychological services to be knowledgeable of multicultural assessment and interventions. Clinicians should also understand the impact of culture on behavior (APA, 1993). The most recent revision of the APA *Standards for Educational and Psychological Testing* (2014) updated and expanded upon requirements for assessment measures to be considered culturally fair. Specifically, APA called for validity studies to be completed for each of the intended examinee subgroups for published assessment instruments and discussed fairness of instruments as a fundamental issue related to test validity. For evaluations conducted in the public school setting, the Individuals with Disabilities Education Act requires that assessment and other evaluation materials be "selected and administered so as not to be discriminatory on a racial or cultural basis" (U.S. Department of Education, 2002).

Purpose of the Present Study

Hunsley and Mash (2007) emphasized the importance of developing assessment instruments and methods that take into consideration the diverse populations on which they will be used as part of an evidence-based approach to assessment. Additional research is needed regarding how variables such as ethnicity influence assessment measures and methods (Hunsley & Mash, 2007). Although the most desirable screeners demonstrate strong psychometric properties, few ASD-specific screening instruments have been validated across ethnic, racial or socioeconomic groups.

Kazdin (2005) noted that the evidence base for a particular assessment tool will never be so large as to include each possible moderating variable that may influence performance on that measure; however, clinicians must be aware of the limitations of using an instrument with groups

not adequately represented in the standardization group. To date, no known study has examined the predictive validity of early autism screening instruments across racial and SES groups.

If early ASD-specific screening measures are to be used for early identification of children at-risk for ASD, examining the psychometric properties of such instruments across race and SES groups is critical. According to Schmidt and Hunter (1974) in their discussion of differential validity, "when a test is not related to the criterion of interest within a population subgroup (i.e., has zero validity for the subgroup), there is no justification for its use with that subgroup" (p. 2). As such, adequately measuring ASD symptoms across race and SES groups is necessary to prevent inaccurate classification and is a prerequisite to comparing measurements across relevant subgroups within the population.

In light of the aforementioned premises, the current study seeks to examine the cross-cultural validity of three commonly used ASD screening instruments. The current study aims to address the following research questions:

(1) Do autism screening ratings by caregivers produce different mean scores across racial/ethnic and socioeconomic status (SES) groups?

(2) When compared to results from a standardized diagnostic evaluation, how well do three screeners differentiate participants with ASD from those without ASD across minority and non-minority groups?

CHAPTER 3

METHODS

Participants

Participants were 121 children and their parent(s) or caregiver(s) referred to the Boling Center for Developmental Disabilities (BCDD) located in Memphis, Tennessee between July 2010 and August 2012. The BCDD is administered by the College of Medicine of the University of Tennessee Health Science Center. The BCDD is one of 60 University Centers for Excellence in Developmental Disabilities in the United States. These centers were established to develop inter-disciplinary training and research with regard to developmental disabilities. During the data collection period, the BCDD conducted 2-3 diagnostic evaluations each week, for which ASD was a referral concern, within an outpatient diagnostic clinic. Children are typically referred to the Boling Center through child find agencies, school districts or community pediatricians for a range of developmental delays, including ASD.

A total of 121 child participants were included based on the inclusion criteria of having been referred for an evaluation at the BCDD between July 2010 and August 2012 as well as completion of all required study materials. Child participants consisted of 81% males (n = 98) and 17.7% (n = 23) females who ranged in age from 1.75 to 6.33 years (M = 3.86; SD = 1.15). Of note, the original intent of the examiner was to include only 12 to 48-month-old child participants; however, due to the small sample size, all participants with complete study data were included. Parent-reported child ethnicities were 44.6% (n = 58) African-American; 39.2% (n = 52) Caucasian; 9.2% (n = 12) Other Race. A review of 2010 U.S. Census data for the city

of Memphis indicated that 29.4% of residents identified as Caucasian and 63.3% identified as Black or African-American. As such, the remaining 7.3% would fall into the 'Other' category as defined in the current study (U.S. Census Bureau, 2010). Cognitive ability estimates were available for 120 child participants. Cognitive ability ranged from standard scores of 40 to 110 (M = 70.40, SD = 16.47).

Medicaid status was proposed to be used as an indicator for socioeconomic status (SES), as Medicaid enrollment is based on federal poverty levels; however, insurance information was not reported for 58.5% (n = 76) of child participants. As such, maternal education level was used as a proxy for SES. Demographic characteristics for child participants divided by diagnostic group are presented in Appendix A. Caregiver demographic data were collected from an information form completed by caregivers. For 86.8% (n = 105) of child participants, biological mother completed rating scales. The remaining 13.2% (n = 16) of child participants had rating scales completed by other caregivers including the biological father, adoptive parent, foster parent, grandparent, guardian, or aunt. Additional caregiver demographic data including parent education level, marital status, and race are presented in Appendix B.

Measures

Autism Diagnostic Observation Schedule (ADOS). The *Autism Diagnostic Observation Schedule* (ADOS; Lord, Rutter, DiLavore, & Risi, 2001) is a semi-structured observational assessment of an individual's communication, social interaction, and play or imaginative use of toys or other materials. The ADOS is built on the presentation of planned social interactions or "presses," which are intended to elicit behaviors or deficits indicative of an ASD. There are four developmentally sequenced modules that comprise the ADOS; however, only one module is chosen and administered based on the child's expressive language level and

age. Administration of the ADOS results in four separate scores. Items are scored between 0 and 2 or 0 and 3, with higher scores indicating more impairment. The ADOS is considered a "gold standard" diagnostic instrument for ASD and was used in the diagnostic evaluations conducted at the BCDD.

Modified Checklist for Autism in Toddlers (M-CHAT). The *Modified Checklist for Autism in Toddlers* (M-CHAT; Robins, Fein, Barton, & Green, 2001) is a parent report version of the Checklist for Autism in Toddlers (CHAT; Baron-Cohen, Allen, & Gillberg, 1992). The CHAT was designed to identify children exhibiting symptoms of AD. In contrast to the CHAT, the goal of the M-CHAT is early identification of children with ASD, as opposed to AD. The M-CHAT is one of the most widely used measures to assess for ASD in preschool-aged children.

The M-CHAT was developed to screen young children, as it does not include expressive language questions and includes few items related to receptive language. The M-CHAT consists of 23 items with a dichotomous response format and is intended for use with children ages 18-48 months; however, one study found the M-CHAT to be useful in detecting ASD in children as young as 16 months (Kleinman et al., 2008). The initial standardization study of the M-CHAT included 1,293 children at age 24 months. Authors reported excellent sensitivity (.87) and specificity (.99) (Robins, Fein, Barton, & Green, 2001); however, the reported sensitivity and specificity have been much lower in subsequent studies. In studies utilizing both at-risk and population samples, the M-CHAT shows adequate internal consistency reliability (Cronbach's α = .85; Kleinman et al., 2008) and has demonstrated good sensitivity (.77-.87) but variable specificity (.43-.99) within several samples of children between the ages of 18-48 months (Robins et al., 2001; Eaves, Wingert, & Ho, 2006a).

Psychometric properties of the M-CHAT have been investigated in studies utilizing both at-risk and population samples. Kleinman et al. (2008) performed an initial screen using the M-CHAT on selected and unselected samples at age 16-30 months and a second screen at 42-54 months. Fifteen participants were missed at the initial screen who screened positive for ASD on the second administration. Pandey et al. (2008) examined the M-CHAT with four groups of participants (i.e., younger/high-risk, younger/low-risk, older/high-risk, and older/low-risk) in an effort to determine if the false-positive rate was significantly different across age and risk levels. Authors reported that positive predictive power (PPP) was lowest for the younger/low-risk group (0.28), followed by PPP of 0.61 for the older/low-risk group. Higher PPP was observed for younger (0.79) and older (0.74) children at-risk for ASD; however, no significant differences in PPP by age were observed (Pandey et al., 2008). Kozlowski, Matson, Worley, Sipes, and Horovitz (2012) utilized a sample of 18-30-month-olds (N = 243) with positive screens on the M-CHAT. Authors reported that 141 participants were diagnosed with ASD while the remaining 102 did not meet criteria for ASD. As such, authors suggest that the M-CHAT may not be useful for at-risk populations (Kozlowski et al., 2012). Ventola et al. (2007), employing a selected sample of children diagnosed with ASD or other developmental disorders, found that 11 of the 23 M-CHAT items differentiated ASD from non-ASD participants. Similarly, Yama, Freeman, Graves, Yuan, and Campbell (2012) examined the M-CHAT with a large (N = 1,604), unselected sample of children ages 20-67 months. After excluding participants older than 48 months, results indicated that 95.93% of the sample screened negative for ASD on the M-CHAT, with the proportion of children with positive screens increasing with age (Yama et al., 2012).

The M-CHAT has been adapted and validated in several languages including the Sinhala language spoken in Sri Lanka as well as Portuguese (Perera, Wijewardena, & Aluthwelage, 2009; Losapio & Ponde, 2008). Chinese and Japanese versions of the M-CHAT have also been developed (Inada, Koyama, Inokuchi, Kuroda, & Kamio, 2011; Wong, Hui, & Lee, 2004). A major advantage of the M-CHAT is that the measure is accessible with no associated cost.

Of note, a revised version of the M-CHAT was published following completion of data collection for the current study. The *Modified Checklist for Autism in Toddlers – Revised with Follow-Up* (M-CHAT-R/F) was modified from the original M-CHAT to improve utility (Robins et al., 2014). Specifically, three questions from the M-CHAT were removed, the order of items was reorganized, and item wording was simplified for improved comprehension by raters. Additionally, authors revised scoring to include examination of total score only as opposed to alternate scoring (i.e., critical item total score). Robins et al. (2014) modified the scoring algorithm to divide total scores by risk level (i.e., Low-risk – Total score ≤ 3; Medium-risk – Total score = 3 – 7; High-risk = ≥ 8). Authors noted that no follow-up is necessary for individuals in the low-risk category; however, individuals with scores indicating medium-risk should complete the M-CHAT-R Follow-up items. Likewise, authors suggested that individuals with scores in the high-risk range should be referred immediately for a diagnostic evaluation (Robins et al., 2014).

Pervasive Developmental Disorders Screening Test – II (PDDST-II). The *Pervasive Developmental Disorders Screening Test – II* (PDDST-II) is a set of three screening instruments, consisting of three forms: a Primary Care Screener (PCS), a Developmental Clinic Screener (DCS), and an Autism Clinic Severity Screener (Siegel, 2004). The PDDST-II was developed for use with children between the ages of 12 and 48 months. The PDDST-II was designed as a

parent-report rating scale and requires 10-20 minutes to complete (Siegel, 2004). The DCS is a parent-report measure consisting of 14 items and has shown moderate sensitivity (.73) and low specificity (.49) within an autism diagnostic service setting (Siegel, 2004). Although reportedly widely used as a screener (Wiggins, Bakeman, Adamson, & Robins, 2007), few studies independently evaluating the validity of the PDDST-II exist in the literature. In an unpublished thesis manuscript, Garland-Daniels (2002) discussed an examination of the validity of the PDDST-II in a sample of children referred for a developmental evaluation (N = 292). The author reported that, at a cut-off score of 10, the PDDST-II demonstrated a sensitivity of 0.63 and a specificity of 0.78. The author also investigated validity with a cut-off score of 8; however, nearly all participants with general developmental delay were identified at a cut-off score of 8, suggesting low specificity. The author posited that a cut-off score of 10 on the PDDST-II may be more appropriate for differentiating children with ASD from those with general developmental delay (Garland-Daniels, 2002). It should be noted that the PDDST-II manual recommends using a cut-off score of 5 for both the PCS and DCS forms (Siegel, 2004). Of note, the PDDST-II includes questions regarding regression which are not often included on other early autism screeners. The PDDST-II Primary Care Screener (PCS) was utilized in the current study.

Social Communication Questionnaire (SCQ). The SCQ has shown variable sensitivity (.47 - .89) and specificity (.29 - .89) within samples of young children referred for evaluation (Allen et al., 2007; Wiggins et al., 2007). Similar to the M-CHAT, the *Social Communication Questionnaire* (SCQ; Rutter, Bailey, & Lord, 2003) is a 40-item parent-report instrument answered in a yes/no response format based on the Autism Diagnostic Interview – Revised, an instrument designed to assist in diagnosis of ASD for children at risk for developmental

problems (ADI-R; Lord, Rutter, & Le Couteur, 1994). There are two versions of the SCQ; the *Lifetime* version examines a child's entire developmental history while the *Current* form addresses the child's behavior in the last 3 months. The *Current* version was designed for use in evaluating treatment and educational plans while the *Lifetime* version provides information useful in a diagnostic assessment.

The SCQ was designed for use with children ages 48 months and older. Social Communication Questionnaire items correspond to DSM-IV-TR diagnostic criteria for Autistic Disorder. Total Scores on the SCQ can range from 0-39 for children with language and 0-32 for those without language. The cut-off score established as an indicator of symptomatology of ASD is a Total Score greater than or equal to 15 for both verbal and nonverbal children. The SCQ can be completed in 10 minutes by a caregiver.

The initial validation study of the SCQ indicated a sensitivity of 0.85 and a specificity of 0.75 for discriminating between ASD and non-ASD with a cut-off score of 15 in a clinical sample; however, ages of ASD and non-ASD participants ranged from 4-32, with few young children included (Berument, Rutter, Lord, Pickles, & Bailey, 1999). Similar to the M-CHAT, the SCQ aims to detect risk for ASD as opposed to autism alone.

Many studies conducted using the SCQ have included school-aged participants. Chandler et al. (2007) evaluated the SCQ with a large, high-risk sample of children with and without ASD and children from the general population. Authors reported sensitivity and specificity of 0.88 and 0.72, respectively, between ASD and non-ASD cases across the entire sample (Chandler et al., 2007). Goin-Kochel and Cohen (2008) examined the discriminative validity of the SCQ using a sample of children with a mean age of 9.5 years (SD = 5.6) who carried a diagnosis of ASD. Authors reported that 88.6% of participants were identified as at-

risk for ASD by the SCQ. Mulligan, Richardson, Anney, and Gill (2009) examined the SCQ using a small population sample of 5-13 year-olds in Ireland. Authors reported a range of scores (1-20) with mean of 3.89 (*SD* = 2.77). Additionally, Mulligan and colleagues (2009) found that some items on the SCQ were answered "autism-positive" for approximately one-third of children in the population sample, suggesting that some items on the SCQ may not discriminate well between ASD and non-ASD.

Witwer and Lecavalier (2007) found sensitivity of .92 and specificity of .62 with a sample of children ages 4-14 who were diagnosed with ASD or with intellectual disability only. The SCQ correctly classified over 80% of the sample (Witwer & Lecavalier, 2007). Bolte, Holtman, and Poustka (2008) found sensitivity of 0.89 and specificity of 0.91 between ASD and non-ASD for the SCQ using a representative child and adolescent psychiatric sample with a mean age of 14.1 years (*SD* = 8.8). A recent examination of the SCQ-Lifetime with school-aged children found much higher sensitivity and specificity compared with previous studies (.75, .99, respectively). Likewise, authors examined the SCQ-Current as a teacher-report screening instrument. Results indicated high specificity (.95) and moderate sensitivity (.60) for discrimination between ASD and non-ASD for teacher ratings (Schanding, Nowell, & Goin-Kochel, 2012).

Several studies have extended the SCQ downward by examining the properties of the measure with preschool-aged children. Eaves et al. (2006a) reported sensitivity of 0.74 and specificity of 0.54 for the SCQ based on a study of a clinical sample of children ages 2-6 years; however, it should be noted that authors did not specify whether the SCQ Current or Lifetime form was used. Eaves, Wingert, Ho, and Mickelson (2006b) examined the psychometric properties of the SCQ with ASD-specific and general developmental clinic samples of children

whose ages ranged from 36 to 82 months. Authors reported an overall sensitivity of .71 for both clinics and specificity of .62 and .53 for the developmental and ASD-specific clinics, respectively (Eaves et al., 2006b). Additionally, authors found that 28% of children with ASD were missed by the SCQ at a Total Score cut-off of 15. Likewise, 38% of children without ASD were identified as at-risk (i.e., false positives). Authors found that only 15 of the 40 items on the SCQ differentiated children with and without ASD in the sample (Eaves et al., 2006b). Lee, David, Rusyniak, Landa, and Newschaffer (2007) examined the SCQ with a large sample of 3-5-year-olds drawn from children receiving special education services in several public school districts. Authors reported sensitivity of 0.59 and specificity of 0.63 for SCQ ratings with ADOS scores in the ASD range as the criterion (Lee et al., 2007). Oosterling et al. (2009) utilized a clinical sample of 8-44-month-olds referred for possible ASD to examine the SCQ. Authors reported sensitivity of 0.66 and specificity of 0.64 between ASD and non-ASD (Oosterling et al., 2009).

As noted by Wei, Chestnut, Barnard-Brak, and Richman (2015), few studies have examined the SCQ-Current and have instead evaluated the SCQ-Lifetime. Corsello et al. (2007) examined the SCQ-Current with preschool-aged participants (i.e., <5 years) and utilized the SCQ-Lifetime with school-aged participants in a large (N = 208) sample of 2-16 year-olds referred to a university clinic specializing in ASD assessment. Authors reported that, using the cut-off score of 15, the SCQ missed a large number of young children (i.e., false negatives); however, given that parents of children under age 5 years completed the SCQ-Current form, it is unclear whether the lower specificity in comparison to the SCQ-Lifetime was due to the difference the SCQ form version or participant age. Nonetheless, authors reported overall sensitivity and specificity of .71 for discriminating between those at-risk for an ASD versus non-

ASD (Corsello et al., 2007). Oosterling et al. (2010) attempted to replicate the findings of Corsello et al. in a clinical sample of children aged 20 – 40 months (*N* = 208). Authors employed the Dutch version of the SCQ-Current. Authors reported a sensitivity of .76 and a specificity of .62 for an Autistic Disorder diagnosis at a cut-off score of 15. Oosterling et al. concluded that the SCQ-Current is not an optimal level two screener given the low specificity observed; however, similar to Corsello et al., it is unclear whether results were confounded by the young age of participants when compared to the older age of participants included in other published studies examining the SCQ-Lifetime form.

In general, the range of reported psychometric properties for both forms of the SCQ is likely due to demographic and diagnostic characteristics of the samples used. As would be expected, psychometric properties appear to be better with selected as opposed to unselected samples. In the current study, a cut-off score of 15 was used to indicate a "positive" screen on the SCQ, as is suggested in the manual (Rutter, Bailey, & Lord, 2003), despite concerns reported in the literature regarding using this cut-off with young children (Corsello et al., 2007; Eaves et al., 2006).

Cognitive assessment. Formal cognitive assessment is routinely included as part of BCDD diagnostic evaluations. For the evaluations in the current study, the *Stanford-Binet Intelligence Scales, Fifth Edition* (SB-5; Roid, 2003) was utilized to assess cognitive ability in the majority of participants. For the SB-5, evaluators reported either a Full Scale IQ (FSIQ) or an Abbreviated IQ (ABIQ) depending on whether the full or abbreviated battery was administered. Depending on the age and developmental level of the participant, evaluators also reported Cognitive domain standard scores for the *Bayley Scales of Infant and Toddler Development, Third Edition* (Bayley-III; Bayley, 2006). At least one full-scale standard score

from a cognitive measure was available for all participants with the exception of one (n = 120). For participants for whom both SB-5 and Bayley-III scores were reported, SB-5 scores were used in analyses.

Demographic Information Form. Data on additional variables (e.g., gender, ethnicity) were obtained from a demographic information questionnaire completed by caregivers. Information provided by parents on the Demographic Information Form included: child date of birth, child race, child gender, health insurance type (i.e., private, Medicaid, no insurance), maternal/paternal marital status, maternal/paternal race, maternal/paternal education level (i.e., no high school, high school graduate, some college, Associate's degree, Bachelor's degree, Master's degree, Doctorate), maternal/paternal occupation, and maternal/paternal income. Variables of interest included child race, child health insurance (i.e., Medicaid, private insurance, no insurance), and child age.

Procedures

Participant recruitment and data collection. After completion of BCDD intake procedures, screening instruments including the SCQ, PDDST-II, and M-CHAT were given to caregivers. It should be noted that the screening instruments completed by parents were not a routine part of BCDD evaluations and were administered solely for research purposes. Caregivers provided consent for research participation prior to completing screening instruments. Screening instruments were included with BCDD materials given to caregivers as part of a clinic information packet issued after referral. Screening instruments were randomly counterbalanced in order to minimize response bias that may arise if instruments were presented in identical order. In order to prevent criterion contamination (i.e., screening results influencing diagnostic decision making), BCDD evaluators did not have access to the screening instruments or

screening results during the evaluation process. Participants' screeners were assigned consecutive numbers that corresponded with BCDD clinic identification numbers in order to match screener results to diagnostic findings. Institutional Review Board (IRB) approval through both the College of Medicine of the University of Tennessee Health Science Center and The University of Georgia was obtained prior to beginning data collection.

BCDD diagnostic evaluations. Clinicians with the BCDD rendered diagnostic decisions via: (a) the Autism Diagnostic Observation Schedule (ADOS), (b) a structured developmental interview currently in use at the BCDD, and (c) diagnostic checklists keyed to the DSM-IV-TR diagnostic definitions of Autistic Disorder, Asperger's Disorder, or Pervasive Developmental Disorders, Not Otherwise Specified, currently in use at the BCDD. When appropriate, non-ASD diagnoses were also made either exclusively or concurrently with ASD and included intellectual disability, Attention Deficit/Hyperactivity Disorder (ADHD), language disorders, or behavior disorders based on DSM-IV-TR diagnostic criteria.

Analytic Method

Data Screening. Following completion of diagnostic evaluations, participant data were entered into a de-identified SPSS data file. Demographic data and item-level responses and scores from each measure were extracted. Missing items were investigated, and participants without a sufficient number of items completed to allow a measure (i.e., M-CHAT, SCQ-Current, PDDST-II, or ADOS) to be scored for one or more measures were excluded from analyses. Likewise, participants for whom a final DSM-IV TR Axis I diagnosis was not available were removed prior to main analyses. Overall, nine participants were excluded from analyses due to missing data. Missing data for items on the M-CHAT and SCQ-Current were replaced with the mean value for that participant's ratings on the measure when, per each

measure's scoring manual, an allowable number of items were left unanswered. The mean value for individual item ratings on the M-CHAT was used to replace one missing item for two participants. For SCQ-Current ratings, the mean value was used to replace one missing item for two participants. No items were missing from any PDDST-II form for any rater.

Research Question One: Comparing Screener Total Scores Across Race and SES Groups. The first research question was addressed by comparing raw screening scores across race groups and maternal education levels to determine if caregivers produce different mean scores. Socioeconomic status (SES) was originally intended to serve as a second independent variable; however, due to insurance type (a proxy for SES) having been reported for only 36.6% (n = 45) of child participants, Medicaid status was excluded from examination as an independent variable. Instead, the researcher chose to include maternal education level as an independent variable. Maternal education level was chosen given that the majority of screeners were completed by the biological mother (n = 105). Likewise, the researcher hypothesized that screener total scores would differ across maternal education levels.

Four two-way analysis of variance (ANOVAs) were employed to examine the relationship between race and maternal education on screener total scores. Within the two-way ANOVA, parent-reported child race and maternal education level served as independent variables, with total raw score for each of three screening instruments serving as the dependent variable. Race was coded into three categories (i.e., African-American, Caucasian, Other); however, due to small numbers of participants in the 'Other' category, only two levels of race (i.e., African-American, Caucasian) were utilized for two-way ANOVAs. For both mothers and fathers, rates of response at each level of postsecondary education (i.e., "Some college," "Associate's degree," "Bachelor's degree", "Master's degree," "Doctorate") were low; therefore,

postsecondary education levels were collapsed into "Some College or Higher." The additional levels of maternal education included: "no high school diploma" and "high school graduate." Four separate two-way ANOVAs were conducted to examine mean total scores across race groups and maternal education levels on the M-CHAT, M-CHAT Critical Items, PDDST-II, and SCQ-Current. Main effects of race and maternal education level as well as the interaction between the two variables on screener total scores were examined. Level of significance was defined as $p < .05$.

Research Question Two

The second research question was examined using receiver operating characteristic (ROC) curve analyses to compare test accuracy findings across race and maternal education levels. Guidelines delineated in Youngstrom (2014) for conducting and interpreting ROC analyses were utilized. First, the dichotomous criterion variable (i.e., ASD diagnosis) was defined. A broad definition of ASD was used to include DSM-IV-TR diagnoses of Autistic Disorder, Asperger's Disorder, and PDD-NOS subsumed under the category of a "positive" diagnosis. All other final diagnoses (e.g., Generalized Anxiety Disorder) were coded as "negative" for ASD. Of note, as mentioned previously, the criterion diagnosis was made blind to the predictor (i.e., screener) test results in order to prevent criterion contamination.

The researcher compared and contrasted screeners' predictions of results from a standardized diagnostic assessment, which includes a gold standard diagnostic measure. The researcher compared test accuracy by testing for differences between areas under the ROC curve (AUCs) generated by each screening measure to examine the predictive validity of screening measures across racial/ethnic groups and maternal education levels.

In order to compare differences in accuracy across the three screeners, receiver operating characteristic (ROC) analysis, which provides information about sensitivity and specificity at all possible cut-off scores, was utilized. Areas under the ROC curve (AUCs) for each screening instrument were calculated, which provided an index of each screener's accuracy or discriminant validity in correctly classifying participants with ASD and without ASD. Areas under the ROC curve range from .50 (random accuracy) to 1.0 (perfect accuracy) and are interpreted as the probability of correctly classifying a pair of individuals, one with ASD and one without. Youngstrom (2014) described the AUC as the probability that a randomly selected case with a particular disorder would have a higher score on the index test (i.e., screener) than a randomly selected case without the disorder. For each test, the researcher determined if the screener performed at better than chance levels. Commonly used criteria for AUCS suggest that values ≥0.90 are "excellent," ≥0.80 "good," ≥0.70 "fair," and <0.70 "poor" (Cicchetti, Volkmar, Klin, & Showalter, 1995). However, Youngstrom pointed out that such AUC criteria are less useful in the social sciences. He further noted that the AUC is constrained by the reliability and validity of the reference standard (i.e., ASD diagnosis), which is often imperfect with respect to mental health assessment. As such, it is nearly impossible to obtain an AUC of 1.0 given the inherent error in the criterion diagnosis. Youngstrom also explained that many of the most widely used and best-performing behavior checklists have reported AUC values between 0.7 – 0.8.

Next, the researcher compared the AUCs produced by the three screening instruments to determine if the screening instruments differed in diagnostic accuracy. Areas under the curve were compared using Hanley and McNeil's (1983) methodology, which allows for AUC comparisons derived from the same individual in pairwise fashion, correcting for correlations that exist between AUCs. Six comparisons were conducted (i.e., SCQ - M-CHAT; M-CHAT -

PDDST-II; SCQ - PDDST-II; M-CHAT – M-CHAT Critical; PDDST-II – M-CHAT Critical; M-CHAT Critical - SCQ) via *z* tests using the following formula (Hanley & McNeil, 1983):

$$z = AUC_1 - AUC_2 / [(SE_1)^2 + (SE_2)^2 - 2rSE_1SE_2]^{1/2}$$

where AUC_1 and SE_1 refer to the observed area and estimated standard error for test 1; AUC_2 and SE_2 refer to the observed area and estimated standard error for test 2, and *r* references the correlation between AUC_1 and AUC_2.. Sensitivity and specificity were examined for each measure.

CHAPTER FOUR

RESULTS

Preliminary Analyses

To investigate whether a relationship existed between the two independent variables (Race and Maternal Education), a chi-square analysis was employed to ensure that maternal education level did not differ across race categories. Results indicated no significant race group differences across maternal education level, χ^2 (2, N = 118) = 4.7, *ns*.

Independent samples *t* tests were conducted to determine whether mean total scores for each screener differed across diagnostic groups (i.e., ASD vs. non-ASD). Interestingly, *t* statistics revealed no significant differences in mean total scores on the PDDST-II, $t(119) = 0.60$, *ns*, SCQ, $t(119) = 0.37$, *ns*, or M-CHAT, $t(119) = 1.72$, *ns*, between diagnostic groups (ASD, non-ASD). In contrast, the *t* statistic revealed a significant difference between ASD and non-ASD diagnostic groups for mean total M-CHAT critical items, $t(119) = 2.97$, $p = .004$. Mean total scores for each screener across diagnostic groups are presented in Table 1. Correlations among screener total scores are presented for the total sample in Table 2 and separated by diagnostic group in Tables 3 and 4. All screeners were positively correlated at the $p < .01$ level.

An independent samples *t* test was also conducted to compare mean total cognitive standard scores between ASD and non-ASD groups to determine whether intellectual functioning differed significantly between diagnostic groups. The *t* statistic revealed a significant difference in mean total cognitive standard scores between ASD and non-ASD groups, $t(118) = 4.47$, $p < .0001$, suggesting that participants in the ASD diagnostic group demonstrated significantly lower cognitive functioning than those in the non-ASD group.

One-way ANOVAs were also conducted to investigate mean differences in total cognitive standard scores across race and maternal education groups. No significant differences were observed among levels of race for cognitive standard score, $F(2, 117) = 1.11$, *ns*. In contrast, a significant difference in mean cognitive standard scores was observed across levels of maternal education, $F(2, 114) = 3.55$, $p = 0.03$. Post hoc comparisons using the Tukey HSD indicated that the mean cognitive total score for the 'No High School' group was significantly lower than the mean cognitive total score for those in the 'Some College or Higher' group ($M_{Difference} = 12.19$, $SE = 4.58$, $p = .024$). However, mean cognitive total scores for high school graduates did not differ significantly from those in the 'No High School' or 'Some College or Higher' groups.

Table 1
Mean Total Scores and Standard Deviations for M-CHAT, PDDST-II, and SCQ-Current

	ASD (n = 70)			Non-ASD (n = 51)		
	M	(*SD*)	Range	*M*	(*SD*)	Range
M-CHAT	6.93	(3.7)	0 - 16	5.63	(4.6)	0 - 17
M-CHAT Critical Items	2.07	(1.5)	0 – 5	1.24	(1.5)	0 - 5
PDDST-II	9.43	(4.2)	0 - 18	9.88	(3.9)	0 - 19
SCQ-Current	16.67	(5.7)	5 - 27	16.24	(7.4)	2 - 31

Note. *M* = mean, *SD* = standard deviation

Table 2
Pearson's Correlations among M-CHAT, PDDST-II, and SCQ-Current Total Scores, Total Sample (N = 121)

	M-CHAT	M-CHAT Critical	PDDST-II	SCQ-Current
M-CHAT	--	.85**	.55**	.80**
M-CHAT Critical Items	.85**	--	.42**	.66**
PDDST-II	.55**	.42**	--	.57**
SCQ-Current	.80**	.66**	.57**	--

*Note. **indicates significance at the p < .01 level*

Table 3
Pearson's Correlations among M-CHAT, PDDST-II, and SCQ-Current Total Scores for ASD Diagnostic Group (n = 70)

	M-CHAT	M-CHAT Critical	PDDST-II	SCQ-Current
M-CHAT	--	.85**	.65**	.81**
M-CHAT Critical Items	.85**	--	.49**	.70**
PDDST-II	.55**	.49**	--	.57**
SCQ-Current	.80**	.70**	.57**	--

*Note. **indicates significance at the p < .01 level*

Table 4
Pearson's Correlations among M-CHAT, PDDST-II, and SCQ-Current Total Scores for Non-ASD Diagnostic Group (n = 51)

	M-CHAT	M-CHAT Critical	PDDST-II	SCQ-Current
M-CHAT	--	.85**	.48**	.81**
M-CHAT Critical Items	.85**	--	.40**	.67**
PDDST-II	.48**	.40**	--	.59**
SCQ-Current	.81**	.67**	.59**	--

*Note. **indicates significance at the p < .01 level*

Correlations between screener total score and cognitive ability estimates were also evaluated to investigate the possibility that ratings on screeners were related to level of intellectual functioning. Correlations are presented in Table 5.

Table 5
Pearson's Correlations Between M-CHAT, M-CHAT Critical Items, PDDST-II, and SCQ-Current Total Scores and SB-5/Bayley-III Standard Score, Total Sample

	M-CHAT	M-CHAT Critical	PDDST-II	SCQ
SB-5/ Bayley-III	-.09	-.16	.04	-.07

Note. ***indicates significance at the $p < .01$ level*

Research Question One

Previous research (e.g., Mandell et al., 2002) suggests a disparity in identification of ASD across racial groups and SES levels. The first research question sought to investigate whether mean screener scores differed across groups. One-way analyses of variance (ANOVAs) were conducted to investigate significant differences in screener total scores across maternal education levels. No significant differences were observed among levels of maternal education for M-CHAT Total Score, $F(2, 115) = 1.15$, *ns*, for M-CHAT Critical Item Total, $F(2, 115) = 1.84$, *ns*, or for SCQ-Current Total Score, $F(2, 115) = 1.28$, *ns*. In contrast, a significant difference for PDDST-II total scores across maternal education levels was observed, $F(2, 115) = 3.72$, $p = .03$. Post hoc analyses were conducted given the statistically significant F test. Post hoc comparisons using the Tukey HSD indicated that the mean PDDST-II total score for high school graduates was significantly higher than the mean PDDST-II total score for those with some college or higher ($M_{Difference} = 1.92$, $SE = .80$, $p = .047$). However, PDDST-II mean total scores for 'No High School' group did not differ significantly from those for high school graduates or those with some college or higher levels of education.

One-way ANOVAs were conducted to investigate differences in screener total scores across race groups. No significant differences among screener total scores were observed across race groups for M-CHAT, $F(2, 118) = 1.05$, *ns*, M-CHAT Critical Items, $F(2, 118) = 2.30$, *ns*, PDDST-II, $F(2, 118) = 1.12$, *ns*, or SCQ, $F(2, 118) = 1.67$, *ns*, Total Scores. Mean total scores across race and maternal education levels are presented in Table 6.

Table 6
Mean Total Scores and Standard Deviations for M-CHAT, PDDST-II, and SCQ-Current by Race and Maternal Education Level

	No High School			High School Grad			Some College or Higher		
	n	*M*	*SD*	*n*	*M*	*SD*	*n*	*M*	*SD*
M-CHAT									
African-American	9	7.0	5.5	25	7.8	4.2	22	5.8	4.3
Caucasian	6	7.5	4.6	16	6.3	4.4	28	5.6	3.5
Other	1	1.0	--	7	5.4	4.0	4	5.3	3.9
M-CHAT Critical									
African-American	9	2.2	2.1	25	2.3	1.6	22	1.3	1.5
Caucasian	6	2.2	1.5	16	1.6	1.7	28	1.6	1.6
Other	1	1.0	--	7	1.3	1.5	4	1.0	1.2
PDDST-II									
African-American	9	11.6	4.0	25	11.0	4.6	22	8.8	4.4
Caucasian	6	10.2	2.9	16	10.3	3.5	28	8.5	4.1
Other	1	9.0	--	7	9.0	4.1	4	7.5	2.6
SCQ-Current									
African-American	9	15.9	6.8	25	18.8	6.1	22	15.7	6.2
Caucasian	9	17.0	7.4	16	17.1	8.3	28	15.5	5.6
Other	1	19.0	--	7	13.1	6.7	4	13.0	5.6

Note. M = mean, *SD* = standard deviation

Additionally, a two-way analysis of variance (ANOVA) was employed with race and maternal education level serving as independent variables and each screener's total raw score serving as the dependent variable. A 3 (Maternal Education: No High School, High School Graduate, Some College or Higher) x 2 (Race: African-American, Caucasian) non-randomized blocking design was used for the two-way ANOVAs. It should be noted that, due to the small number of participants in the 'Other' race group, this level was omitted from two-way ANOVAs. The means, standard deviations, and ranges for screener total scores as a function of the two factors are presented in Table 6. The assumption of normality was not met for levels of maternal education. Specifically, the Wilks-Shapiro statistic revealed a non-normal distribution for the 'Some College or Higher' group on the M-CHAT ($W = 0.94$, $p = .013$). The 'No High School' group was not normally distributed on the PDDST-II ($W = 0.70$, $p = .0001$). Neither the 'High School Graduate' nor the 'Some College or Higher' levels were normally distributed on the M-CHAT Critical. Likewise, the Wilks-Shapiro statistic revealed a non-normal distribution for the 'African-American' group on the M-CHAT ($W = 0.95$, $p = .03$) and for both the 'African-American' ($W = 0.88$, $p = .0001$) and 'Caucasian' ($W = 0.87$, $p = .0001$) groups on the M-CHAT Critical. Results of Levene's test for equality of variances suggest that the assumption of homogeneity of variance was not violated for the following analyses for all dependent variables. Given that the assumptions of normality were violated, Kruskal-Wallis tests were conducted for all screeners. Results of the Kruskal-Wallis tests were consistent with ANOVA findings.

For the M-CHAT, the two-way ANOVA indicated no significant interaction between race and maternal education level, $F(2, 100) = 0.46$, $p = .63$, $\eta^2 = .01$. The main effect for maternal education on M-CHAT total score was not significant, $F(2, 100) = 1.25$, $p = .29$, $\eta^2 = 02$. Likewise, the main effect for race on M-CHAT total score was not significant, $F(1, 100) = .13$, $p = .72$, $\eta^2 = .001$.

For M-CHAT Critical Items Total Score, the two-way ANOVA indicated no significant interaction between race and maternal education level $F(2, 100) = 1.03$, $p = .36$, $\eta^2 = .02$. The main effect for maternal education on M-CHAT Critical Item total score was not significant, $F(2, 100) = 1.78$, $p = .18$, $\eta^2 = .03$. The main effect for race was not significant, $F(1, 100) = 0.12$, $p = .73$, $\eta^2 = .001$.

For PDDST-II total scores, the two-way ANOVA indicated no significant interaction between race and maternal education level, $F(2, 100) = 0.10$, $p = .90$, $\eta^2 = .002$. No significant main effect was found for race, $F(1, 100) = 0.79$, $p = .38$, $\eta^2 = .01$. In contrast, a significant main effect for maternal education was observed, $F(2, 100) = 3.13$, $p = .048$, $\eta^2 = .06$. Tukey's HSD indicated that the PDDST-II mean total score was significantly lower for those in the 'Some College or Higher' group when compared to those in the 'High School' group ($M_{Difference} = 2.09$, $SE = .87$, $p = .048$).

For SCQ total scores, the ANOVA indicated no significant interaction between race and maternal education level, $F(2, 100) = 0.29$, $p = .75$, $\eta^2 = .01$. Neither the main effect for maternal education on SCQ total scores, $F(2, 100) = 1.48$, $p = .23$, $\eta^2 = .03$, nor the main effect for race on SCQ total scores, $F(1, 100) = 0.03$, $p = .86$, $\eta^2 = .0001$, was significant.

It should be noted that, due to sample size and observed effect size, power was likely inadequate to detect significant main effects and interactions. For example, observed power

equaled .31 for the maternal education level effect, .05 for the race effect, and .09 for the interaction effect for the two-way ANOVA with SCQ total score as the dependent variable. Typically, statistical power estimates above .80 are considered adequate. For the current analyses, it is unclear whether the null hypothesis could not be rejected due to inadequate power or due to there being no significant differences in mean screener total scores across race and maternal education groups.

Research Question Two

The second research question was addressed through receiver operating characteristic (ROC) curve analysis. The researcher intended to compare screener accuracy across race and SES groups; however, as mentioned previously, low response rates were observed for Medicaid eligibility. As such, screener accuracy was examined across race and maternal education levels to determine if screeners performed differently across groups. The ability of each screener to predict a dichotomous diagnostic outcome (i.e., ASD vs. non-ASD final diagnosis) was examined. The dichotomous outcome variable of ASD or non-ASD was used as the outcome variable for all ROC analyses.

In order to compare and contrast screeners' predictions of results from a standardized diagnostic assessment, the researcher compared test accuracy by testing for differences between areas under the ROC curve. Area under the curve (AUC) serves as an estimate of a screener's accuracy or discriminant validity. Areas under the curve range from .50 (random accuracy) to 1.0 (perfect accuracy) and are interpreted as the probability of correctly classifying a pair of participants, one with ASD and one without.

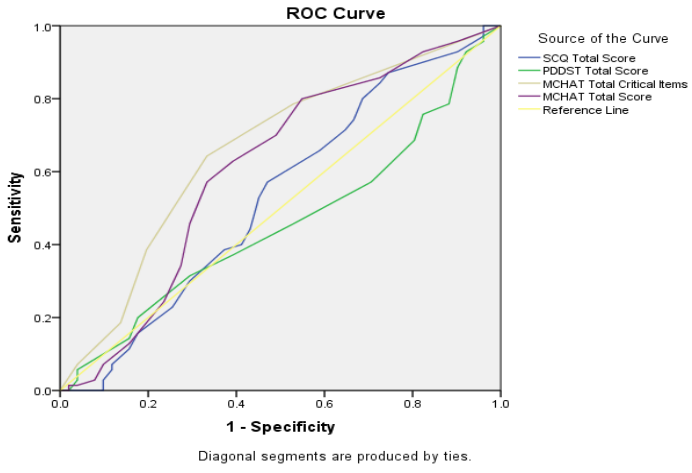
With respect to the ROC curve, scores closest to the upper left corner of the graph represent the best balance between sensitivity and specificity. The upper left point of the graph

is the perfect classification, indicating that all true positives are identified without any false positives (Pintea & Moldovan, 2009). The diagonal line represents the null hypothesis that the screening instrument performs randomly, in that a random instrument would correctly identify half of the positive screens and half of the negative screens as either having ASD or not having ASD. Cut-off points above the line perform better than random or chance, whereas points below the line do not perform better than chance.

Results of the ROC analysis for the total sample (N = 121) produced the following AUC values: AUC = .53 for the SCQ-Current, AUC = .46 for the PDDST-II, AUC = .61 for the M-CHAT Total, and AUC = .66 for the M-CHAT Critical Item total. Confidence intervals for SCQ-Current and PDDST-II total scores included the null hypothesis of .50. In contrast, the null hypothesis was not included in AUC 95% confidence intervals for M-CHAT and M-CHAT critical item confidence intervals. Based on the asymptotic significance index value, M-CHAT and M-CHAT critical item AUCs were significantly different from .50 at the $p < .05$ level. However, based on commonly used criteria for AUCS (i.e., values ≥0.90 are "excellent," ≥0.80 "good," ≥0.70 "fair," and <0.70 "poor"), discriminant validity for each of the screeners would fall into the "poor" category across all groups for the current sample. ROC curves for each screener including all participants (N = 121) is presented in Figure 1.

Figure 1

ROC Curves for M-CHAT, M-CHAT Critical Items, SCQ, and PDDST-II Total Scores, Total Sample



Results of ROC analysis indicated that none of the AUC values for any of the four screener total scores across levels of race or maternal education was significantly different from .50 at the $p < .05$ level based on the asymptotic significance index. Likewise, all 95% confidence intervals for M-CHAT, SCQ, and PDDST-II total scores with the exception of the AUC for M-CHAT/Other race and M-CHAT/No High School included the null hypothesis of .50. In contrast, AUCs for M-CHAT Critical Item total score for the Caucasian, Other Race, and Some College or Higher subsamples did not include the null hypothesis in the 95% confidence interval. Based on commonly used criteria for AUCS (i.e., values ≥0.90 are "excellent," ≥0.80 "good," ≥0.70 "fair," and <0.70 "poor"), each of the screeners' discriminant validity would fall in to the "poor" category across all groups. ROC curves for each screener across each racial category are presented in Figures 2-4. Likewise, screener AUCs, standard errors, and 95% confidence intervals across groups are presented in Table 7. Receiver Operating Characteristic curves for each screener across each level of maternal education are presented in Figures 5 and 6.

Figure 2

ROC Curves for M-CHAT, M-CHAT Critical Items, SCQ, and PDDST-II Total Scores, African-American

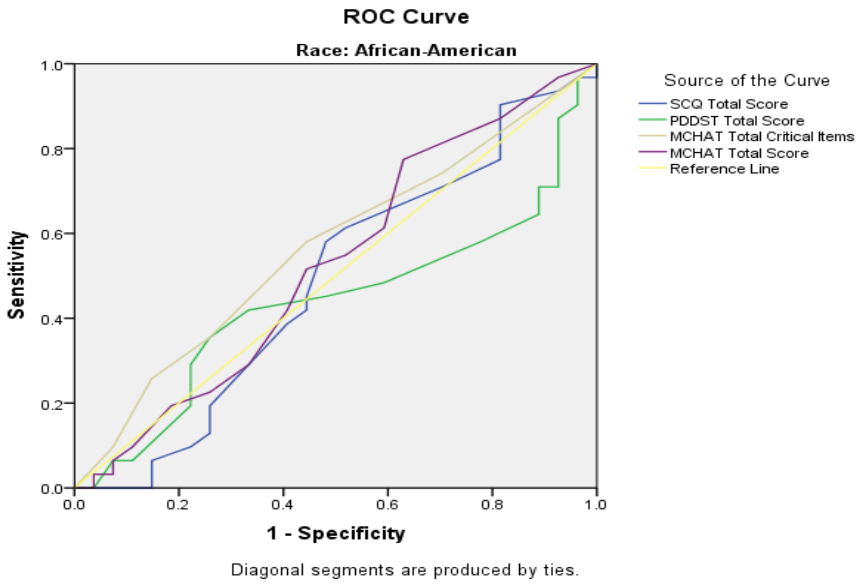


Figure 3

ROC Curves for M-CHAT, M-CHAT Critical Items, SCQ, and PDDST-II Total Scores, Caucasian

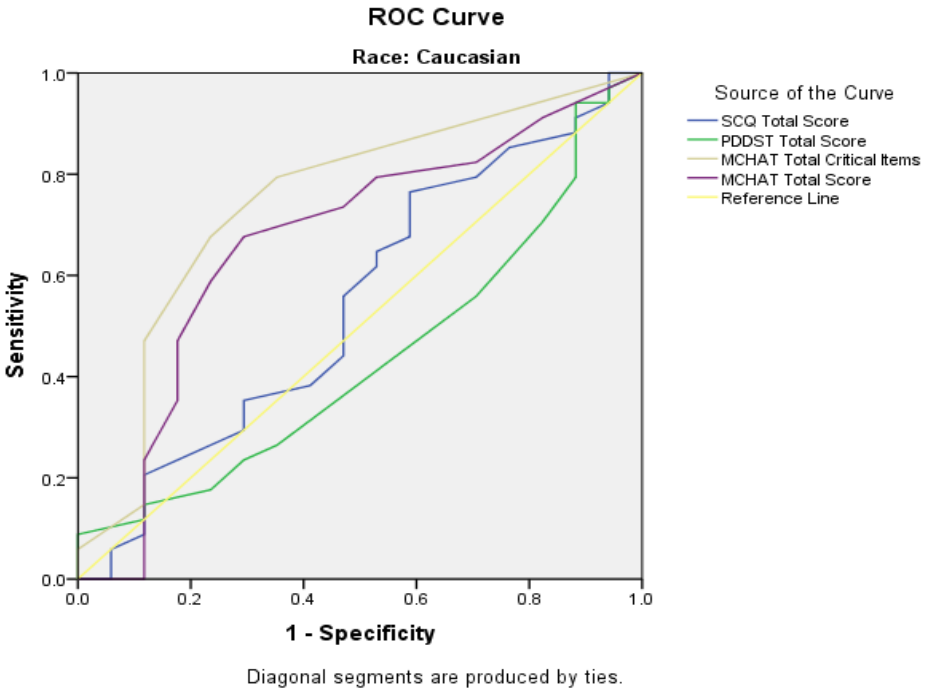


Figure 4

ROC Curves for M-CHAT, M-CHAT Critical Items, SCQ, and PDDST-II Total Scores, Other

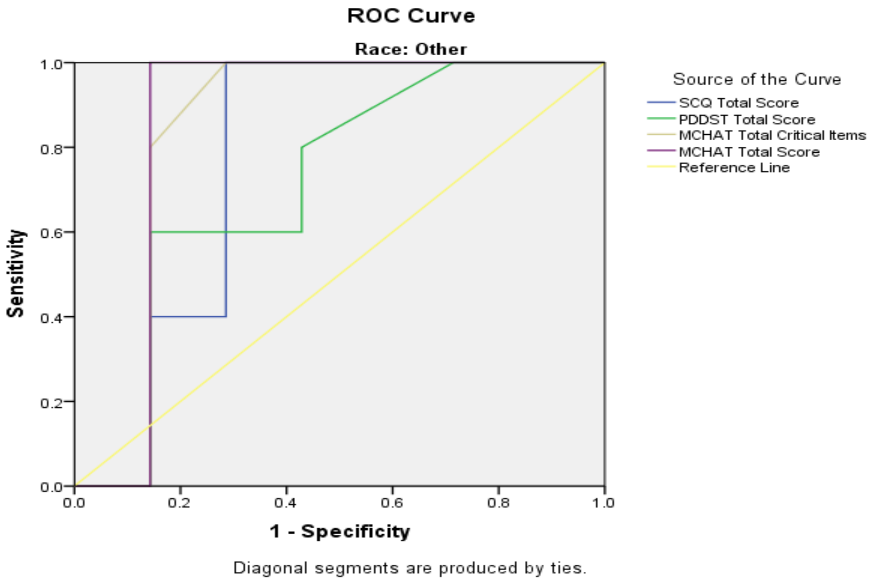


Figure 5

ROC Curves for M-CHAT, M-CHAT Critical Items, SCQ, and PDDST-II Total Scores, No High School

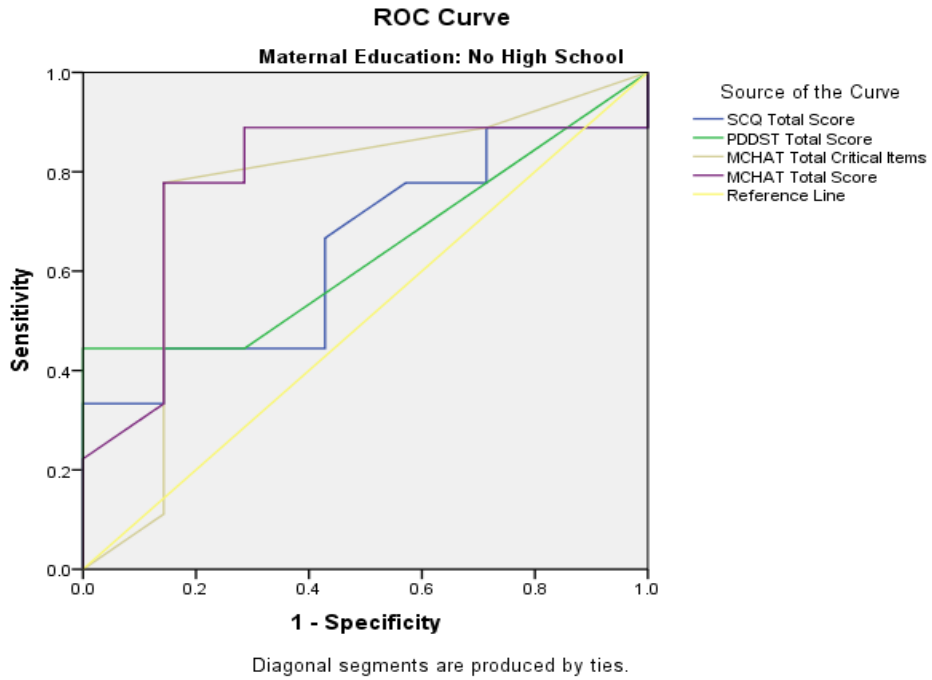


Figure 6

ROC Curves for M-CHAT, M-CHAT Critical Items, SCQ, and PDDST-II Total Scores, High School Graduate

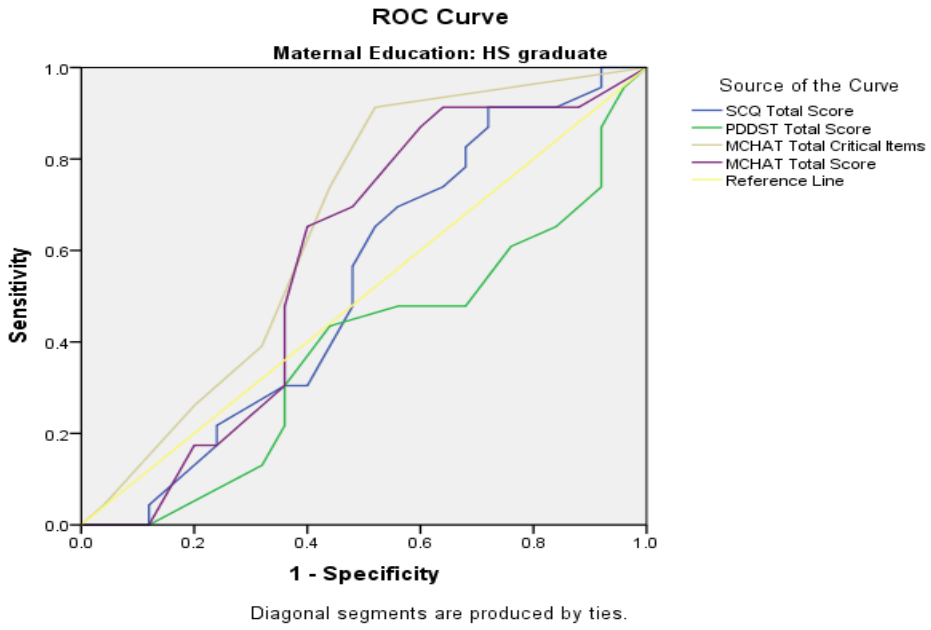
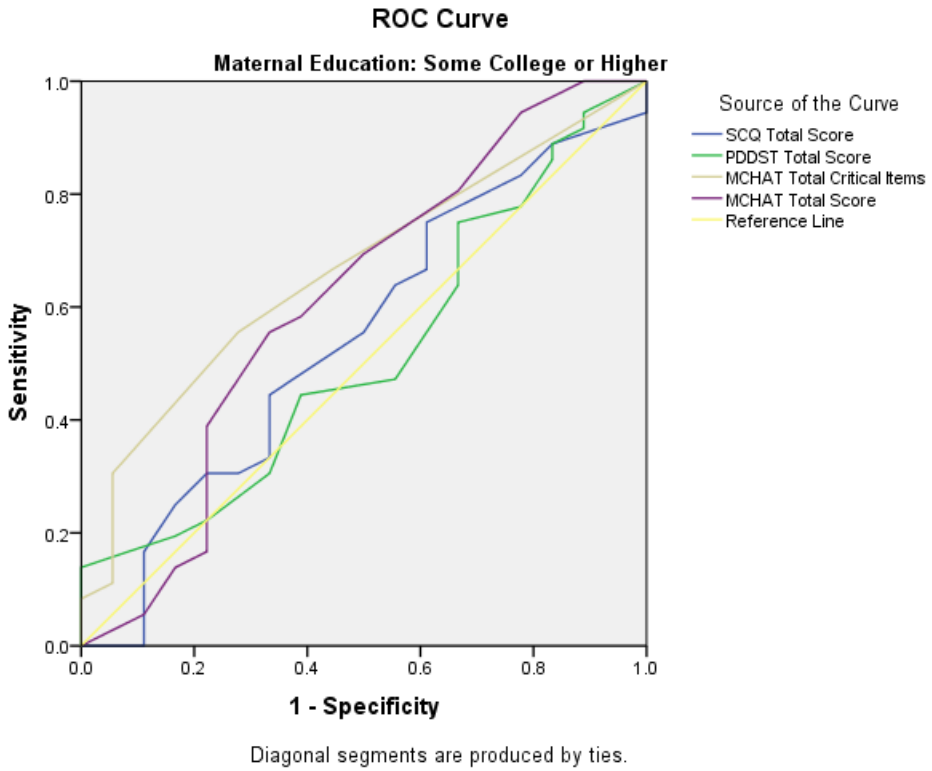


Figure 7
ROC Curves for M-CHAT, SCQ, and PDDST-II Total Scores, Some College or Higher



Screener performance was contrasted across groups. The researcher compared the AUCs produced by the screening instruments within each race and maternal education level group to determine if the screening instruments differed in diagnostic accuracy across the levels of each independent variable. Areas under the curve comparisons were computed using statistical comparisons described by Youngstrom (2014). The formula delineated by Youngstrom makes it possible to compare diagnostic performance between two samples by using the two AUC coefficients and their standard errors to perform *z*-tests. Six comparisons were computed for each screener (i.e., African-American /Caucasian, African-American/Other, Caucasian/Other, No High School/High School Graduate, No High School/Some College or Higher, High School Graduate/Some College or Higher). Results of *z*-test comparisons are presented in Table 8 with corresponding *p* values.

Table 7
Screener Areas Under the Curve (AUC) Across Groups

	AUC	Standard Error	95% CI
M-CHAT			
African-American	.53	0.08	.38 - .68
Caucasian	.67	0.09	.50 - .84
Other	.86	0.13	.60 – 1.0
No High School	.79	0.13	.54 – 1.0
High School	.58	0.09	.42 - .75
Some College or Higher	.61	0.09	.44 - .79
Total Sample	.61	0.05	.50 - .72
M-CHAT Critical			
African-American	.57	0.08	.42 - .72
Caucasian	.75	0.08	.60 - .90
Other	.84	0.13	.58 – 1.0
No High School	.75	0.14	.48 – 1.0
High School	.66	0.08	.50 - .82
Some College or Higher	.67	0.08	.52 - .81
Total Sample	.66	0.05	.56 - .76
PDDST-II			
African-American	.44	0.08	.29 - .59
Caucasian	.44	0.09	.27 - .60
Other	.71	0.16	.41 – 1.0
No High School	.64	0.14	.37 - .92
High School	.39	0.08	.23 - .55
Some College or Higher	.52	0.08	.35 - .68
Total Sample	.46	0.05	.35 - .56
SCQ			
African-American	.49	0.08	.33 - .64
Caucasian	.55	0.09	.37 - .72
Other	.77	0.15	.48 – 1.0
No High School	.64	0.14	.37 - .92
High School	.53	0.09	.36 - .69
Some College or Higher	.55	0.09	.38 - .71
Total Sample	.53	0.06	.42 - .64

Note. CI = confidence interval, AUC = area under the curve

Table 8
Screener Accuracy Comparisons across Race and Maternal Education

Screener/Contrast	z	*p*
M-CHAT		
African-American/Caucasian	1.21	.23
African-American/Other	2.14*	.03
Caucasian/Other	1.20	.23
No High School/High School Graduate	1.32	.19
No High School/Some College or Higher	1.11	.27
High School Graduate/Some College or Higher	0.24	.81
M-CHAT Critical Items		
African-American/Caucasian	1.66	.08
African-American/Other	1.80	.07
Caucasian/Other	0.62	.53
No High School/High School Graduate	0.60	.55
No High School/Some College or Higher	0.56	.58
High School Graduate/Some College or Higher	0.08	.94
PDDST-II		
African-American/Caucasian	0.02	.99
African-American/Other	1.59	.11
Caucasian/Other	1.57	.12
No High School/High School Graduate	1.57	.12
No High School/Some College or Higher	0.77	.44
High School Graduate/ Some College or Higher	1.10	.27
SCQ		
African-American/Caucasian	0.49	.62
African-American/Other	1.69	.09
Caucasian/Other	1.30	.19
No High School/High School Graduate	0.71	.48
No High School/Some College or Higher	0.58	.56
High School Graduate/Some College or Higher	0.18	.85

Note. *****indicates significance at the $p < .05$ level; M-CHAT = Modified Checklist for Autism in Toddlers, PDDST-II = Pervasive Developmental Disorders Screening Test, Second Edition; SCQ = Social Communication Questionnaire.

Table 9

Positive and Negative Screens at Established Cut-Off Scores for Each Screener

	Positive Screens	Negative Screens
Total Sample (*N* = 121)		
M-CHAT Critical Only	62 (51.2%)	59 (48.8%)
M-CHAT	97 (80.2%)	24 (19.8%)
PDDST-II	108 (89.3%)	13 (10.7%)
SCQ	77 (63.6%)	44 (36.4%)
African-American (*n* = 58)		
M-CHAT Critical Only	30 (51.7%)	28 (48.3%)
M-CHAT	49 (84.5%)	9 (15.5%)
PDDST-II	52 (89.7%)	6 (10.3%)
SCQ	42 (72.4%)	16 (27.6%)
Caucasian (*n* = 51)		
M-CHAT Critical Only	27 (52.9%)	24 (47.1%)
M-CHAT	40 (78.4%)	11 (21.6%)
PDDST-II	45 (88.2%)	6 (11.8%)
SCQ	30 (58.8%)	21 (41.2%)
Other (*n* = 12)		
M-CHAT Critical Only	5 (41.7%)	7 (58.3%)
M-CHAT	8 (66.7%)	4 (33.3%)
PDDST-II	11 (91.7%)	1 (8.3%)
SCQ	5 (41.7%)	7 (58.3%)
No High School (*n* = 16)		
M-CHAT Critical Only	8 (50.0%)	8 (50.0%)
M-CHAT	12 (75.0%)	4 (25%)
PDDST-II	16 (100.0%)	0
SCQ	11 (68.8%)	5 (31.3%)
High School Graduate (*n* = 48)		
M-CHAT Critical Only	28 (58.3%)	20 (41.7%)
M-CHAT	41 (85.4%)	7 (14.6%)
PDDST-II	43 (89.6%)	5 (10.4%)
SCQ	33 (68.8%)	15 (31.3%)
Some College or Higher (*n* = 54)		
M-CHAT Critical Only	25 (46.3%)	29 (53.7%)
M-CHAT	41 (75.9%)	13 (24.1%)
PDDST-II	46 (85.2%)	8 (14.8%)
SCQ	30 (55.6%)	24 (44.4%)

Note. M-CHAT = Modified Checklist for Autism in Toddlers, PDDST-II = Pervasive Developmental Disorders Screening Test, Second Edition; SCQ = Social Communication Questionnaire.

Comparing results of AUCs for the M-CHAT between race groups resulted in a significant difference in diagnostic accuracy for the M-CHAT across participants identified as African-American or Other in that the M-CHAT demonstrated significantly better diagnostic accuracy for participants in the Other race category when compared to those identified as African-American. Neither of the other two race group comparisons for the M-CHAT was statistically significant. Similarly, no significant differences emerged for pairwise race group AUC comparisons for PDDST-II and SCQ total scores. Areas under the curve comparisons were also conducted to determine whether diagnostic accuracy differed across maternal education level groups. The comparisons between No High School, High School Graduate, and Some College or Higher maternal education level resulted in no significant differences in AUC values across each of the three screeners.

Sensitivity and specificity as well as the percentage of positive and negative screens for each screener are presented in Appendix C and Table 9 for both the total sample and target demographic groups. Utilizing the Cicchetti et al. (1995) criteria for clinical significance (i.e., <.70 = poor, .70 - .79 = fair, .80 - .89 = good, .90 – 1.00 = excellent), none of the screeners demonstrated acceptable diagnostic accuracy for both sensitivity and specificity. Screener accuracy estimates fell above the poor range for both sensitivity and specificity within the two smallest subsamples (Other Race (n = 12); No High School (n = 16)). Specifically, in the Other Race subsample, specificity fell in the good range while sensitivity fell in the excellent range for the M-CHAT at a cut-off score of 5 but not at the suggested cut-off score of 3. In the same subsample, M-CHAT critical item total resulted in excellent sensitivity and fair specificity at a cut-off score of 1 and good sensitivity and specificity at the suggested cut-off of 2. Within the Other Race subsample, the SCQ demonstrated excellent sensitivity and fair specificity at a cut-

off score of 13 but not at the suggested cut-off. The M-CHAT also demonstrated fair sensitivity and specificity at a cut-off of 5, while the M-CHAT critical item total demonstrated fair sensitivity and good specificity at the suggested cut-off of 2.

Additionally, AUCs were compared for each of four screener total scores to determine if the instruments differed in diagnostic accuracy within the entire sample and then within demographic groups. As pointed out by Youngstrom (2014), it is important to conduct statistical comparisons of AUCs rather than simply examining confidence intervals given that small differences in AUCs can be statistically significant when the predictors are correlated. Likewise, comparing different screening instruments within the same sample provides strong evidence for use of one instrument versus the other. Areas under the curve were compared using Hanley and McNeil's (1983) methodology, which allows for AUC comparisons derived from the same individual in pairwise fashion, correcting for correlations that exist between AUCs. Areas under the curve are compared through *z* tests using the following formula (Hanley & McNeil, 1983):

$$z = AUC_1 - AUC_2 \,/\, [(SE_1)^2 + (SE_2)^2 - 2rSE_1SE_2]^{1/2}$$

where AUC_1 and SE_1 refer to the observed area and estimated standard error for test 1; AUC_2 and SE_2 refer to the observed area and estimated standard error for test 2, and *r* references the correlation between AUC_1 and AUC_2. The Hanley and McNeil method compares the areas under the curve for two measures applied to the same individual. The method requires calculating the correlation between AUC values for both the positive and negative diagnostic groups. The average of the two correlations is included in the equation provided by Hanley and McNeil. Hanley and McNeil provide a table that combines the average correlation between the two measures with the average area of the curve for the two measures to determine the correlation coefficient between the two AUCs. Areas under the curve comparisons were

computed using the pROC macro for MedCalc version 15.2. For the total sample and each subsample, AUC comparisons were computed for all possible combinations of screener total scores including M-CHAT, M-CHAT Critical Items, PDDST-II, and SCQ-Current. Pairwise comparisons of screeners for the entire sample as well as race and maternal education subsamples are listed in Table 10 and Appendices D and E.

When examining AUC comparisons for the total sample, the M-CHAT Total Score and the M-CHAT Critical Item Total Score predicted final diagnosis (i.e., ASD or non-ASD) significantly better than the SCQ-Current. Likewise, the M-CHAT Critical Item Total Score predicted final diagnosis significantly better than the PDDST-II. However, no other comparisons in the total sample produced significant results, suggesting that the screeners did not perform significantly better in differentiating ASD from non-ASD across the entire sample.

Table 10
Pairwise Comparisons of AUCs for Total Sample (N = 121)

Contrast	AUC Difference	Standard Error	z	*p*
M-CHAT/ M-CHAT Critical	0.05	0.03	1.57	.12
M-CHAT/ PDDST-II	0.07	0.05	1.36	.17
M-CHAT/ SCQ	0.08	0.03	2.29*	.02
M-CHAT Critical/ PDDST-II	0.12	0.06	2.14*	.03
M-CHAT Critical/ SCQ	0.13	0.04	3.08*	.002
PDDST-II/ SCQ	0.01	0.05	0.22	.82

Note. *indicates significance at the $p < .05$ level

With respect to screener AUC comparisons within race groups, no significant differences were observed in pairwise comparisons within the African-American and Other Race subsamples. Such findings suggest that no screener performed significantly better than another in differentiating individuals with ASD from those without ASD in the African-American and Other Race subsamples. In contrast, pairwise comparisons of screeners within the Caucasian subsample revealed significant differences. Specifically, the M-CHAT Critical Item Total Score performed significantly better in differentiating ASD from non-ASD when compared with the M-CHAT and the SCQ. Likewise, the M-CHAT performed significantly better than the SCQ.

Within maternal education groups, no significant differences between screeners were found in the No High School group. In contrast, within the High School Graduate group, the M-CHAT Critical Item Total Score performed significantly better than the SCQ in differentiating ASD from non-ASD participants. Within the Some College or Higher group, the M-CHAT Critical Item Total Score performed significantly better than both the PDDST-II and the SCQ.

Given that there is evidence to support differences in age of identification of ASD across racial groups, the researcher sought to determine if mean chronological age at first autism evaluation within a specialty care setting differed across race groups. A one-way analysis of variance was employed to test for significant differences in chronological age across race groups. The F statistic approached significance, $F(2, 118) = 3.37$, $p = .08$; however, differences were not statistically significant at the $p < .05$ level. An additional one-way ANOVA was employed to test for significant differences in chronological age across maternal education levels. Differences were not statistically significant, $F(2, 118) = 1.91$, $p = .15$.

Table 11
Analysis of Variance (ANOVA) of Mean Differences in Chronological Age Across Levels of Maternal Education and Race

Factor	*df*	*F*	η^2	*p*
Race	2	3.37	.04	.08
Maternal Education	2	1.91	.03	.15

Note. df = degrees of freedom, η^2= p

CHAPTER FIVE

DISCUSSION

With prevalence studies now estimating that 1 in 68 children in the United States has been diagnosed with an ASD by age 8, research examining ASD diagnostic tools is vital, as improved tools may lead to earlier identification (CDC, 2014). Early identification and intervention for young children with ASD has been shown to improve social, communicative, and cognitive skills (Osterling et al., 2002). Despite evidence to support improved outcomes with early intervention, ASD is still being diagnosed, on average, much later than the time at which it can be reliably diagnosed. Additionally, research suggests that differences exist between minority racial and SES groups with regard to age at first diagnosis of ASD. Population-wide screening for ASD has been recommended to aid in early identification as well as to reduce disparities in timing of diagnosis, particularly for underserved groups.

The present study aimed to examine the cross-cultural validity of three early ASD screening instruments developed for use with preschool-aged children to inform the validity of use of these instruments with diverse populations. First, the researcher sought to determine whether caregivers from varying racial/ethnic and educational backgrounds would rate their children's behavior differently on standardized ASD screeners. Additionally, the researcher examined differences in screener total scores across race to determine if any observed differences could be due to maternal education level. Next, the researcher evaluated the psychometric properties of the screeners in the total sample as well as within demographic subgroups. Sensitivity, specificity and areas under the curve (AUCs) were examined for each

screener. Likewise, the researcher compared performance of the screeners between demographic groups and directly compared screener performance within the total sample and separate demographic groups. Finally, chronological age at the time of the ASD evaluation within the specialty care clinic setting was evaluated across race and maternal education level to determine if significant age differences existed.

An unexpected finding of the current study was that no significant differences in M-CHAT, PDDST-II, or SCQ total scores were found between diagnostic groups (i.e., ASD vs. non-ASD). However, a significant difference was found between groups for the M-CHAT Critical Item Total Score. It would be expected that mean total scores would differ across these groups if screeners are adequate to differentiate between those with and without the disorder. Specifically, it would be expected that mean total scores for each screener would be significantly higher for ratings of participants ultimately diagnosed with ASD; however, results suggest no differences in caregiver ratings of child behavior between those children who were diagnosed with ASD and those who were not.

It should be noted that, as screeners, the measures may not be adequate indicators of severity. The analyses employed assumed that a higher total score was indicative of a "more positive" result; however, the screeners were not designed as measures of ASD severity. Instead, the screeners were developed to detect "hits" and "misses" as measured by total scores above or below the determined cut-off. It should also be noted that, by using a "broad" operational definition of autism (i.e., including all DSM-IV-TR PDDs rather than Autistic Disorder alone), total raw scores across screeners may have been lower than if a more "narrow" definition of autism were used. However, using a broad definition of autism allows for detecting all severity

levels of a disorder (Youngstrom, 2014). Nonetheless, it was expected that mean screener total scores would differ significantly for those participants with and without ASD.

Comparison of intellectual functioning between diagnostic groups revealed significantly lower cognitive standard scores for participants diagnosed with ASD when compared to those in the non-ASD group. Such a finding suggests that participants with general developmental delay such as that associated with lower cognitive functioning may be more likely to screen positive for ASD when compared to participants with higher cognitive scores. However, screener total score was not significantly correlated with cognitive standard score for any of the three screeners.

Even though previous analyses indicated that screener total scores did not differ across race or maternal education level, a two-way ANOVA was employed to determine if there was an interaction effect between race and maternal education. No main effects for race or maternal education level were found. Additionally, no significant interaction effect was observed. It should be noted that the findings of no significant main effects or interaction could be due to inadequate statistical power. It is unclear whether the null hypothesis could not be rejected due to inadequate power or due to there being no significant differences in mean screener total scores across race and maternal education groups.

Screener accuracy was examined across race and maternal education levels to determine if screeners performed differently across groups. The ability of each screener to predict a dichotomous diagnostic outcome (i.e., ASD vs. non-ASD final diagnosis) was examined. Receiver operating curve analysis revealed that few AUC estimates fell above chance levels for each of the screeners for the present sample. Likewise, few screener AUC estimates for the total sample and within subgroups were significant at the $p < .05$ level.

Next, the researcher compared the AUCs produced by the three screening instruments within each race and maternal education level group to determine if the screening instruments differed in diagnostic accuracy across the levels of each independent variable. No significant differences in diagnostic accuracy across maternal education levels for each screener were observed. Likewise, no significant differences in diagnostic accuracy across race groups for the PDDST-II and SCQ were observed. However, a significant difference in AUC estimates for M-CHAT total scores between those identified as African-American and those in the Other race group. This result suggests that the diagnostic accuracy of the M-CHAT was significantly better for those identified in the Other race group when compared to those identified as African-American.

With respect to sensitivity and specificity, none of the screeners demonstrated acceptable diagnostic accuracy for both sensitivity and specificity. Screener accuracy estimates fell above the poor range for both sensitivity and specificity within the two smallest subsamples (Other Race (n = 12); No High School (n = 16)). Specifically, in the Other Race subsample, specificity fell in the good range while sensitivity fell in the excellent range for the M-CHAT at a cut-off score of 5 but not at the suggested cut-off score of 3. In the same subsample, M-CHAT critical item total resulted in excellent sensitivity and fair specificity at a cut-off score of 1 and good sensitivity and specificity at the suggested cut-off of 2. Within the Other Race subsample, the SCQ demonstrated excellent sensitivity and fair specificity at a cut-off score of 13 but not at the suggested cut-off. The M-CHAT also demonstrated fair sensitivity and specificity at a cut-off of 5, while the M-CHAT critical item total demonstrated fair sensitivity and good specificity at the suggested cut-off of 2.

When examining AUC comparisons for the total sample, the M-CHAT Total Score and the M-CHAT Critical Item Total Score predicted final diagnosis (i.e., ASD or non-ASD) significantly better than the SCQ-Current. However, no other comparisons in the total sample produced significant results, suggesting that the screeners did not perform significantly better in differentiating ASD from non-ASD across the entire sample. No differences in screener performance were observed for the African-American or Other Race groups. For the Caucasian subsample, the M-CHAT Critical Item Total Score performed significantly better than the M-CHAT or the SCQ. Likewise, the M-CHAT performed significantly better than the SCQ in this subgroup. Comparing results of AUCs for the M-CHAT between race groups resulted in a significant difference in diagnostic accuracy for the M-CHAT across participants identified as African-American or Other in that the M-CHAT demonstrated significantly better diagnostic accuracy for participants in the Other race category when compared to those identified as African-American.

Finally, given evidence to support disparities in age of first ASD evaluation for minority groups, the researcher sought to determine whether the age at the present ASD evaluation differed significantly across race and maternal education level. Results indicated no significant differences in age at the time of the evaluation.

Limitations

The results and conclusions of this study should be interpreted keeping in mind the following limitations. Generalization of the results of the current study may be limited. Results should be interpreted keeping in mind the difficulties inherent in utilizing a referred or clinical sample with assessment measures created for population screening. Each participant was referred to the BCDD for evaluation after having been determined to be “at-risk” for ASD. As

such, more "positive" screens were observed in the current sample than would be expected in a sample representative of the general population. However, authors of many widely used ASD screening or diagnostic tools (e.g., M-CHAT, SCQ) utilized referred samples in their standardization studies.

Another major limitation was the size of the sample utilized in the current study. Results should be interpreted keeping in mind that analyses were conducted despite limited statistical power to reject the null hypothesis. Likewise, response rates for items related to insurance provider on the Demographic Information Form precluded the researcher from conducting proposed analyses. Though still a proxy, Medicaid status would have been a more accurate measure of SES than maternal education level. Similarly, important variables such as parent reading level, primary language spoken, or prior knowledge of ASDs that may affect parent interpretation and comprehension of screening items were not investigated. Additionally, estimates of cognitive functioning may have been affected by the use of different measures (i.e., Bayley-III, SB-5) across participants. Additionally, final diagnoses were heavily based on the ADOS. Although the ADOS is considered a gold standard assessment tool for ASD diagnosis, the ADOS has not been well-validated across diverse samples. As such, the validity of ADOS results may be questionable for participants from minority backgrounds.

Clinical Implications

In spite of the major limitations of the current study, investigation of three widely-used ASD screening instruments reveals that such instruments need improvement in their ability to discriminate between ASD and non-ASD participants in a clinically referred sample at-risk for ASD. Although it is nearly impossible to obtain perfect diagnostic accuracy given the inherent error in the procedures for diagnosing ASD, improved screener performance is warranted.

The aim of the current study was to compare screener accuracy across subgroups. With the exception of a significant difference in accuracy for the M-CHAT between African-American and Other Race participants, no significant differences were observed between demographic subgroups on any screener. This finding may suggest that the screeners perform generally the same across race and education level groups. Additionally, pairwise comparisons of the screeners within the total sample and within each subsample indicate that no significant differences in diagnostic accuracy exist within the African-American, Other Race, and No High School groups. Significant differences in diagnostic accuracy were observed within the High School Graduate, Some College or Higher, and Caucasian groups with the M-CHAT and M-CHAT Critical Item Total Score generally performing better than the SCQ and PDDST-II. Such differences in screener performance across cultural groups are limited in generalizability due to the limitations of the current sample.

In theory, directly comparing the psychometric properties of screeners within the same sample has the potential to inform use of one instrument over the other in clinical practice. In the current sample, none of the screeners performed well across all accuracy standards. However, the M-CHAT and M-CHAT Critical Item Total Score were better able to discriminate ASD from non-ASD based on measures of sensitivity and specificity as well as in direct comparisons with the SCQ.

Directions for Future Research

It is clear that there is a need for improved assessment practices when evaluating individuals with diverse backgrounds for ASD. Future studies may look to evaluate whether a combination of screening instruments improves diagnostic accuracy. Additionally, future research should focus on identifying which ASD assessment tools are the least culturally biased.

As new ASD assessment instruments are developed, researchers should ensure that instruments are standardized with diverse samples and that results of standardization with diverse populations are explicitly described in published manuals and standardization studies.

REFERENCES

American Academy of Pediatrics (2006). Identifying infants and young children with developmental disorders in the medical home: An algorithm for developmental surveillance and screening. *Pediatrics, 108*, 405-420.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing.* Washington, D.C.: Author.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders, fourth edition, text revision.* Washington, DC: American Psychiatric Association.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, fifth edition.* Arlington, VA: American Psychiatric Association.

American Psychological Association (1993). Guidelines for providers of psychological services to ethnic, linguistic, and culturally diverse populations. *American Psychologist, 48*, 45-48.

Arunyanart, W., Fenick, A., Ukritchon, S., Imjajitt, W., Northrup, V., & Weitzman, C. (2012). Developmental and autism screening: A survey across six states. *Infants and Young Children, 25*, 175-187.

Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *British Journal of Psychiatry, 161*, 839-843.

Barton, M. L., Dumont-Mathieu, T., & Fein, D. (2012). Screening young children for autism spectrum disorders in primary practice. *Journal of Autism and Developmental Disorders, 42*, 1165-1174.

Bayley, N. (2006). Bayley Scales of Infant and Toddler Development – Third Edition. San Antonio, TX: Harcourt Assessments, Inc.

Blacher, J., Cohen, S. R., & Azad, G. (2014). In the eye of the beholder: Reports of autism symptoms by Anglo and Latino mothers. *Research in Autism Spectrum Disorders, 8*¸1648-1656.

Begeer, S., El Bouk, S., Boussaid, W., Terwogt, M. M., & Koot, H. M. (2009). Underdiagnosis and referral bias of autism in ethnic minorities. *Journal of Autism and Developmental Disorder, 39*, 142-148.

Bertrand, J., Mars, A., Boyle, C., Bove, F., Yeargin-Allsopp, M., & Decoufle, P. (2001). Prevalence of autism in a United States population: The Brick Township, New Jersey, investigation. *Pediatrics, 108*, 1155-1161.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *The British Journal of Psychiatry, 175*, 444-451.

Bhasin, T. K., & Schendel, D. (2007). Sociodemographic risk factors for autism in a US metropolitan area. *Journal of Autism and Developmental Disorders, 37*, 667-677.

Bolte, S., Holtmann, M., & Poustka, F. (2008). The social communication questionnaire (SCQ) as a screener for autism spectrum disorders: Additional evidence and cross-cultural validity. *Journal of the American Academy of Child and Adolescent Psychiatry, 47*, 719-720.

Centers for Disease Control and Prevention. (2012). Prevalence of autism spectrum disorders – autism and developmental disabilities monitoring network. *Morbidity and Mortality Weekly Report, 61*.

Centers for Disease Control and Prevention (2014). Prevalence of autism spectrum disorders among children aged 8 years – autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *Morbidity and Mortality Weekly Report, 63*(2).

Chandler, S., Charman, T., Baird, G., Simonoff, E., Loucas, T., Meldrum, D., . . . Pickles, A. (2007). Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 1324-1332.

Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., . . . Pickles, A. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *British Journal of Psychiatry, 191,* 554-449.

Chawarska, K., Klin, A., Paul, R., & Volkmar, F. (2007). Autism spectrum disorder in the second year: Stability and change in syndrome expression. *Journal of Child Psychology* and Psychiatry, 48, 128-138.

Cho, S., Singer, G. H., & Brenner, B. (2003). A comparison of adaptation to childhood disability in Korean immigrant and Korean mothers, *Focus on Autism and Other Developmental Disabilities, 18,* 9-19.

Chuan-Yu, C., Liu, C. Y., Su, W. C., Huang, S. L., & Lin, K. M. (2008). Urbanicity-related variation in help-seeking and services utilization among preschool-age children with autism in Taiwan. *Journal of Autism and Developmental Disorders, 38*, 489-497.

Chung, K. M. et al. (2012). Cross cultural differences in challenging behaviors of children with autism spectrum disorders: An international examination between Israel, South Korea, the United Kingdom, and the United States of America. *Research in Autism Spectrum Disorders, 6*, 881-889.

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

Coonrod, E., & Stone, W. (2004). Early concerns of parents of children with autistic and nonautistic disorders. *Infants and Young Children, 17*, 258-269.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., . . . Lord, C. (2007). Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, Allied Disciplines, 48*, 932-940.

Cuccaro, M. L., Brinkley, J., Abramson, R. K., Hall, A., Wright, H. H., Hussman, J. P., . . . Pericak-Vance, M. A. (2007). Autism in African American families: Clinical-phenotypic findings. *American Journal of Medical Genetics Part B, 144B*, 1022-1026.

Cuccaro, M. L., Wright, H. H., Rownd, C. V., & Abramson, R. K. (1996). Brief report: Professional perceptions of children with developmental difficulties: The influence of race and socioeconomic status. *Journal of Autism and Developmental Disorders, 26*, 461-469.

Daley, T. C. (2002). The need for cross-cultural research on the pervasive developmental disorders. *Transcultural Psychiatry, 39*, 531-550.

Daniels, A. M., & Mandell, D. S. (2013). Children's compliance with American Academy of Pediatrics' well-child care visit guidelines and the early detection of autism. *Journal of Autism and Developmental Disorders, 43,* 2844-2854.

Dawson, G. (2008). Early behavioral intervention, brain plasticity, and the prevention of autism spectrum disorder. *Development and Psychopathology, 20,* 775-803.

De Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Volkmar, F., & Minderaa, R. (2003). Measuring pervasive developmental disorders in children and adolescents with mental retardation: A comparison of two screening instruments used in a study of the total mentally retarded population from a designated area. *Journal of Autism and Developmental Disorders, 33,* 595-605.

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

Dumont-Mathieu, T. & Fein, D. (2005). Screening for autism in young children: The modified checklist for autism in toddlers (M-CHAT) and other measures. *Mental Retardation and Developmental Disabilities, 11,* 253-262.

Durkin, M. S., Maenner, M. J., Meaney, F. J., Levy, S. E., DiGuiseppi, C., Nicholas, J. S., . . . Schieve, L. A. (2010). Socioeconomic inequality in the prevalence of autism spectrum disorder: Evidence from a U.S. cross-sectional study. *PLoS ONE, 5,* e11551.

Eaves, L. C., Wingert, H., & Ho, H. H. (2006a). Screening for autism: Agreement with diagnosis. *Autism, 10,* 229-242.

Eaves, L. C., Wingert, H. D., Ho, H. H., & Mickelson, E. C. (2006b). Screening for autism spectrum disorders with the social communication questionnaire. *Journal of Developmental and Behavioral Pediatrics,* 27, S95-S103.

Fenske, E. C., Zalenski, S., Krantz, P. J., & McClannahan, L. E. (1985). Age at intervention and treatment outcome for autistic children in a comprehensive intervention program. *Analysis and Intervention in Developmental Disabilities, 5,* 49-58.

Fombonne, E. (2003). The prevalence of autism. *Journal of the American Medical Association, 289*, 87-89.

Garland-Daniels, S. A. (2002). Early screening: Can autism be identified in young children through parental report? Unpublished manuscript, University of Alabama, Tuscaloosa, Alabama. Retrieved from https://www.researchgate.net/publication/34511355 _Early_screening__can_autism_be_identified_in_young_children_through_parental_ report.

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*, 107-116.

Goin-Kochel, R. P., & Cohen, R. (2008). Screening cases within a statewide autism registry: A comparison of parent reports using DSM-IV-TR criteria versus the SCQ. *Focus on Autism and Other Developmental Disabilities, 23*, 148-154.

Goin-Kochel, R. P., Mackintosh, V. H., & Myers, B. J. (2006). How many doctors does it take to make an autism spectrum diagnosis? *Autism, 10,* 439-451.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology, 148*, 839-843.

Harris, B., Barton, E. B., & Albert, C. (2014). Evaluation autism diagnostic and screening tools for cultural and linguistic responsiveness. *Journal of Autism and Developmental Disorders, 44*, 1275-1287.

Hartley-McAndrew, M., Doody, K. R., & Mertz, J. (2014). Knowledge of autism spectrum disorders in potential first-contact professionals. *North American Journal of Medicine and Science, 7*¸ 97-102.

Hilton, C. L., Fitzgerald, R. T., Jackson, K. M., Maxim, R. A., Bosworth, C. C., Shattuck, P. T., . . . Constantino, J. N. (2010). Brief report: Under-representation of African Americans in autism genetic research: A rationale for inclusion of subjects representing diverse family structures. *Journal of Autism and Developmental Disorders, 40*, 633-639.

Hunsley, J., & Mash, E. J. (2007). Evidence-based assessment. *Annual Review of Clinical Psychology, 3*, 29-51.

Inada, N., Koyama, T., Inokuchi, E., Kuroda, M., & Kamio, Y. (2011). Reliability and validity of the Japanese version of the modified checklist for autism in toddlers (M-CHAT). *Research in Autism Spectrum Disorders, 5*, 330-336.

Jacobson, J. W., Mulick, J. A., & Green, G. (1998). Cost-benefit estimates for early intensive behavioral intervention for young children with autism: General model and single state case. *Behavioral Interventions, 13*, 201-226.

Johnson, C. P., & Meyers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics, 120*, 1183-1215.

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

Kazdin, A. E. (2005). Evidence-based assessment for children and adolescents: Issues in measurement development and clinical applications. *Journal of Clinical Child and Adolescent Psychology, 34* 348-358.

Kleinman, J. M., Robins, D. L., Ventola, P. E., Pandey, J., Boorstein, H. C., Esser, E. L., . . . Fein, D. (2008). The Modified Checklist for Autism in Toddlers: A follow-up study investigating the early detection of autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*, 827-839.

Kolage, K. M., Smaldone, A. M., & Cohn, E. G. (2014). How will DSM-5 affect autism diagnosis? A systematic literature review and meta-analysis. *Journal of Autism and Developmental Disorders, 44*, 1918-1932.

Kozlowski, A. M., Matson, J. L., Worley, J. A., Sipes, M., & Horovitz, M. (2012). Defining characteristics for young children meeting cutoff on the modified checklist for autism in toddlers. *Research in Autism Spectrum Disorders, 6*, 472-479.

Lee, L., David, A. B., Rusyniak, J., Landa, R., & Newschaffer, C. J. (2007). Performance of the social communication questionnaire in children receiving preschool special education services. *Research in Autism Spectrum Disorders, 1*, 126-138.

Liptak, G. S., Benzoni, L. B., Mruzek, D. W., Nolan, K. W., Thingvoll, M. A., Wade, C. M., & Fryer, G. E. (2008). Disparities in diagnosis and access to health services for children with autism: Data from the national survey of children's health. *Journal of Developmental and Behavioral Pediatrics, 29*, 152-160.

Lord, C., Rutter, M., DiLavore, P., & Risi, S. (2001). Autism Diagnostic Observation Schedule (ADOS). Western Psychological Services.

Lord, C., Rutter, M., & Lecouteur, A. (1994). Autism Diagnostic Interview-Revised – a revised version of the diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 30*, 205-223.

Losapio, M. F., & Ponde, M. P. (2008). Translation into Portuguese of the M-CHAT scale for early screening. *Revista de Psiquiatria do Rio Grande do Sul, 30*, 221-229.

Mandell, D. S., Ittenbach, R. F., Levy, S. E., & Pinto-Martin, J. A. (2007). Disparities in diagnoses received prior to a diagnosis of autism spectrum disorder. *Journal of Autism and Developmental Disorders, 37*, 1795-1802.

Mandell, D. S., Listerud, J., Levy, S. E., & Pinto-Martin, J. A. (2002). Race differences in the age at diagnosis among Medicaid-eligible children with autism. *Journal of the American Academy of Child and Adolescent Psychiatry, 41*, 1447-1453.

Mandell, D. S., Morales, K. H., Xie, M., Lawer, L. J., Stahmer, A. C., & Marcus, S. C. (2010). Age of diagnosis among Medicaid-enrolled children with autism, 2001-2004, *Psychiatric Services, 61*, 822-829.

Mandell, D. S., & Novak, M. (2005). The role of culture in families' treatment decisions for children with autism spectrum disorders. *Mental Retardation and Developmental Disabilities Research Reviews, 11*, 110-115.

Mandell, D. S., Novak, M. M., & Zubritsky, C. D. (2005). Factors associated with age of diagnosis among children with autism spectrum disorders. *Pediatrics, 116*, 1480-1486.

Mandell, D. S., Wiggins, L. D., Carpenter, L. A., Daniels, J., DiGuiseppi, C., Durkin, M. S., . . . Kirby, R. S. (2009). Racial/ethnic disparities in the identification of children with autism spectrum disorders. *American Journal of Public Health, 99*, 493-498.

Matson, J. L., Worley, J. A., Fodstad, J. C., Chung, K. M., Suh, D., Jhin, H. K., . . . Furniss, F. (2011). A multinational study examining the cross cultural differences in reported symptoms of autism spectrum disorders: Israel, South Korea, the United Kingdom, and the United States of America. *Research in Autism Spectrum Disorders, 5*, 1598-1604.

Matson, J. L., Worley, J. A., Kozlowski, A. M., Chung, K. M., Jung, W., Yang, J. W. (2012). Cross cultural differences of parent reported social skills in children with autistic disorder: An examination between South Korea and the United States of America. *Research in Autism Spectrum Disorders, 6*, 971-977.

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

Mulligan, A., Richardson, T., Anney, R. J., & Gill, M. (2009). The social communication questionnaire in a sample of the general population of school-going children. *Irish Journal of Medical Science, 178*, 193-199.

Osterling, J., Dawson, G., & Munson, J. (2002). Early recognition of one year old infants with autism spectrum disorder versus mental retardation: A study of first birthday party home videotapes. *Developmental and Psychopathology, 14*, 239-252.

Oosterling, I. J., Rommelse, N., de Jonge, M., van der Gaag, R. J., Swinkels, S., Roos, S., . . . Buitelaar, J. (2010). How useful is the Social Communication Questionnaire in toddlers at risk of autism spectrum disorder? *Journal of Child Psychology and Psychiatry, 51*, 1260-1268.

Oosterling, I. J., Swinkels, S. H., van der Gaag, R. J., Visser, J. C., Dietz, C., & Ruitelaar, J. K. (2009). Comparative analysis of three screening instruments for autism spectrum disorders in toddlers at high risk. *Journal of Autism and Developmental Disorders, 39*, 897-909.

Pandey, J., Verbalis, A., Robins, D. L., Boorstein, H., Klin, A., Babitz, T., Chawarska, K., . . . Fein, D. (2008). Screening for autism in older and younger toddlers with the modified checklist for autism in toddlers. *Autism, 12*, 513-535.

Perera, H., Wijewardena, K., & Aluthwelage, R. (2009). Screening of 18-24-month-old children for autism in a semi-urban community in Sri Lanka. *Journal of Tropical Pediatrics, 55*, 402-405.

Pierce, K., Carter, C., Weinfeld, M., Desmond, J., Hazin, R., Bjork, R., & Gallagher, N. (2011). Detecting, studying, and treating autism early: The one-year well-baby check-up approach. *The Journal of Pediatrics, 159*, 458-465.

Pintea, S., & Moldovan, R. (2009). The receiver-operating characteristic (ROC) analysis: Fundamentals and applications in clinical psychology. *Journal of Cognitive and Behavioral Therapies, 9*, 49-66.

Pinto-Martin, J. A., Dunkle, M., Earls, M., Fliedner, D., & Landes, C. (2005). Developmental stages of developmental screening: Steps to implementation of a successful program. *American Journal of Public Health, 95*, 1928-1932.

Pinto-Martin, J. A., Young, L. M., Mandell, D. S., Poghosyan, L., Giarelli, E., & Levy, S. E. (2008). Screening strategies for autism spectrum disorders in primary care. *Journal of Developmental and Behavioral Pediatrics, 29*, 345-350.

Reijnevald, S. A., Harland, P., Brugman, E., Verhulst, F. C., & Verloove-Vanhorick, S. P. (2005). Psychosocial problems among immigrant and non-immigrant children: Ethnicity plays a role in their occurrence and identification. *European Journal of Child and Adolescent Psychiatry, 14*, 145-152.

Robins, D. L., Casagrande, K., Barton, M., Chen, C. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the Modified Checklist for Autism in Toddles, Revised with Follow-up (M-CHAT-R/F). *Pediatrics, 133*, 37-45.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders, 31*, 131-144.

Rogers, S. J., & Vismara, L. A. (2008). Evidence-based comprehensive treatments for early autism. *Journal of Clinical Child and Adolescent Psychology, 37*, 8-38.

Roid, G. H. (2003). Stanford-Binet Intelligence Scales – Fifth Edition. Itasca, IL: Riverside.

Rosenberg, R. E., Daniels, A. M., Law, J. K., & Kaufmann, W. E. (2009). Trends in autism spectrum disorder diagnoses: 1994-2007. *Journal of Autism and Developmental Disorders, 39*, 1099-1111.

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

Sattler, J. M. (2008). *Assessment of children: Cognitive Foundations*. La Mesa, CA: Jerome M. Sattler Publisher.

Saulnier, C. A., & Ventola, P. E. (2012). *Essentials of autism spectrum disorders evaluation and assessment.*Hoboken, New Jersey: Wiley and Sons, Inc.

Schanding, G. T., Nowell, K. P., & Goin-Kochel, R. P. (2012). Utility of the Social Communication Questionnaire – Current and Social Responsiveness Scale as teacher-report screening tools for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*, 1705-1716.

Schmidt, F. L. & Hunter, J. E. (1974). Racial and ethnic bias in psychological tests: Divergent implications of two definitions of test bias. *American Psychologist, 29*, 1-8.

Sell, N. K., Giarelli, E., Blum, N., Hanlon, A. L., & Levy, S. E. (2012). A comparison of autism spectrum disorder DSM-IV criteria and associated features among African American and white children in Philadelphia County. *Disability and Health Journal, 5,* 9-17.

Shattuck, P. T., Durkin, M., Maenner, M., Newschaffer, C., Mandell, D. S., Wiggins, L., . . . Cuniff, C. (2009). 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,* 474-483.

Siegel, B. (2004). *Pervasive Developmental Disorder Screening Test – II Manual*. San Antonio, TX: Harcourt Assessment, Inc.

Snow, A. V., & Lecavalier, L. (2008). Sensitivity and specificity of the Modified Checklist for Autism in Toddlers and the Social Communication Questionnaire in preschoolers suspected of having pervasive developmental disorders. *Autism, 12,* 677-694.

Thomas, K. C., Ellis, A. R., McLaurin, C., Daniels, J., & Morrissey, J. P. (2007). Access to care for autism-related services. *Journal of Autism and Developmental Disorders, 37,* 1902-1912.

Travers, J. C., Tincani, M., & Krezmien, M. P. (2011). A multiyear national profile of racial disparity in autism identification. *The Journal of Special Education, 47,* 41-49.

U.S. Census Bureau (2010). *State and county quick facts: Memphis, Tennessee.* Retrieved from http://quickfacts.census.gov/qfd/states/47/4748000.html.

Ventola, P., Kleinman, J., Pandey, P., Wilson, L., Esser, E., Boorstein, H., . . . Fein, D. (2007). Differentiating between autism spectrum disorders and other developmental disabilities in children who failed a screening instrument for ASD. *Journal of Autism and Developmental Disorders, 37,* 4325-4336.

Volkmar, F. R., Lord, C., Bailey, A., Schultz, R. T., & Klin, A. (2004). Autism and pervasive developmental disorders. *Journal of Child Psychology and Psychiatry, 45,* 135-170.

Wei, T., Chesnut, S. R., Barnard-Brak, L, & Richman, D. (2015). Psychometric analysis of the Social Communication Questionnaire using an item-response theory framework: Implications for use of the Lifetime and Current forms. *Journal of Psychopathology and Behavioral Assessment, 37*, 469-480.

Welterlin, A., & LaRue, R. H. (2007). Serving the needs of immigrant families of children with autism. *Disability and Society, 22*, 747-760.

Wiggins, L. D., Baio, J., & Rice, C. (2006). Examination of the time between first evaluation and first autism spectrum diagnosis in a population-based sample. *Developmental and Behavioral Pediatrics, 27*, 79-87.

Wiggins, L. D., Bakeman, R., Adamson, L. B., & Robins, D. L. (2007). Utility of the social communication questionnaire in screening for autism in children referred for early intervention. 34 *Focus on Autism and Other Developmental Disabilities, 22*, 33-38.

Williams, M. E., Atkins, M., & Soles, T. (2009). Assessment of autism in community settings: Discrepancies in classification. *Journal of Autism and Developmental Disorders, 39*, 660-669.

Witwer, A. N., & Lecavalier, L. (2007). Autism screening tools: An evaluation of the Social Communication Questionnaire and the Developmental Behaviour Checklist-Autism Screening Algorithm. *Journal of Intellectual and Developmental Disability, 32*, 179-187.

Wong, V., Hui, L. H. S., & Lee, W. C. (2004). A modified screening tool for autism for Chinese children. *Pediatrics, 114*, 166-176.

Yama, B., Freeman, T., Graves, E., Yuan, S., & Campbell, M. K. (2012). Examination of the properties of the modified checklist for autism in toddlers (M-CHAT) in a population sample. *Journal of Autism and Developmental Disorders, 42*, 23-34.

Yeargin-Allsopp, M., Rice, C., Karapurkar, T., Doernberg, N., Boyle, C., & Murphy, C. (2003). Prevalence of autism in a US metropolitan area. *Journal of the American Medical Association, 289*, 49-55.

Young, R. L., & Rodi, M. L. (2014). Redefining autism spectrum disorder using DSM-5: The implications of the proposed DSM-5 criteria for autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*, 758-765.

Youngstrom, E. A. (2014). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology, 39*, 204-221.

Zuckerman, K. E., Mattox, K. M., Sinche, B. K., Blaschke, G. S., & Bethell, C. (2014). Racial, ethnic, and language disparities in early childhood developmental/behavioral evaluations: A narrative review. *Clinical Pediatrics, 53*, 619-631.

APPENDICES

Appendix A

Description of Child Participants – Sample Sizes and Percentages (N = 121)

Demographic Variables	*M*	*SD*	*Range*	Sample Size (Percentage) No ASD (*n* = 51)	ASD (*n* = 70)
Age (years)					
ASD (*n* = 51)	3.74	1.14	1.75 - 6.08		
No ASD (*n* = 70)	4.03	1.17	1.92 – 6.33		
Cognitive ability					
ASD (*n* = 51)	65.03	14.76	40 – 110		
No ASD (*n* = 69)	77.67	16.00	48 - 109		
Sex					
Male				42 (34.7)	56 (46.3)
Female				9 (7.4)	14 (11.6)
Race					
African-American				27 (22.3)	31 (25.6)
Caucasian				17 (14.0)	34 (28.1)
Other				7 (5.8)	5 (4.1)
Insurance Type					
Medicaid				8 (6.6)	19 (15.7)
Private insurance				3 (2.5)	15 (12.4)
No response				40 (33.1)	36 (29.8)
ADOS					
Module 1				21 (17.4)	48 (39.7)
Module 2				24 (19.8)	17 (14.0)
Module 3				3 (2.5)	1 (0.8)
Final DSM-IV-TR Axis I Diagnosis					
Autistic Disorder				- -	44 (36.4)
Asperger's Disorder/ PDD-NOS				- -	26 (21.5)

Disruptive Behavior Disorder-NOS	16 (13.2)	- -
Mixed Developmental Disorder	13 (10.7)	- -
Mixed Receptive-Expressive Language Disorder	12 (9.9)	- -
Phonological Disorder	4 (3.3)	- -
Attention-Deficit/Hyperactivity Disorder	2 (1.7)	- -
Generalized Anxiety Disorder	2 (1.7)	- -
Expressive Language Disorder	1 (0.8)	- -
Oppositional Defiant Disorder	1 (0.8)	- -

Note. *M* = mean, *SD* = standard deviation

Appendix B
Description of Parents/Caregivers – Sample Sizes and Percentage

Demographic Variables	Sample Size (Percentage)	
	No ASD (n = 51)	ASD (n = 70)
Rater		
Mother	44 (36.4)	61 (50.4)
Father	3 (2.5)	1 (0.8)
Grandparent	1 (0.8)	3 (2.5)
Adoptive Mother	0 (0)	0 (0)
Aunt	2 (1.7)	0 (0)
Foster parent	1 (0.8)	0 (0)
Guardian	0 (0)	1 (0.8)
Both parents	0 (0)	1 (0.8)
Maternal Marital Status		
Married	8 (6.6)	24 (19.8)
Single	12 (9.9)	8 (6.6)
Divorced	2 (1.7)	0 (0)
Separated	2 (1.7)	0 (0)
No Response	27 (22.3)	38 (31.4)
Paternal Marital Status		
Married	19 (15.7)	41(33.9)
Single	12 (9.9)	10 (8.3)
Divorced	2 (1.7)	3 (2.5)
Separated	4 (3.3)	0 (0)
No Response	12 (9.9)	16 (13.2)
Maternal Race		
Caucasian	18 (14.9)	34 (28.1)
African-American	22 (18.2)	24 (19.8)
Other	2 (1.7)	3 (2.5)
No Response	9 (7.4)	9 (7.4)
Paternal Race		
Caucasian	15 (12.4)	30 (24.8)
African-American	15 (12.4)	24 (19.8)
Other	3 (2.5)	1 (0.8)
No Response	18 (14.9)	13 (10.7)
Maternal Education		
No High School	7 (5.8)	9 (7.4)
High School Graduate	25 (20.7)	23 (19.0)

Some College or Higher	18 (14.9)	36 (29.8)
No Response	1 (0.8)	2 (1.7)
Paternal Education		
No High School	8 (6.6)	8 (6.6)
High School Graduate	19 (15.7)	29 (24.0)
Some College or Higher	15 (12.4)	26 (21.5)
No Response	9 (7.4)	7 (5.8)

Note. Column number indicates number of participants in that condition; number in parentheses indicates percentage of total sample

Appendix C
Psychometric Properties of Screeners Based on ROC Analysis

	Specificity	Sensitivity
Total Sample (*N* = 121)		
M-CHAT (cut-off = 3)	.86	.27
M-CHAT (cut-off = 4)*	.80	.45
M-CHAT Critical (cut-off = 2)*	.64	.67
PDDST-II (cut-off = 5)	.21	.88
PDDST-II (cut-off = 8)*	.43	.71
SCQ (cut-off = 10)*	.87	.25
SCQ (cut-off = 15)	.57	.53
African-American (*n* = 58)		
M-CHAT (cut-off = 3)	.87	.19
M-CHAT (cut-off = 4)	.77	.37
M-CHAT Critical (cut-off = 2)*	.58	.56
PDDST-II (cut-off = 5)	.29	.93
PDDST-II (cut-off = 7)*	.35	.89
SCQ (cut-off = 15)	.39	.52
SCQ (cut-off = 26)*	1.00	.15
Caucasian (*n* = 51)		
M-CHAT (cut-off = 1)*	.79	.65
M-CHAT (cut-off = 3)	.82	.29
M-CHAT Critical (cut-off = 1)*	.79	.65
M-CHAT Critical (cut-off = 2)	.68	.76
PDDST-II (cut-off = 5)	.18	.88
PDDST-II (cut-off = 8)*	.44	.71
SCQ (cut-off = 11)	.79	.29
SCQ (cut-off = 12)*	.76	.41
Other (*n* = 12)		
M-CHAT (cut-off = 3)	1.00	.57
M-CHAT (cut-off = 5)*	1.00	.86
M-CHAT Critical (cut-off = 1)*	1.00	.71
M-CHAT Critical (cut-off = 2)	.80	.86
PDDST-II (cut-off = 5)	1.00	.29
PDDST-II (cut-off = 11)*	.60	.86
SCQ (cut-off = 13)*	1.00	.71
SCQ (cut-off = 15)	.60	.71
No High School (*n* = 16)		
M-CHAT (cut-off = 3)	.89	.29
M-CHAT (cut-off = 5)*	.78	.71

M-CHAT Critical (cut-off = 2)*	.78	.86
PDDST-II (cut-off = 10)*	.44	1.00
SCQ (cut-off = 15)	.78	.43
SCQ (cut-off = 23)*	.22	1.00
High School Graduate (*n* = 48)		
M-CHAT (cut-off = 3)	.91	.20
M-CHAT (cut-off = 4)*	.91	36
M-CHAT Critical (cut-off = 1)*	.91	.48
M-CHAT Critical (cut-off = 2)	.74	.56
PDDST-II (cut-off = 5)	.26	.92
PDDST-II (cut-off = 9)*	.52	.68
SCQ (cut-off = 10)*	.91	.28
SCQ (cut-off = 15)	.74	.44
Some College or Higher (*n* = 54)		
M-CHAT (cut-off = 3)	.82	.32
M-CHAT (cut-off = 6)*	.55	.63
M-CHAT Critical (cut-off = 2)*	.55	.74
PDDST-II (cut-off = 5)	.87	.16
PDDST-II (cut-off = 14)*	.13	1.00
SCQ (cut-off = 12)*	.76	.37
SCQ (cut-off = 15)	.58	.47

Note. * indicates the best cut-off score for sensitivity and specificity; Published cut-off scores for each screener: SCQ = 15, PDDST-II = 5, M-CHAT = 3, M-CHAT Critical Items = 2

Appendix D
Pairwise Comparisons of AUCs by Race Group

Group/ Contrast	AUC Difference	Standard Error	z	*p*
African-American				
M-CHAT/ M-CHAT Critical	0.04	0.05	0.84	.40
M-CHAT/ PDDST-II	0.03	0.13	0.26	.80
M-CHAT/ SCQ	0.02	0.15	0.12	.91
M-CHAT Critical/ PDDST-II	0.004	0.13	0.03	.97
M-CHAT Critical/ SCQ	0.05	0.14	0.39	.70
PDDST-II/ SCQ	0.05	0.08	0.60	.55
Caucasian				
M-CHAT/ M-CHAT Critical	0.08	0.03	2.46*	.01
M-CHAT/ PDDST-II	0.10	0.15	0.71	.48
M-CHAT/ SCQ	0.12	0.06	2.23*	.03
M-CHAT Critical/ PDDST-II	0.18	0.13	1.38	.17
M-CHAT Critical/ SCQ	0.20	0.07	3.08*	.002
PDDST-II/ SCQ	0.02	0.16	0.13	.90
Other				
M-CHAT/				

M-CHAT Critical	0.01	0.02	0.71	.48
M-CHAT/ PDDST-II	0.14	0.11	1.26	.21
M-CHAT/ SCQ	0.09	0.09	0.93	.36
M-CHAT Critical/ PDDST-II	0.13	0.10	1.29	.20
M-CHAT Critical/ SCQ	0.07	0.08	0.91	.36
PDDST-II/ SCQ	0.06	0.10	0.57	.57

Note. *****indicates significance at the $p < .05$ level

Appendix E
Pairwise Comparisons of AUCs by Maternal Education Group

Group/ Contrast	AUC Difference	Standard Error	z	*p*
No High School				
M-CHAT/ M-CHAT Critical	0.03	0.05	0.66	.51
M-CHAT/ PDDST-II	0.14	0.19	0.78	.44
M-CHAT/ SCQ	0.14	0.11	1.30	.19
M-CHAT Critical/ PDDST-II	0.11	0.21	0.53	.60
M-CHAT Critical/ SCQ	0.11	0.13	0.86	.39
PDDST-II/ SCQ	0	0.14	0	1.00
High School Graduate				
M-CHAT/ M-CHAT Critical	0.08	0.04	1.82	.07
M-CHAT/ PDDST-II	0.03	0.14	0.22	.83
M-CHAT/ SCQ	0.06	0.06	0.98	.33
M-CHAT Critical/ PDDST-II	0.04	0.13	0.33	.74
M-CHAT Critical/ SCQ	0.13	0.06	2.12*	.03
PDDST-II/ SCQ	0.09	0.15	0.60	.55
Some College or Higher				
M-CHAT/				

M-CHAT Critical	0.07	0.05	1.30	.19
M-CHAT/ PDDST-II	0.10	0.08	1.33	.19
M-CHAT/ SCQ	0.07	0.05	1.26	.21
M-CHAT Critical/ PDDST-II	0.17	0.08	2.22*	.03
M-CHAT Critical/ SCQ	0.14	0.07	1.98*	.05
PDDST-II/ SCQ	0.03	0.07	0.43	.67

Note. *****indicates significance at the $p < .05$ level