PARENTAL KNOWLEDGE IN SCREENING FOR AUTISM

by

JAMES ALEXANDER RANKIN

THEODORE S. TOMENY, COMMITTEE CHAIR
ANGELA BARBER
WILLIAM HART
RANDALL SALEKIN
MENGYA XIA

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2020

ABSTRACT

Timely diagnosis of autism spectrum disorder (ASD) is vitally important for improving the prognosis of young children with this condition. One of the greatest challenges facing healthcare providers for individuals with ASD and their families is shortening the time between when symptoms first appear and when an assessment for ASD is conducted. Current practice guidelines suggest pediatric screening should occur before 24 months of age for all children to help in detecting ASD as early as possible. Currently, screeners such as the Modified Checklist for Autism in Toddlers – Revised (M-CHAT-R) have been developed and validated for use in primary care settings. However, an underlying assumption behind screeners such as the M-CHAT-R is that parents are able to adequately understand the items on a screener questionnaire and relate those items back to their child's behavior. Using an item response theory framework, the current study found that the majority of behaviors characteristic of ASD assessed during the screening process are easy or very easy for parents to correctly identify. This study also found that greater parental knowledge of both child development norms and knowledge of ASD helped parents to accurately identify symptoms of ASD, but only when these symptoms were severe. Results of the current study help to highlight a fundamental divide in screening wherein more severe cases of ASD are well captured by current screening measures, but mild, less severe cases of ASD may require closer examination in future studies on screening accuracy.

LIST OF ABBREVIATIONS AND SYMBOLS

α	Cronbach's alpha
ADOS-2	Autism Diagnostic Observation Schedule, Second Edition
ASD	Autism Spectrum Disorder
ASK-ASD	A Survey of Knowledge of Autism Spectrum Disorder
β	Structural Beta Weight
CFI	Comparative Fit Index
CTT	Classical Test Theory
df	Degrees of Freedom
=	Equal To
F	Fisher's *F* ratio
IRT	Item Response Theory
KCDI	Knowledge of Child Development Inventory
<	Less Than
M	Mean: The sum of a set of values divided by the number of values in the set
MANOVA	Multivariate Analysis of Variance
M-CHAT-R	Modified Checklist for Autism in Toddlers - Revised
p	p-value: the probability that the null hypothesis is true
PSOC	Parenting Sense of Competence Scale
r	Pearson product-moment correlation coefficient
RMSEA	Root mean square error of approximation

RRBI's	Restricted and repetitive behaviors and interests
SD	Standard Deviation: Value of variation from the mean
SEM	Structural Equation Modeling
SRMR	Standardized root mean residual
t	T-test
χ^2	Chi-Square

ACKNOWLEDGEMENTS

This dissertation would not have come to completion without a host of endearing support, compassion, and guidance throughout this process. First, I would like to thank my graduate mentor Dr. Theodore Tomeny. I attribute my growth as a scholar and scientist largely due to his nurturing and endearing support. Ted's mentorship was more than just tangible advice throughout this process. He unlocked the best of my abilities and ultimately lead me to become the scholar I desired to become deep down. I will greatly miss our weekly meetings filled with wonderful discussions that never ceased to maintain and grow my passion for science.

Secondly, I would like to thank Dr. Randall Salekin. Dr. Salekin has been a great source of mentorship, encouragement, and support in the process of this dissertation and in my graduate training. Dr. Salekin was truly instrumental in developing my confidence as a clinician and as a researcher. His endearing encouragement and belief in me and my abilities often came through in the hardest times of graduate school.

Third, I would like to thank my officemates in the 'Broffice'. Their unbelievable accomplishments and knowledge continually served as inspiration. Every day my officemates encouraged me to be the best scholar and scientist I could be, while also encouraging me to never lose my own individual identity. I learned many lessons from them and shared a great number of fond memories in the 'Broffice' with them.

Fourth, I would like to thank my friends, family, and labmates for their encouragement, support, and joyful memories away from graduate school. My graduate studies, would not have been possible without a great deal of sugar-free Red Bull and heavy metal music. Roll Tide.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Diagnosis of ASD

Autism spectrum disorder (ASD) is a neurodevelopment disorder characterized by deficits in social interaction, understanding, and communication, as well as a pattern of restrictive and repetitive behaviors and interests (APA, 2013). Currently, in the United Stated 1 in every 59 children experiences ASD (Baio et al., 2018). Current measures of ASD such as the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2; Lord et al., 2012) and the Autism Diagnostic Interview, Revised (ADI-R; Rutter, Le Couteur, Lord, 2003) are well validated (e.g., Le Couteur, Haden, Hammal, & McConachie, 2008; Foley-Nicpon, Fosenburg, Wurster, & Assouline, 2017) and allow for diagnoses of ASD to be made as early as 2 years of age (Baio et al., 2018). In most cases, symptoms of autism are first noticed around 12-18 months, with deficits in verbal abilities and absence of eye contact often first appearing around this age. Research suggests that some parents may notice symptoms of ASD as early as 14 months of age, but many parents do not notice symptoms until much later (Chawarska et al., 2007). Historically, diagnoses of ASD have typically been made at much later ages than when symptoms first appear (Baio et al., 2018). Average age at diagnosis has been improving over time (Hertz-Picciotto & Delwiche, 2009); however, current average ages at which children with ASD are diagnosed (around 4 years of age) are still often far later than recommended for best treatment outcomes (Baio et al., 2018; Daniels & Mandell, 2014). This is particularly true for low SES children and Medicaid-enrolled children, who often do not receive an ASD diagnosis until after 5 years of age (Fountain, King, & Bearman, 2011; Mandell et al., 2010).

Reasons for this delay in diagnosis may be due to a lack of available resources to families, low knowledge about ASD among parents, as well as inadequate knowledge of ASD symptoms among professionals who may refer families to colleagues who may not be able to provide an ASD-specific evaluation (Daniels & Mandell, 2014; Daniels, Halladay, Shih, Elder, & Dawson, 2014). Children with atypical development who have older siblings, regardless of whether or not the sibling has ASD, tend to be identified earlier (Herily, Knoch, Vibert, & Fein, 2015). Alternatively, children of parents with less education attainment tend to be diagnosed after school-age more often than children of parents with greater educational attainment (Yergin-Allsopp et al., 2003). This body of literature becomes alarming when one considers the growing and consistent research that suggests that the earlier an ASD diagnosis is made, the better the outcomes are for the child with ASD (e.g., Charman & Baird, 2002).

Importance of Early Detection and Diagnosis of ASD

Research on the treatment of ASD has often suggested that ASD persists across the lifespan, and that no current treatment appears to completely abate symptoms of the condition (Roane, Fisher, & Carr, 2016). Whereas areas such as comorbid anxiety, social skills, and verbal communication can be improved through various interventions, no current treatment offers broad long-term improvement in core symptoms of ASD to a point that the majority of symptoms are significantly reduced. This same research literature identifies one factor that appears to be consistently linked to improved outcomes for individuals with ASD: early diagnosis (Dumont-Mathieu & Fein, 2005). The earlier a child is diagnosed, the earlier intervention services can be started. Interventions such as Naturalistic Developmental Behavioral Interventions (NDBIs; Schreibman et al., 2015), are recommended to begin as early as possible so that parents can be educated on how to help their child and address specific communication, cognitive, and other

skill gaps (Wallace & Rogers, 2010). The earlier these interventions begin, the better the effects of the treatment are at bridging developmental gaps experienced by children with ASD (Wallace & Rogers, 2010). With a consistent literature base supporting improved outcomes for those who are diagnosed early, high-quality, universal screening for ASD is imperative.

Importance of Pediatric Screening of ASD Symptoms

Given the significant impact of early detection on the lives of those affected by ASD, successful early screening has been a vital effort for providers of pediatric care. Current recommended practice is universal screening, in which all children are screened for ASD in pediatric care settings before age 24 months (Kim et al., 2016). This screening should involve comprehensive interviews (Robins et al., 2014). However, due to time constraints common to pediatric care settings, ASD is currently screened via questionnaires (Kim et al., 2016). It would be ideal if pediatricians were able to identify all warning signs of ASD as part of any routine check-up. However, with the limited amount of time pediatricians are able to spend with parents, pediatricians often struggle to have enough time to accurately screen for any sign of developmental psychopathology (Dosreis et al., 2006). Currently, screening for developmental concerns, such as ASD, is often conducted via parent-report questionnaires completed by parents prior to or during their appointments. However, the validity of this practice presumes 1) that parents understand the questions being asked of them and 2) that parents are able to accurately apply the questionnaire items to their children's behavior.

Validity of Informant Reports

For longstanding periods in clinical child psychology, parents have served as primary informants of the functioning of children in their daily lives (De Los Reyes et al., 2015; Dumont-Mathieu & Fein, 2005). Research provides support for parental report as a valuable insight into

the child's behavior and emotional expression at home, a context wherein psychologists often have little direct observation of a child (De Los Reyes et al., 2015; Dumont-Mathieu & Fein, 2005). Although parents have the opportunity to view their child's behavior for much longer periods of time than psychologists, the question of whether parents are accurate reporters of their child's behavior and emotional expression has persisted for years (De Los Reyes, 2011).

Limitations of parent report measures. Concerns regarding the validity of utilizing untrained reporters (parents) to rate a child's behavior has come in to question frequently in the field of child psychology (Hartman, Rhee, Willcutt, & Pennington, 2007; Howells-Wrobel & Lachar, 1998). Measures of psychopathology that utilize parent report assume that the respondent understands the questions being asked in the questionnaire. While these measures are explicitly designed so that respondents should be able to understand question items (i.e., these questionnaires are written in simple, explicit terms, often incorporating feedback from stakeholders), it is difficult to assess how often respondents are able to accurately answer questions correctly. Seminal work by Baranek (1999) examined home-video footage of children, coded the behaviors, and then compared their findings with parental report. Unsurprisingly, Baranek (1999) found that parental report was vastly different from what behaviors were actually occurring at home.

A variety of factors have been considered in attempts to explain sources of measurement error inherent in the use of parent report questionnaires. This error can result from parents not understanding an item's wording or not being able to understand the concept. For instance, parents are often inaccurate in their reporting of child development norms, even after being trained on what those norms are (Robbins, 1963; Ertem et al., 2007; Lee, 2009; Reis, 1988). Relatedly, a parent might understand what a term means, but may not be able to identify that

phenomenon. For example, the parent may know what proper eye contact would look like but might not attend to these behaviors in their own children when not specifically focusing on noticing these behaviors. For instance, parents often have difficult in accurately recalling past events, even when noticeable (Hassan, 2006). Parents of children with ASD often provide inaccurate reports on the timing of language development and social regression when asked to report on symptoms retrospectively (Ozonoff, Li, Deprey, Hanzel, & Iosif, 2018). While this literature may portray parents as uninformed, a longstanding body of research suggests that, with training, parents can achieve high levels of understanding of child development and of ASD symptoms (Matson, Mahan, & Matson, 2009). Parent training has long stood as one of the most efficacious treatments for a variety of childhood emotional and behavioral concerns, largely due to its role in educating and training parents in noticing behavior and responding appropriately to both adaptive and maladaptive patterns of behavior (Chronis, Chacko, Fabiano, Wymbs, & Pelham, 2004; Kazdan, 2009; Matson et al., 2009; Serketich & Dumas, 1996).

Beyond understanding items and being able to identify behaviors, a third concern regarding measurement error in parent report questionnaires are a host of rater biases. Parent reports tend to contain more biases than teacher ratings (Hartman et al., 2007). Bias source can be specific to an individual, such as a parent holding an especially negative view of their child's behavior, which often occurs when parents are depressed (Jacob & Johnson, 1997). Some sources of rater biases may be more inherent in parents as a group. For instance, Hartman and colleagues (2007) found that when parents rate the behavior of twins, they tend to report larger differences between the twins' behavior relative to teachers, as parents have fewer points of comparison relative to teachers. Some work suggests that rater biases alone explain differences in informants' reports. However, others argue differences in informants' reports are better

encompassed by a mix of rater biases, unique perspectives of different informants, and contextual variation in the behaviors being reported on (De Los Reyes, 2011; Van der Valk, Van Den Oord, Verhulst, & Broomsma, 2001).

Overall, clinical psychology persists in using parent report questionnaires despite these concerns. Research still contends that parents can accurately report on behaviors at home to a high enough degree to suggest incorporating parent report questionnaires into the assessment process (Achenbach & Edelbrock, 1981; Barton, Dumont-Mathieu, & Fein, 2012; De Los Reyes, Henry, Tolan, & Wakschlag, 2009; Norris & Lecavalier, 2010). Whereas a strong literature base acknowledges these sources of error and demonstrates utility of parent report measures even in the face of this expanded rate of error (De Los Reyes et al., 2015), deeper testing of parent's actual ability to identify and rate behaviors characteristic of ASD is still needed.

Use of the MCHAT

Many screening measures for ASD that utilize parent report have been developed. Research on screening measures of ASD suggest that not all screeners are equal, and that a factor that explains differences in validity among screening measures is often which items they choose to include (Hampton & Strand, 2015). For instance, screening measures that probe more thoroughly into social interaction skills tend to have higher validity than measures which probe other areas equally (Hampton & Strand, 2015).

The Modified Checklist for Autism in Toddlers-Revised (M-CHAT-R; Robins et al., 2014) is one of the most commonly used measures in broad, level-1 pediatric screening of ASD (McPheeters et al., 2016), as recommended by the American Academy of Pediatrics (Johnson & Myers, 2007), the American Academy of Neurology, and the Child Neurology Society (Filapek, 2000). The M-CHAT-R is also used internationally in countries like Spain (Canal-Bedia et al.,

2011), Japan (Inada et al., 2011), and South Korea (Seung et al., 2015). Use of the M-CHAT-R in primary care settings is viewed as a low-cost (Gura, Champagne, & Blood-Siegfried, 2011) way to reduce time between when symptoms first appear and when an ASD diagnosis is made, particularly for racial/ethnic minority groups (Herlihy et al., 2015). With the widespread use of the M-CHAT-R, a considerable research base has examined the psychometric properties of this measure and other related screeners.

Psychometric properties of the M-CHAT-R. Overall, the M-CHAT-R has demonstrated good psychometric properties beyond its ease of use and its cost utility (Robins et al., 2014). For instance, nearly all children who score high on the M-CHAT-R have some developmental concern warranting intervention (Chlebowski, Robins, Barton, & Fein, 2013; Robins et al., 2014; Toh, Tan, Lau, & Kiyu, 2018). Initial Positive Predictive Value (PPV) for the M-CHAT-R is low in large scale studies (.36) but is much higher when families of children who screen positive are contacted for a follow-up interview (.74; Chelbowski et al., 2013). Research suggests that the M-CHAT-R is less accurate with children younger than 21 months old but performs well in children who are between 22 and 48 months of age (Toh et al., 2018; Yama, Freeman, Graves, Yuan, & Campbell, 2012). Another population in which psychometric properties of the M-CHAT-R are not as strong includes children born prematurely, for which the M-CHAT-R has a high false positive rate (48%) and a low positive prediction rate (20%; Kim et al., 2016). Research also suggests that while children who are diagnosed with ASD score higher on the M-CHAT-R, children who have other developmental delays or atypical features also score high on the M-CHAT-R (Sturner et al., 2016). In a large-scale study of more than 16,000 children, only 48% of children who screened positive on the M-CHAT-R received an ASD diagnosis (Robins et al., 2014), suggesting that simply scoring high on the M-CHAT-R may not

be enough to accurately screen for who may meet criteria for an ASD diagnosis. Such a high false positive rate is problematic for parents, by creating unwarranted stress and anxiety, and for professionals, by creating more demand for advanced testing, thereby increasing waitlist times.

M-CHAT-R follow-up interview. Considering findings regarding the psychometric properties of the M-CHAT-R, Sturner et al. (2016) argue that pairing the M-CHAT-R screener with a follow-up interview appears necessary for the M-CHAT-R to be used as an effective (i.e., accurate) screener in pediatric settings, and that pediatricians are capable of doing these interviews with minimal training. Another study suggests that parent-reported information provided during pediatric appointments that supplements the M-CHAT-R is useful in re-specifying misclassified children on the M-CHAT-R (Yama et al., 2011). Unfortunately, few pediatric providers have the time or resources to conduct follow-up assessments with parents, and rarely do they occur (Dosreis et al., 2006).

One of the goals of the follow-up process for the M-CHAT-R is clarifying parental misconceptions regarding their child's behavior and the behaviors they are asked about (Robins et al., 2014). Currently, limited research has been conducted on parents' underlying understanding of the behaviors they are asked to rate. Although often assumed, no research, to date, has assessed whether clarifying potential parental misunderstandings explains the positive gains in the M-CHAT-R's utility after the follow-up process. Parental knowledge of normative behavior and behavior related to ASD may affect how parents rate their own children on the M-CHAT-R. De Giacomo and Fombonne (1998) suggest that lack of knowledge about normative child development among parents results in a delay in the recognition of ASD symptoms in children. Research also suggests that some parental factors are likely related to the age of recognition of symptoms of ASD (Chawarska et al., 2007). For instance, prior experience raising

a child appears to reduce delays in noticing ASD symptoms, even to some degree amongst parents with low knowledge of child development norms (De Giacomo & Fombonne, 1998). As previously stated, parents' report of child behaviors at home are often vastly different from what a trained professional would report (Baranek, 1999). Clinicians and physicians are widely regarded as being generally more knowledgeable about developmental disorders than parents (Dumont-Mathieu & Fein, 2005). As such, the interview process following the M-CHAT may bridge this gap between parental knowledge and clinician expertise. Particularly, there may be some symptoms of ASD, and thus some screener items, that parents may have difficulty rating correctly. However, with assistance from a physician or clinician, a parent may be better able to rate these items correctly, thus improving the accuracy of the screener. Psychometric research examining the differential functioning of items featured on various ASD screeners remains somewhat scarce. Therefore, a deeper, psychometrically-focused examination of the items contained within the M-CHAT-R may be required.

Item Response Theory vs Classical Test Theory

For years, the foundational psychometric theory which underlaid the development of any test/questionnaire was classical test theory (Crocker & Algina, 1986). Classical Test Theory (CTT) is the idea that a test-taker has an observed score on a test, which is an estimate of their true score of the ability being tested. The equation: True Score = Observed Score + Error, is well known and frequently taught in measurement-centered courses in graduate training of psychologists. Item Response Theory (IRT) is an area of psychometrics that has been growing since its conceptualization in the 1970's. IRT posits that the probability of a particular response (most often a "correct" response) is a mathematical function integrating estimates of both item characteristics and person characteristics (Baker & Kim, 2017). Specifically, IRT integrates item

characteristics (such as the difficulty and discriminatory power of an item) to predict a person's probability of providing a particular response given a certain level of a singular latent trait or dimension (e.g., depression level; Baker & Kim, 2017).

While both CTT and IRT have largely been applied to the development of tests, their theoretical implications impact the development of informant report questionnaires as well (Cappelleri, Lundy, & Hays, 2014). Some research suggests that IRT provides greater information relative to CTT in how to improve rating scales (Petrillo, Cano, McLeod, & Coon, 2015). Recently, IRT has gained greater use in clinical psychology measurement with various applications including examining the continuity of traits, examining response process, and examining differences in traits between groups (see Reise & Waller, 2009). Yet, it is still reported that IRT's application to clinical assessment is limited (Thomas, 2011; Reise & Rodriquez, 2016). In a systematic review of the use of IRT in clinical assessment, Reise and Waller (2009) suggest that IRT is often applied to understand existing scales, but it is often done to examine item discrimination only. For instance, a researcher may be interested in how specific items on a questionnaire differentiate those who are depressed versus those who are not depressed. Item difficult, however, is often of little interest to clinical researchers using parent-report questionnaires, as there often is not a "correct" or "incorrect" answer choice dichotomy. One of IRT's most fundamental applications is in understanding the difficulty of each specific item within an item pool. IRT may help provide an added dimension to the refinement and further development of ASD screeners by serving as a theoretical basis for explaining why not all items are created equal in the case of ASD screeners. Specifically, IRT may help in understanding how difficult it is for a parent to identify a particular behavior asked about within a screener item. This information can help refine screening questionnaires to focus on behaviors

that parents can readily identify and highlight target areas where parental knowledge may need to be increased prior to screening for ASD.

Theoretical Integration

Considering the importance of early detection of ASD, it is imperative for the autism field to find any way possible to enhance current screening procedures. The current state of screening is good but leaves room for improvement. Specifically, research has established that some of the items used in ASD screening measures may not perform as well as other items (Hamptom & Strand, 2015). Parent report, while validated and useful, is often underlain with error (Hartman et al., 2007). When rating ASD symptoms, parents tend to display some inaccuracy (Baranek, 1999), and these do appear in the screening process. Reviews of screening measures have identified that the M-CHAT, and the subsequent M-CHAT-R, as being far superior than other available screeners for ASD historically (Dumont-Mathieu & Fein, 2005). The M-CHAT-R is also the most commonly used screening measure in pediatric practices (Siu et al., 2016). As such, examination of the M-CHAT-R represents examination of the best-to-date, most widely used tool available for screening.

While a highly useful measure, the M-CHAT-R often requires the input of an expert interviewer to clarify parental report (Robbins et al., 2014). Taken together, parents are often accurate reporters, but show some difficulties in their reporting. This difficulty may vary across different items that measure different aspects of ASD (Hampton & Strand, 2015), thus creating differences in how easily parents are able to accurately complete different measures. Given this evidence, difficulty experienced by parents in either understanding, identifying, or reporting of behaviors characteristic of ASD may help explain variability in screener performance. This may further explain why parents often require follow-up interviews to ensure diagnostic accuracy.

Beneath any screening questionnaire is a latent ability in the rater to understand, notice, and report on the behaviors being asked about (Reise & Walker, 2009). Within the development of most questionnaires, it is often assumed that parents have the requisite knowledge of what the item means and can accurately recall past experiences and apply an item to their child's situation. However, this assumption often goes un-tested. Thus, variability across items may be a point of interest in assessing parental report abilities within the screening process for ASD. However, previous attempts to understand these items, and parental ability to accurately rate these items, may have been hindered by a lack of application of an item response theory framework (Reise & Walker, 2009; Thomas, 2011; Reise & Rodriquez, 2016).

Several factors related to the timing of an ASD diagnosis may impact how well parents are able to identify symptoms. Parental knowledge serves as the first point of consideration in parental variables that may affect their ability to detect symptoms of ASD (Daniels & Mandell, 2014; Daniels et al., 2014). Greater knowledge may help parents better understand how ASD presents in children and what normative development should look like. However, factors that may explain parents' use of their knowledge warrant exploration as well. For example, an important area of research considers the interplay between parental knowledge and parental self-efficacy in explaining parental behavior (Bandura, 1977; Reiner-Hess, Teti, & Hussey-Gardner, 2004). Specifically, literature considers self-efficacy as a variable that impacts how well parents translate their knowledge into effective parenting behaviors (Jones & Prinz, 2005). Thus, parental self-efficacy is a second variable warranting consideration of investigation. Parents' self-reported confidence in their knowledge of ASD symptoms may perform a similar role, wherein parents with higher confidence in their knowledge may more easily, or more quickly, identify a behavior as atypical than parents who lack confidence in their knowledge of ASD. The

number of children a parent has already raised is linked with how quickly a parent expresses concerns when a younger child displays atypical development (De Giacomo & Fombonne, 1998; Herily et al., 2015). The number of children a parent has previously raised may serve as a proxy for points of comparison between children and may provide parents with real-world examples of normative development. Educational attainment has also been linked to timing of diagnosis of ASD (Yeargin-Allsopp et al., 2003), however the specific mechanism in how educational attainment affects identification of ASD symptoms is currently unknown.

Current Study

This study seeks to understand whether parents who currently have a child between 18 to 30 months of age are able to accurately identify atypical behaviors when shown symptoms characteristic of ASD, and to understand how parental knowledge of ASD symptoms relates to parents' abilities to accurately report on screening measures such as the M-CHAT-R. More specifically, the first major aim of this study is to understand whether parents are able to accurately identify behaviors characteristic of ASD, and to identify behaviors that may be easier or more difficult for parents to rate (Aim 1). The second aim of this study is to understand whether parental accuracy in identifying these symptoms is related to parental knowledge (of child development and of ASD), parental self-efficacy, parental education, and the number of children the parent has raised (Aim 2).

Hypotheses

It was hypothesized that (1a) parent ratings would vary from clinician ratings on the M-CHAT-R when rating the same child and (1b) that parent ratings of severe symptoms of ASD would be more accurate than parent ratings of more mild symptoms of ASD. It was further hypothesized that (2a) greater parental knowledge of child development and greater parental

knowledge of ASD would predict accuracy. Second, it was predicted that (2b) greater parenting self-efficacy, greater parental education, and greater number of children older than 48 months would all significantly predict greater accuracy in identifying symptoms of ASD. Lastly, it was predicted that (2c) parental confidence in knowledge of ASD would predict accuracy in identifying symptoms of ASD beyond what was already explained by parental knowledge.

CHAPTER 2

METHOD

Procedure

IRB approval was obtained prior to data collection (Appendix A). A sample of 288 parents were recruited to complete study measures through an online survey panel (e.g., Qualtrics Panels). Participants were mothers or fathers who currently have a child between the ages of 18 to 30 months old. Parents who fit demographic criteria were notified via e-mail of their eligibility for the study, and parents who were interested in the study were directed to a Qualtrics page that contained the study materials. After providing their consent, parents began by providing demographic information such as their age, number of children, children's ages, and annual household income. Next, parents completed a measure of knowledge about child development, a measure of knowledge of ASD symptoms and confidence in identifying ASD symptoms in children, and a measure of parental self-efficacy (see the Measures section for more information). Next, parents were informed that they would view brief videos (around 5 minutes each) of two children, and that they would be asked to answer yes/no questions about each child's behavior following each video. Each parent then viewed the videos of the two children with ASD, one who scored in the moderate range on the ADOS-2 (17 score on Module 1; Comparison Score of 6; 'Mild' condition), and one who scored in the severe range (25 score on Module 1; Comparison Score of 10; 'Severe' condition). While the video for the 'Mild" condition falls under the moderate range on the ADOS-2, it is worth noting that this score is the lower-bound of the range of 'Autism' on the ADOS-2, (while the 'Severe' condition scored on the upper-bound of this range; a score on the ADOS-2 may fall below the 'Autism' range and still be scored as 'Autism Spectrum'). For the purposes of the current study, the former video

will be considered as the *milder* presentation of ASD and will serve as the 'Mild' condition video. These videos consisted of selected components of the ADOS-2, Module 1, that are intended to elicit behaviors that are screened on the M-CHAT-R (see Videos section below for more details). Parents then rated each child on the selected M-CHAT-R items directly after watching the child's respective video. Order of video presentation (i.e., 'mild' or 'severe' case first) was counterbalanced.

Videos. Prior to data collection, a set of autism experts (professionals with a terminal degree who regularly interact with/serve children with ASD) aided in selection of which components of the ADOS-2 administration were shown in the video stimuli. The experts first viewed all of the items on the M-CHAT-R and were then provided with a list of each task on the ADOS-2 and the corresponding description and behaviors elicited in the task. Experts then rated each ADOS-2 item as 'essential' 'useful but not essential' or 'not useful'. Content validity ratios were used to determine which tasks should be featured in the videos and which tasks should be discarded (for more on this approach, see Lawshe, 1975). Resultingly, any parts rated as 'essential' by at least one expert were included in both videos. The videos were culled from diagnostic assessments performed at an ASD specialty clinic. Among families willing to participate, two videos best matched on age, race, and number of parents present in the room were chosen. Two autism experts experienced in assessing children for ASD then watched the same videos viewed by participants and rated each of the children using the M-CHAT-R. These scores were used as benchmark ratings.

Measures

Demographic information form (Appendix B)**:** Participants completed a demographic form that included relevant information about the participants and their families. Specifically, the

demographic form included questions about number and ages of children and the parent's education level, income, and level of experience with children with developmental disabilities.

The Modified Checklist for Autism in Toddlers - Revised (M-CHAT-R; Robins et al., 2014). The M-CHAT-R is a widely used pediatric screener for ASD with evidence of good psychometric properties (see section above titled "Psychometric Properties of the M-CHAT-R"; Robins et al., 2014). The M-CHAT-R assess topics such as eye contact, response to name, repetitive motor movements, gaze shifting, social motivation, requesting, and shared enjoyment. Response options for the M-CHAT-R are dichotomous ("Yes" or "No"). The M-CHAT-R is typically completed by parents of children between 2-4 years of age. For the purposes of the current study, the M-CHAT-R was administered to parents as they rated a series of children viewed in a set of videos (i.e., these children were not the parents' own children). Therefore, the wording of the M-CHAT-R items was amended throughout (e.g., "your child" was changed to "this child"). For any items in which both experts rating the child agreed that the item could not be rated based on the video content (i.e., there was no opportunity for the child to display or not display a particular behavior), the respective item was omitted and parents were not asked to respond to the item.

A Survey of Knowledge of Autism Spectrum Disorder (ASK-ASD; Hansen, 2015). The ASK-ASD is a 51-item measure of perceived and actual knowledge of ASD etiology, epidemiology, symptoms, diagnosis, and prognosis/treatment. Participants are first asked to respond whether a statement is "True" or "False" to indicate the accuracy of their knowledge about each item. Participants then indicate their level of confidence in each answer on a Likert scale of 1 = Not At All Confident, 2 = Confident, and 3 = Very Confident. Scores from the ASK-ASD have demonstrated test-retest reliability, adequate internal consistency, fair validity

coefficients, and a two-factor structure in previous examination of its psychometric properties (Hansen, 2015). The total score derived from the ASK-ASD was utilized in analyses. Internal consistency was excellent for the ASK-ASD confidence scale (α = .968) and good for the ASK-ASD knowledge scale (α = .869).

Knowledge of Child Development Inventory (KCDI; Larsen & Juhasz, 1986). The KCDI is a 56-item multiple choice measure of parents' factual knowledge of emotional, cognitive, physical, and social development. The KCDI captures parental knowledge of child development norms from birth to age three. The KCDI is a multiple-choice questionnaire with four response options per question. The measure was developed to be used with both parents and adolescents and requires an eighth-grade reading level (Larsen & Juhasz, 1986). The KCDI has demonstrated good criterion validity, content validity, and internal consistency (Larsen & Juhasz, 1986). For the current study, only the measures of emotional and cognitive development were utilized. Internal consistency was good for the two scales of the KCDI combined (α = .812), internal consistency of the emotional scale was acceptable (α = .765), while internal consistency of the cognitive scale was poor (α = .582).

Parenting Sense of Competence Scale (PSOC; Johnston & Mash, 1989). The PSOC is a 17-item measure of parental perceptions of competency in parenting. Items are answered on a 6-point scale ranging from "strongly disagree" to "strongly agree". Items load onto one of two scales, "Satisfaction" or "Efficacy", with higher scores indicative of greater parenting self-esteem. The PSOC has demonstrated good internal consistency (Johnston & Mash, 1989), construct validity (Karp, Lutenbacher, & Wallston, 2015), and convergent and divergent validity (Ohan, Leung, & Johnston, 2000). Internal consistency for the PSOC total scale was good in the current study (a = .852).

Data Analytic Plan

Preliminary data analysis. Prior to analysis, data were screened for violations of univariate normality, violations of homoscedasticity, and for outliers. To measure parental accuracy, two experts rated each child on the M-CHAT-R via the video stimuli. From there, when both experts rated the child the same on a particular item, parental accuracy was modeled as whether the parent produced the same rating as the experts (correct) or whether the parent produced the opposite rating of the experts (incorrect). However, when there was disagreement between expert scores, the respective item was dropped from analyses. As such, analyses only focused on items where expert consensus could be reached.

Main hypotheses. For the current study, item response theory (IRT) was utilized to identify items that parents may struggle to rate accurately (for a review see Baker & Kim, 2017). IRT is used to explain performance on test items based on a latent trait/ability. IRT was implemented because the current project utilized item responses that contain a correct (response matches clinician rating) or incorrect (response differs from clinical rating) dichotomy to capture a parental latent ability of 'ability to correctly identify ASD symptoms'. While the M-CHAT-R probes for some of the most common, most salient symptoms of ASD (Robins et al., 2014), the M-CHAT-R does not cover the full range of ASD symptoms. Further, parental ability to identify ASD symptoms is likely to vary from child to child. As such, the M-CHAT-R items are considered only approximations of parental ability to accurately identify symptoms of ASD. Therefore, ratings on the 20 M-CHAT-R items across two children serve as observed instances of a largely latent construct of parental ability to identify ASD symptoms.

A critical component of any IRT model is the item characteristic curve, which plots the latent (unobserved) ability on the X-axis and the probability of a correct response on an item on

the Y-axis. Each test item has its own item characteristic curve (ICC; also referred to as the item response function). Two key features of the ICC present the fundamental information to be understood from IRT. The first feature, "Item Difficulty," describes the location of the curve along the ability scale. Item difficulty, mathematically speaking, is the point on the ability scale at which the probability of a correct response to an item is 50%, which signifies how difficult it is for an average test taker to get that particular item correct. Item difficulty is denoted by *b* and can range from negative infinity to infinity, but typically ranges from -3 to 3 in practice (Baker & Kim, 2017). The second feature, "Item Discrimination" (denoted by *a*), describes the steepness of the curve. Item discrimination informs how well an item can differentiate between different test-takers based on ability. A steeper curve means that an item was better able to differentiate people of varying abilities. Specifically, a steep curve results in greater probability of a correct response being tied to individuals who score higher in the ability, while people who score lower in the ability have a lower probability of a correct response. A flat item characteristic curve essentially represents similar probability of a correct response between people both high and low in the ability, thus providing little differentiation in performance based on ability.

Within the current study, the principal aim of using IRT was to identify items that parents do not typically rate correctly. The function of the M-CHAT-R is not to be a test of ability to identify ASD, but rather to serve as a broad screener of ASD in the parent's own child, completed in primary care settings as a first-line detection tool. Thus, item discrimination within IRT using M-CHAT-R responses holds little practical relevance. As such, the current study utilized a special case of IRT, a one-latent parameter model (i.e., the Rasch model) to estimate item difficulty. The Rasch model was developed by George Rasch as a psychometric model to understand dichotomous items on a test or questionnaire (see Rasch, 1960). In this model, item

discrimination (*a*) is fixed at a value of 1, and only item difficulty parameters can be estimated (Baker & Kim, 2017). In the current study, a Rasch model examining parental ability to accurately identify ASD symptoms was employed. Specifically, "parental ability to identify ASD symptoms" served as the underlying latent ability being measured. The test of this ability was the accuracy ratings on the M-CHAT-R items (e.g. did the parent provide the same response as the clinician on this item). Thus, item difficult reflects how difficult it was for a parent with an average ability to recognize ASD symptoms to accurately rate each item (i.e. behavior) on the M-CHAT-R.

In the present study, IRT yields several benefits over only reporting the percentage of parents who rated an item correctly. First, IRT allows a descriptive indicator ranging from Very Easy to Very Hard that corresponds to the level of difficulty of the individual item. Second, it provides an indication of the *probability* of a correct response, based on various levels of a latent ability (i.e., IRT difficulty parameters provide the probability of a correct response for an individual of average ability). Third, IRT incorporates a parent's performance across items in calculating how likely it is that the parent would get a particular item correct. Collectively, this IRT approach provides greater ability to generalize across a sample to understand the performance on a set of items for a *mean-level responder.* In the current study, IRT difficulty parameters and probabilities are reported in addition to the percentage of the sample that rated the item correctly, for each item.

Secondly, structural equation modeling (SEM) was employed to ascertain what factors are related to parental ability to identify ASD behaviors accurately. Parent knowledge of ASD (as measured by the ASK-ASD Knowledge scale) and of child development (as measured by the Emotional and Cognitive Development subscales of the KCDI) were loaded onto a latent

exogenous (i.e., predictor) “Parental Knowledge” factor. The use of a latent factor to measure parental knowledge was due to the likely high correlation between the three domains of knowledge. Parental self-efficacy (as measured by the PSOC), the number of children the parent has previously raised (measured via the demographic form), and parental educational attainment (measured via the demographic form) were treated as observed predictors. Additionally, Parental Confidence in knowledge of ASD (measured by the confidence scale of the ASK-ASD) was modeled as an observed variable predicting M-CHAT-R items.

M-CHAT-R items loaded onto two separate latent endogenous (i.e., criterion) factors reflecting both severity conditions (i.e., ‘Mild’ and ‘Severe’; see Figure 1). Because the complete set of items was not used (e.g., items clinicians disagreed upon or rated as unable to be rated), latent factors were used to model accuracy on the M-CHAT-R items. This approach of using two separate single-order factors to model M-CHAT-R items was compared to several other approaches of modeling, specifically the bi-factor (items load onto their respective factor and a general ‘M-CHAT-R’ factor), general factor (items do not load onto ‘Mild’ or ‘Severe’ factors and only load onto a single ‘M-CHAT-R’ factor), and second-order factor models (the ‘Mild’ and ‘Severe’ factors load onto a second-order factor, ‘M-CHAT-R’).

Two-step modeling was then employed to test fit of a) measurement model via confirmatory factor analysis and b) of structural model via structural equation modeling. First, separate measurement models were specified for exogenous latent factors (i.e. predictive, independent factors; see Figure 1) and for endogenous latent factors (i.e. criterion, dependent factors; see Figure 1). From there, paths stemming from both exogenous latent factors (i.e., Parental Knowledge and Parental Confidence) to both endogenous latent factors were specified. Diagonally weighted least squares (DWLS), with robust standard errors and a mean and variance

adjusted test statistic (known as WLSMV in R), was selected as the method of estimating the goodness of fit of the model. This option was chosen because it is the preferred option for estimating non-normal data in small sample sizes (n = 200-400; Li, 2016). Analyses were completed using the Lavaan package in R (Rosseel, 2012). The following model fit criteria were used: the chi-square/df approach suggested by Kline (2005), Robust CFI, Robust RMSEA, and SRMR. Model comparison statistics included change in Chi-Square and change in Robust CFI.

CHAPTER 3

RESULTS

Descriptive Data:

Sample demographics. A sample of 288 participants was obtained through data collection (correlations between variables of interest are presented in Table 1; sample demographics are presented in Table 2). The average age of the parents in the sample was 34 years old (SD = 7.99) and the age of parents in the sample ranged from 19 years of age to 65 years of age. 137 parents identified as male, while 151 parents identified as female. The majority of the sample identified as White (167; 58%), while 58 parents identified as African American/Black (20%), 32 parents identified as Asian/Pacific Islander (11%), 15 parents identified as Hispanic/Latino (5%), and 3 parents identified as Native American (1%). Six parents identified as multiracial (2%), while 7 parents preferred not to respond or suggested that their race was not listed (2.5%). The majority of the sample (256 parents; 89%) reported they were not Hispanic/Latino, while 32 parents (11%) reported they were Hispanic/Latino. Relative to United States Census data (U.S. Census Bureau, 2017), this sample underrepresented White racial groups (76.5% in the census) and Hispanic/Latino racial groups (18% in the census), while oversampling Asian/Pacific Islander racial groups (6% in the census) and African American/Black racial groups (13% in the census; U.S. Census Bureau, 2017).

In terms of parental education, 40 parents (14%) identified their highest education as High School Graduate/GED, while 76 parents (26%) identified their highest education as Some College/Associates Degree . Furthermore, 13 parents (4.5%) reported having a Technical School Degree, 83 parents (29%) reported having a Bachelor's degree, 64 parents (22%) reported having a Master's Degree, and 11 parents (4%) reported having a Professional Degree . The percentage

of respondents in the sample with a bachelor's degree or higher was twice that of US Census data (60% in this sample versus 30% in the census; U.S. Census Bureau, 2017). Parents reported having between 2 to 3 children on average (M = 2.347, SD = 1.391; range 1 to 9). The median household income reported was $65,000.

87 parents (30%) reported that they had no prior exposure to someone with ASD, while 34 parents reported having a child with ASD (12%), 15 parents reported having a sibling with ASD (5%), 28 parents reported having a friend with ASD (10%), and 20 parents reported that they work with children or adults with ASD (7%).

M-CHAT items. Overall, both experts displayed substantial agreement in their ratings of the two children in the video clips on the M-CHAT-R based on Cohen's Kappa (.771 for 'mild'; .657 for 'severe'; .714 across all ratings). Across both the mild and severe videos, experts agreed that three items (8, 12, and 20) could not be accurately rated given the information present in the videos. In the severe video, clinicians agreed on all but three items (items 2, 4, and 5), while clinicians agreed on all but two items in the mild video (items 10 and 11). Thus, 15 items were retained from the mild video and 14 items were retained for the severe video (12 items overlapped and were retained in both videos).

Percentages of parents who matched clinician ratings (coded as '1' for accurate/matches experts or '0' for inaccurate/does not match clinicians) on each item is presented in Table 3. Item 13 ("*does this child walk*") was the item rated accurately most often by parents across both mild and severe item sets (93% of parents rated each item correctly). Regarding which item was most difficult in each item set, item 1 ("*if the examiner points at something across the room, does the child look at it*") was the item rated accurately least often across the mild items (42% of parents rated this item correctly), while item 6 ("*does this child point with one finger to ask for*

something or to get help") was the item rated accurately least often across the severe items (64% of parents rated this item correctly). Across all 12 items that were retained from both the severe and mild videos, parental accuracy on the mild video was significantly correlated with parental accuracy on the severe video for each item, though the magnitude of this correlation was often small (*r* ranged between .117 to .300). Yet, paired samples *t*-tests suggest that parents were less accurate rating the same items on the mild video versus severe video on all items except items 6 and 13 (see Table 3).

Sex and race differences across variables. Exploratory analyses were undertaken to explore differences in variables based on sex and race differences. Independent groups *t*-tests were used to determine whether there was a difference in mean accuracy on the M-CHAT-R ratings according to sex. Women were significantly more accurate on 3 of the 15 mild items (item 2: "*have you ever wondered if this child might be deaf*", item 7: "*does this child point with one finger to show something interesting*", and item 11: "*when the examiner smiles at this child, does he smile back at them*") and were more accurate on 11 of the 14 severe items (all items except item 6, item 9: "*does this child show you thing*", and 17: "*does this child try to get the examiner to watch him*"). Men were not significantly more accurate than women on any items. MANOVA was used to determine whether there were differences in accuracy based on race. A non-significant omnibus MANOVA revealed that M-CHAT-R items did not significantly vary as a function of race/ethnic group (Wilks' Lambda = .288, $p = .546$).

Independent groups *t*-tests were used to determine whether there were differences in any ASK-ASD, PSOC, or KCDI variables of interest (i.e. variables which were used in the final SEM model) according to sex. Women scored significantly higher on knowledge of emotional development ($t = 4.478$, $p < .001$) and knowledge of cognitive development ($t = 3.372$, $p = .001$)

on the KCDI, while men scored significantly higher on confidence in their ratings on the ASK-ASD (t = 2.438, p = .02). One-way ANOVA was used to determine whether there were any differences in ASK-ASD, PSOC, or KCDI variables based on race/ethnicity. The only significant one-way ANOVA for these variables of interest was on Parent Sense of Competence (measured by the PSOC; $F(7,280) = 2.753$, p = .009). Post-hoc Tukey's test revealed that this result was driven by differences between African American/Black and Asian/Pacific Islander parents, with African American/Black parents scoring higher in Parent Sense of Competence.

Item Response Theory

IRT results are described in Table 3. As with *t*-test results and mean accuracy scores, item difficulty was lowest on item 13 ("*does this child walk*") across both conditions. The probability of a parent of median ability rating item 13 the same as the clinician was greater than 96.9% across mild and severe conditions. The probability of a correct response by a parent of median ability ranged from .379 to .970 across items. With regard to item difficulty, 23 out of 29 items had a negative item difficulty parameter, suggesting that these items were either Very Easy or Easy difficulty and that the probability of a correct response by an individual with median ability is greater than 50%. No item had a difficulty parameter greater than 0.380, suggesting that while some items might be best classified as Average difficulty, no items in this set should be classified as either Hard or Very Hard difficulty (Baker & Kim, 2017).

Structural Equation Modeling

Measurement model. For the exogenous measurement model, good model fit was achieved, with a non-significant Chi-Square (p = .242), good RMSEA (.032; 90% Confidence Interval .000 to .080), good SRMR (.029). and good CFI (.990). Factor loadings on the 'Knowledge' factor were all strong (all factor loadings > .780; see Table 4). Factor loadings on

the ‘Parent’ factor were high for PSOC Sense of Competence (.528) and Education (-.406) but was low for Number of Children (.201). Number of Children was retained for theoretical purposes and due to the strong model fit of the hypothesized model. Several variables were considered for inclusion on the ‘Parent’ factor including parent sex, parent race, and whether or not the parent had any prior exposure to someone who has ASD. Inclusion of any of these three variables worsened model fit and yielded poor model fit overall, thus none of these three factors were included. Additionally, including only prior exposure into the model, via a separate factor for the specific prior exposures to someone with ASD (e.g. child, sibling, friend, through employment) loading onto a ‘Exposure’ factor, did not improve model fit, though model fit was still considered adequate. Given this information, the originally proposed exogenous measurement model was retained.

For the endogenous measurement model, the originally proposed model achieved adequate model fit based on RMSEA (.060; 90% Confidence Interval .054 to .067) and Robust CFI (.928). The chi-square test was significant for the endogenous measurement model ($p <$ .001) though the chi-square/df statistic (1.933) suggested good fit. The SRMR for the endogenous measurement model achieved poor model fit (.135). All factor loadings loaded stronger than .30 on either factor with the exception of item 13 on both factors (.197 on the ‘Mild’ factor; .260 on the ‘Severe’ factor; see Table 4). Due to the limited variability in this item (with greater than 93% of parents providing a correct response on these items), both item 13 variables (mild and severe) were removed from the model, leading to improved model fit via an increase in Robust CFI (.007) and a significant decrease in Chi-Square (Santorra-Bentler Scaled Chi-Square Difference (TRd) = 97.073, $\Delta df = 53$; $p < .001$). From there, several competing methods of modeling the two-factor structure of the M-CHAT-R items were nested for

comparison to ensure that the hypothesized two single-order factor structure provided the strongest fit for the data. The originally proposed model with mild items loading onto a 'Mild' factor and severe items loaded onto a 'Severe' factor was compared to a Bi-Factor (items load onto their respective factor and a general 'MCHAT' factor), General Factor (items do not load onto 'Mild' or 'Severe' factors and only load onto a single 'MCHAT' factor), and Second-Order factor model (the 'Mild' and 'Severe' factors load onto a second-order factor 'MCHAT'). Overall, model convergence was not achieved for the Bi-Factor or Second-Order factor models, and the fit of the General Factor model was inferior compared to the proposed model (χ^2 (324) = 1003.521, Robust CFI = .858).

Structure model. Overall, the proposed structure model ($\chi^2(514) = 785.006$, $p < .001$; see Figure 2) yielded poor (Robust CFI = .881; SRMR = .115) to good model fit (RMSEA = .046, 90% Confidence Interval .040 to .053; $\chi^2/df = 1.527$). SEM results are presented in matrix form in Table 5. Regarding path relationships, the Mild factor was significantly related to the Parent Knowledge factor ($\beta = .184$, $p = .004$), Number of Children ($\beta = .158$, $p = .020$), and Parental Self-Efficacy ($\beta = .202$, $p = .005$), but not Education or Confidence of ASD Knowledge. The Severe factor was significantly related to the Parental Knowledge factor ($\beta = .463$, $p < .001$), Education ($\beta = -.263$, $p < .001$), and Parental Self-Efficacy ($\beta = .513$, $p < .001$). The R-Squared for the Severe factor was .647, while the R-Squared for the Mild factor was .218.

CHAPTER 4

DISCUSSION

Summary of Findings

Overall, IRT results suggest that parents are able to accurately rate the majority of M-CHAT-R items with ease. While not all items are of Easy to Very Easy difficulty, there are no items that were considered Hard or Very Hard for parents to rate. The majority of items were easier for parents to rate accurately if the child they were rating had more severe symptoms of ASD. Across multiple items, women, relative to men, tended to be better able to rate a novel child on the M-CHAT-R.

In confirmatory factor analysis testing of the hypothesized measurement model, knowledge of cognitive development, emotional development, and ASD all loaded strongly and positively on the Parent Knowledge factor. The Parent Knowledge factor was positively related to more accurate ratings of both the child with more severe ASD and the child with milder ASD in the tested SEM model. Parent confidence in knowledge of ASD was not related to accuracy on ratings of either children in the hypothesized SEM model. Parental self-efficacy and number of children were both positively related to accuracy on ratings of the child with milder ASD, while parent education (negatively) and parent self-efficacy (positively) related to accuracy on ratings of the child with severe ASD. Thus, greater education was inversely related to accuracy in the severe condition, number of children was positively related to accuracy in the mild condition, and parent self-efficacy was positively related to accuracy in both conditions.

Sixty-five percent of the variance in accuracy on ratings of the child with more severe ASD was predicted by the tested model. However, this same model only accounted for approximately 22% of the variance in accuracy when considering ratings on the child with

milder ASD. Thus, while the majority of the variance in accuracy ratings of more severe ASD was well captured by hypothesized factors, only a small amount of variance was explained in ratings of milder ASD.

Interpretation of Findings

Excluded M-CHAT-R Items. An important point of interpretation surrounds the particular M-CHAT-R items that were excluded, specifically items 8 ("*is this child interested in other children*"), 12 ("*does this child get upset by everyday noises*"), and 20 ("*does this child like movement activities*") were each excluded. The exclusion of item 8 is unsurprising given no other children were involved in the videos (other than the one child being assessed). Each of these three items may be items that are easier for parents to identify (e.g., noticing that a child struggles with peers, gets upset around noises, or likes being swung often) relative to the more socially complex items retained in the current sample of items (e.g., responsive social smiling, quality of social response; Charwarska et al., 2007). Items 2 ("*have you ever wondered if this child might be deaf*"), 4 ("*does this child like climbing on things*"), and 5 ("*does this child make unusual finger movements near his eyes*") were only retained for the mild condition video, while items 10 ("*does this child respond when his name is called*") and 11 ("*when the examiner smiles at the child, does the child smile back*") were only retained for the severe condition video. The reason some of these items only appeared in one video compared to the other is likely due to variation in whether or not the behavior occurred specifically in the time frame of the video. For instance, items 4 and 5 could be rated as 'yes' easily if they occur, but could not necessarily be rated as 'no' if the behaviors did not occur in the specific time window captured by the videos. However, on the other hand, item 2 would be decipherable in a five-minute period and items 10 and 11 are specifically probed in the sections of the ADOS-2 included in the videos. Thus, some

of the items that were excluded in only one video may not have been excluded due to having a limited range to rate the behaviors, but rather speak to the difficulty in rating the items in a yes/no dichotomy.

Collectively, the M-CHAT maps items onto each social-communication symptom of ASD in the DSM-5 (e.g., A.1 social-emotional reciprocity – item 16: "*if the examiner turns their head to look at something, did the child look around to see what they were looking at*"; A.2 nonverbal communication – item 6: "*does this child point with one finger to show things*"; A.3 developing and maintaining relationships – item 8: "*is this child interested in other children*") and maps items onto two restricted and repetitive behaviors and interests (RRBI's) symptoms of ASD in the DSM-5 (e.g., B.1 repetitive motor movements – item 20: "*does this child like movement activities*"; B.4 sensory abnormalities – item 12: "*does this child get upset by everyday noises*"). However, two items that target RRBI's were excluded from the sample pool of items (items 12 & 20) in both conditions, and the remaining item covering RRBI's (item 5: "*does this child make unusual finger movements near his eyes*") was only included in the Mild condition. Therefore, even though the M-CHAT-R specifically aims to focus more on social behaviors (Robins et al., 2014; Hampton & Strand, 2015), the current pool of items used does not contain adequate coverage of parental ability to identify RRBI's.

Of the items excluded from the Severe condition, but not the Mild condition, two items (items 2: "*have you ever wondered if this child might be deaf*" and 5: "*does this child make unusual finger movements near his eyes*") were not flagged as concerns on the M-CHAT-R rating by the clinicians during the Mild condition. Of the two items that were excluded from the Mild condition, but not the Severe condition, both items (items 10: "*does this child respond when the examiner calls his name*" and 11 "*when the examiner smiles at the child, does the child*

smile back") were flagged as concerns on the M-CHAT-R ratings by the clinicians during the Severe condition. Thus, items 10 and 11 may be items which parents can identify as atypical when the behaviors are absent, but may be less able to rate as normative when the behaviors are present and appear normative during a small window of time (i.e., these behaviors would need to be observed more frequently and in more contexts to truly classify them as normative).

Hypothesis 1a. Hypothesis 1a, that parent ratings would vary from clinician ratings on the M-CHAT-R when rating the same child, was partially supported by IRT analyses. Only two items had probability parameters greater than .90, suggesting that only one item (item 13 in both videos) had a 90% chance of a parent of average ability providing the same rating as the clinicians about a novel child on the M-CHAT-R. However, the majority of items were easy or very easy for parents to rate, and only six items were of average difficulty. Furthermore, the majority of items had negative difficulty parameters, suggesting there was a greater chance that the parent would provide a rating that matched the clinician rating than provide the opposite rating. Therefore, parents were able to identify many symptoms of ASD in a novel child consistent with a pair of clinician ratings on the child; however, parent ratings were not an exact match to clinician ratings. This leaves room for variance in accuracy to be explained by external factors and provides justification for the use of SEM to explore factors that predict greater parental accuracy in identifying symptoms of ASD.

When considering specific items, most items were accurately rated by 72% to 78% of parents when rating the Severe condition. Parents struggled more with items related to pointing and showing (range: 64% to 67%) but did better with the item related to whether or not the child walks (93%). When considering the Mild condition, most items were accurately rated by 45% to 55% of parents. Two items were rated below 45%, with these items focused on showing and how

well a child follows shift in gaze (both items had 42% accuracy). Items related to hearing, climbing, finger posturing, and walking had accuracies above 70% in the Mild condition. Thus, while most items in the Severe condition were rated at above 70% accuracy, only items related to motor movements (e.g., walking or climbing), or hearing, remained at or above 70% accuracy in the Mild condition. Thus, it appears that when considering subtler symptoms of ASD, parents are better at rating items which are explicitly observable and, in some cases, not specific to ASD (e.g., a child walking). Conversely, parents may struggle with more socially complex symptoms (e.g., how well a child follows gaze shift). Previous literature suggests that parents tend to be more accurate when rating communication behaviors, but struggle with more social-emotional behaviors (e.g., eye contact, social smiling, facial expressions; Charwarska et al., 2007). The results of this study corroborate these findings using a new paradigm testing parental accuracy in reporting or identifying symptoms.

Unexpectedly, three of the four items that were rated at 70% accuracy or higher in the Mild condition were the three items that were only included in the Mild condition. Two of these three items were not flagged positive by the clinicians. Thus, these two items (items 2: "*have you ever wondered if this child might be deaf*" and 5: "*does this child make unusual finger movements near his eyes*") may be items which can easily distinguish a child as developmentally typical when these concerns/behaviors are absent, but may be difficult to distinguish a child as atypical when these concerns/behaviors are present if only based on a 5 minute period.

The majority of items were easy or very easy to rate, and even the items that were relatively more difficult to rate were not difficult (i.e. hard) to rate overall. This provides some added confidence in the psychometric properties of the M-CHAT-R. It is important to note, however, that the way in which the M-CHAT-R was used in this study is vastly different from

how parents rate the M-CHAT-R in practice. Albeit, the current psychometric base has solely focused on how the M-CHAT-R performs during natural use. This study is the first to take an experimental approach to measuring the psychometric properties of the M-CHAT-R with the added benefit of experimental control (i.e., what symptoms the parents see) over naturalistic features (i.e., parents rated a child they would not have normally rated). Therefore, when interpreting these results, it is highly important to consider the current literature on the M-CHAT-R in its natural use.

Hypothesis 1b. Overall, the majority of items were more difficult for parents to rate on the child who displayed subtler, more mild symptoms of ASD compared to the child with more severe and overt symptoms of ASD. This provides support for Hypothesis 1b, that rating the M-CHAT-R items on a child who experiences more severe, overt symptoms would be easier for parents than rating the M-CHAT-R items on a child who experiences subtler symptoms of ASD. This suggests that there is a sizeable difference in parents' ability to detect milder presentations of ASD versus more overt presentations of ASD. However, this may also be an extraneous difference resulting from the brevity of the videos shown. Perhaps five minutes of exposure is enough time for a parent to accurately detect symptoms of ASD in a child who more clearly exhibits symptoms. Yet, for children who have more complex or more subtle symptoms, the time required to rate these symptoms may be greater. For the two clinicians, though, the number of items that they agreed on for the mild and severe presentations were similar (14/20 versus 15/20, respectively). It is also important to note that the description of "mild" and "severe" is based on ADOS-2 score differences and may not fully reflect the wide heterogeneity found in ASD. Thus, it is important to consider these findings as being specific to the two levels of ASD severity presented in this study. Moreover, a larger number of videos would be needed to build

generalizations on how accurate parents are in rating the M-CHAT-R items across the full range of ADOS-2 scores.

While not hypothesized as a predictor of parental accuracy, sex differences arose in the accuracy of parent ratings of a novel child on the M-CHAT-R compared to clinician ratings. Sex differences in accuracy may have been driven by women having greater knowledge of child development (which was positively related to accuracy) relative to men, while men reported greater confidence in their knowledge of ASD (which was negatively related to accuracy) relative to women. Historical gender norms have typically favored greater focus on mothers' attentiveness toward child development norms relative to that of fathers (Bartlett, Guzman, & Ramos-Olazagasti, 2018), including medical providers giving greater guidance to mothers over fathers during well-child visits (Garfield & Isacco, 2006). Furthermore, mothers tend to present more often to early childhood pediatrician visits than fathers (Garfield & Isacco, 2006), and this may provide a focal point for mothers to refine their knowledge on child development norms. Greater confidence in fathers about their knowledge of ASD (though no significant difference between fathers and mothers in actual knowledge of ASD arose) may stem from gender norm expectations for men to display confidence and assertiveness (Leaper & Smith, 2004). This effect is generally present in multiple areas of knowledge, known as the Dunning-Kruger effect (Ehrlinger & Dunning, 2003). Nevertheless, when sex was added to the hypothesized SEM model (which obtained good model fit), the resulting model did not converge, which introduces need for some caution in interpreting sex differences in identifying symptoms of ASD.

Hypothesis 2. Overall, the hypothesized model obtained good model fit based on several fit indices for both measurement models (exogenous factors and endogenous factors) and the structural model. Alternative modeling of the M-CHAT-R factors (i.e. the endogenous

measurement model), including general factor, bi-factor, and second-order factor models, did not fit the data better than the originally hypothesized two single-order factor model. Relatedly, addition of exploratory predictor variables (parental age, sex, race, or prior exposure to ASD) into the structural model did not improve model fit and typically worsened model fit. Based on R-Squared, the hypothesized predictor variables accounted for more variability in the accuracy of parent's ratings of the child with more severe ASD compared to accuracy of the child with milder ASD.

Hypothesis 2a. Parental knowledge of ASD, emotional development, and cognitive development all loaded onto the 'Knowledge' factor strongly and positively. The Knowledge factor was positively and significantly related to both the Severe factor and the Mild factor. This suggests that parental knowledge played a useful role in accurately identifying symptoms of ASD (in a vignette of a child with ASD). However, parental knowledge played a smaller role in a parent's ability to identify symptoms of ASD (in a vignette of a child with ASD) when symptoms presented more subtly or milder. Thus, Hypothesis 2a (that parental knowledge of ASD and child development would relate to accuracy in rating a novel child on M-CHAT-R items) was supported.

Whether this reflects that milder symptoms of ASD are more difficult for parents to identify, or whether milder symptoms of ASD are less descriptive of ASD is an important differential consideration. An existing body of literature suggests that subtler ASD symptoms can also be present in siblings of children with ASD, children with ADHD, children with language delays, and children with disruptive behavior concerns (Buhler et al., 2011; de Bildt et al. 2009; Nijmeijer et al. 2009; Geluk et al., 2012; Reisinger, Cornish, & Fombonne, 2011). Thus, it may be that children who display only subtler symptoms of ASD are more difficult to

differentiate from these other groups of children. For instance, children with language impairments tend to show similar symptoms of ASD, but at milder levels (Reisinger et al., 2011). Conversely, one of the major criticisms of the M-CHAT-R is that, in addition to children with ASD, it will often detect children with some form of developmental delay other than ASD (Toh et al., 2017). Thus, the lower accuracy in rating children with less severe ASD could be reflective of the complex nature of differential diagnosis in milder presentations of ASD (Dickerson-Mayes, Black, & Tierney, 2013). Research suggests that, under the DSM-IV classifications, there was less research on PDD-NOS (which was typically defined by less severe presentations of autism) compared to research on Asperger's syndrome or autistic disorder (Matson & Boisjoli, 2007). Thus, even parents who are highly educated about ASD may have less insight into how milder forms of ASD may appear versus more severe, clearer presentations.

Hypothesis 2b. The proposed observed predictor variables each were related to accuracy for at least one of the factors (mild or severe). Unsurprisingly, greater parental self-efficacy was related to both the mild and severe factors such that greater parental self-efficacy was related to greater accuracy. Parental self-efficacy, as it relates to parental behavior is well and widely studied (for a review, see Coleman & Karraker, 1998). High parental self-efficacy influences parental behaviors through greater motivation, engagement, and action in parenting (Coleman & Karraker, 1998), and likely influences parents' attentiveness to a child's development. Thus, parents who report greater self-efficacy may be more likely to be actively engaged in noticing a child's behavior and better able to report on abnormal patterns of behavior.

The number of children the parent raised was positively related to accuracy in the Mild condition, while parental education was negatively related to accuracy in the Severe condition. Thus, when rating the clearer presentation of ASD, parents who had greater education tended to

provide less accurate ratings on M-CHAT-R items compared to parents who had less education. This finding occurred while partialling out the variance in accuracy explained by parental knowledge and parental confidence. This result is counter-intuitive and difficult to explain. However, some research on parental expectations and academic achievement may provide some insight into this contradictory finding. Alexander, Entwisle, and Bedinger (1994) found that parents of higher educational attainment tended to have achievement expectations that more closely align with their child's abilities when compared to parents with less educational attainment, who often had expectations that exceeded their child's ability. Parents with less educational attainment may thereby have more rigid expectations regarding their children's behavior, which may lead them to more readily flag behavior characteristic of ASD as abnormal. Future research is needed to more thoroughly explore the mechanism explaining this relation and should particularly explore how the M-CHAT-R performs specifically in parents of greater educational attainment. Although unknown, the specific area of study of those parents with college degrees could provide greater insight into this finding (e.g. do parents who studied psychology perform better than parents who received a degree in business?).

The number of children the parent had previously raised was positively related to accuracy in the mild condition only. This suggests that prior experience raising children conveys some benefit in identifying symptoms of ASD that may be subtler to detect. While knowledge and self-efficacy, apart from prior experiences raising a child, may be enough to accurately rate symptoms of ASD when the symptoms are more clear, first-hand experience of raising children may help in rating subtler symptoms that may be more difficult to distinguish. Thus, experience may be helpful in distinguishing atypical behavior from normative behavior, when knowledge of normative behavior alone may not be enough to accurately understand a child's behavior (e.g.,

when symptoms are less clear or subtler).

Synthesizing these findings, broadly, Hypothesis 2b (that greater parenting self-efficacy, greater parental education, and greater number of children older than 48 months would all significantly predict greater accuracy in identifying symptoms of ASD) was partially supported. Considering specific variables, only parental self-efficacy performed as expected (positively related to accuracy in both conditions). While the directionality of the effect of the number of children the parent raised was in line with expectations, this variable was only related to accuracy in the Mild condition. On the other hand, parental education performed in the opposite direction than expected. Thus, Hypothesis 2b was broadly supported in that variables capturing parental factors beyond knowledge related to accuracy; however, only one of three variables related in the expected direction consistently.

Hypothesis 2c. Hypothesis 2c (that parental confidence would predict accuracy in identifying symptoms of ASD beyond what was already explained by parental knowledge) was not supported. Parental confidence was not related to accuracy. Because ASD knowledge, as well as parental self-efficacy, were included in the hypothesized SEM moJRdel, the effect of parental confidence on parental accuracy is considering the effect of greater knowledge and greater self-efficacy on parental accuracy already partialled out. The correlation between confidence in knowledge of ASD symptoms and actual knowledge of ASD symptoms was non-significant and neared a correlation of zero. Given that parental self-efficacy was highly related to accuracy, it is plausible that confidence in ability as a parent is a stronger driver of parental behavior and decision making than their declarative confidence in their knowledge of certain aspects of child development. Another point may be that the confidence scale of the ASK-ASD ("not at all confident", "confident", or "very confident" response options) is too restrictive to

accurately model confidence.

Feasibility of Methodology

Based on overall content validity ratio, among the 40 items (20 M-CHAT-R items x 2 children), 29 items were rated the same by the experts. Thus, a majority of the items were able to be retained and were able to be rated by the experts through the components of the ADOS-2 in the videos shown. This is somewhat unsurprising given the purpose of the ADOS-2 is to serve as a social press to elicit behaviors characteristic of ASD. However, this is simultaneously a bit surprising given the relatively limited length of time the parent is able the view the child in the videos. Generally, this provides preliminary support for the use of vignettes to display behaviors characteristic of ASD to examine parental ability to recognize symptoms of ASD in research studies. However, obviously these vignettes do not present the amount of information a parent sees in their own child, particularly considering the variety of contexts in which a parent views a child's behavior. Therefore, it is helpful to consider these vignettes in the context of parents identifying a discrete behavior, rather than rating the frequency or quality of a behavior (i.e. these videos would not allow a parent to rate questions about how often a child exhibits a behavior or how problematic a certain behavior may be). For simple items (e.g. does this child walk?), a single observation in a five-minute period would provide enough information to rate the item. However, for ASD-related behaviors that may occur infrequently, but are still diagnostic of ASD (e.g., insistence upon sameness, difficulty with transitions), five minutes may not provide enough time for the behavior to occur.

One of the strengths of the M-CHAT-R is its focus on asking about behaviors which do not require extensive recall of overall child behavior and can be rated on behaviors that have occurred within a recent amount of time. It is well accepted that behaviors that are objective and

salient are far easier for parents to rate accurately compared to behaviors which are subtle or leave room for subjective appraisal (e.g., Mattsson, Forsner, & Arman, 2011).

Limitations

Several important limitations require mention. First, only two examples of ASD were used and thus a limited range of variation in symptoms of ASD were featured. On this note, the child in the video for the 'Mild' condition was scored on the ADOS-2 as being in the moderate range rather than the mild range. While this score was still at the lower-end of the scale for 'Autism' on the ADOS-2, this video may not fully capture the most common presentation of 'milder' ASD. Second, a video with a child without ASD was not shown as a control condition, as this would have allowed examination of how much parental over-reporting of symptoms may occur. Third, only Caucasian males were featured in video stimuli, thus limiting the generalizability of findings. Fourth, and perhaps most importantly, the video stimuli lack considerable ecological validity because the M-CHAT-R was designed for parents to rate their own child. Specifically, the M-CHAT-R is designed to probe behaviors that occur across contexts and across time. Five minutes does not provide the same amount of opportunities for a behavior characteristic of ASD to occur compared to the amount of opportunities a parent has to rate their own children. Thus, this small window may not present enough time for some atypical behaviors to occur. For instance, a study by Gabrielsen et al. (2015) showed clinicians 10-minute video samples using methods similar to the current study and found that children with ASD showed normative behavior in 89% of the time in the videos. As a result, Gabrielsen et al. (2015) found that expert raters missed 39% of children with ASD in these videos. With that said, a substantial amount of the unexplained variance in ratings (10% in Severe ratings; 64% in Mild ratings) may be due to this phenomenon rather than systematic random error. Fourth, the experts,

while having experience and highest qualifications in working with individuals with ASD, may have not provided accurate ratings across all items due to the novel task being asked of them.

Beyond the video stimuli, several other methodological limitations exist. First, the KCDI was developed more than 30 years ago and has not been updated since. While it is likely that child cognitive and emotional development has largely remained unchanged over this period, changes in socio-cultural norms surrounding normative development may have changed over time. Second, while the ASK-ASD has been validated by previous research (Hansen, 2015), it does not possess established population norms. Thus, the current study cannot conclude whether this sample has high or low knowledge of ASD relative to the general population. Third, several limitations exist given the online data collection method employed for sample recruitment in the current study. This study sample consisted only of participants who complete surveys through online survey panels. Therefore, even though the sample was recruited specifically to recruit a diverse sample and have an even split between men and women, our sample may be biased in several ways due to the influence of a select group of the population being recruited. Specifically, this sample consists of people who have regular access to the internet, meaning that families in more rural areas where internet is less accessible or from lower SES backgrounds may be underrepresented. Secondly, our study also is biased in that families unable to dedicate time to complete our survey were not included, and we cannot speculate how these relations may play out in parents who would not have the time needed to take an online survey. Third, parents did not complete study measures in a controlled setting, and parents may have been distracted while completing the survey.

At least three limitations exist with regard to the current sample. The sample was not directly representative of the general population, with some racial groups (Asian/Pacific

Islanders and African-Americans) being oversampled and other groups being underrepresented (Caucasians and Hispanic/Latinos) relative to the general US population. This sample also was more highly educated than the general population. Lastly, the sample appeared to have a high degree of prior exposure to ASD. While it is not known what percentage of the general population has previous exposure to ASD, it is highly possible that a greater portion of parents who had prior exposure participated in this study. The current sample may then reflect a group that was more highly motivated to take part in a study about ASD and who may possess more knowledge of and sensitivity to identifying symptoms of ASD.

Several statistical limitations also exist. First, the sample size in the current study was enough to test the hypothesized model but did not allow for more sophisticated models to be tested (e.g., non-linear interactions between knowledge and confidence could not be modeled). Second, only the M-CHAT-R was used, and these items represent only a small number of symptoms indicative of ASD and represents only one of the ASD screeners currently in use. Thus, the use of latent trait (in IRT) and latent factor (in SEM) models instead of simple observed data was done because these items only represent a limited sample of ASD behaviors. Third, the SEM model obtained good RMSEA, but had poor CFI and SRMR, thus the model requires improvement and refinement and presently should be interpreted with some degree of caution. Lastly, as the response options were either yes/no, parents were unable to identify times when they guessed on an item. Thus, some degree of correct responses may be more attributable to guessing, and the IRT parameters may overestimate parents' ability to accurately identify symptoms of ASD.

Future Directions

The present study provides multiple avenues for future exploration. First, the current

study should be replicated with a larger sample and more diverse stimulus. For instance, a study with more video trials covering children of different ages, races, sexes, and severity of ASD could further develop a profile of how the method employed in this study performs under different conditions. A specific area of improvement could be examining if parents identify symptoms of ASD better in children of the similar versus different demographic backgrounds. Specifically, parents may have better understanding of normative development, and thereby what is atypical development, within their own culture relative to other cultures. Given the higher rate of ASD in boys (Baio et al., 2018), whether symptoms are more difficult to identify in girls would be a reasonable question for future studies to explore.

A more comprehensive study would also explore how parents rate their own children on the M-CHAT-R in addition to a novel child. For instance, a study could have parents complete the same video ratings present in this study in addition to rating their own child on the M-CHAT-R. Clinicians could then administer an ADOS-2 and separate experts could score the child on the M-CHAT-R based on the same parts of the ADOS-2 selected in the current study. However, such a study would likely require significant resources to achieve the sample size required to complete statistical approaches similar to those used in the current study.

Given established structural paths found in this study, examination of mediational pathways is warranted. Specifically, surprising findings linking lower educational attainment and greater accuracy might be better explained through covariance or indirect relationships (e.g., through focus on career versus family development) not considered in the current study.

Lastly, whether parental accuracy in detecting symptoms of ASD is related to parental ability to detect symptoms of other psychopathologies (e.g., symptoms of ADHD, symptoms of anxiety) would help determine if efforts to increase parental knowledge and recognition of

symptoms needs to occur specifically related to ASD or more broadly (e.g., typical vs atypical development).

Perhaps most significantly, given results suggesting that rating milder symptoms of ASD may be more difficult, more research is needed to better understand how to identify and correctly classify children who experience only mild symptoms of ASD. This area of research remains understudied and the results of this study underscore the need for increased future research in this area.

Clinical & Research Implications

These findings present several important implications for research and clinical practice alike. Most strikingly, given the vast difference in explained variance between the mild and severe factors, in addition to severe items being significantly easier than milder items, some refinement of our current screening initiatives should be made. Screening has routinely focused on identifying ASD, however, there may be two separate screening related questions inherent within this focus. First, how can we identify clear-cut examples of ASD most efficiently, and, secondly, how can we differentiate mild ASD from other conditions that may present somewhat similarly (e.g., language delays, global developmental delay, behavioral concerns). This project adds some level of validity to the performance of the M-CHAT-R in screening for more clear cases of ASD (e.g., severe presentations). However, where the M-CHAT-R struggled was in the cases which presented more mildly. Much of the criticism in the literature on the M-CHAT-R surrounds its oversensitivity in detecting other developmental concerns (Kim et al., 2016). While the focus of screening most often emphasizes the principles of sensitivity and specificity, the field may benefit from more focus on 1) identifying clear ASD and 2) identifying children who need additional assessment to determine proper diagnoses (e.g., identifying the children who are

experiencing symptoms that *may* or may not be ASD). From there, screening and detection efforts may benefit from a two-pronged approach, where children showing clear evidence of ASD are referred to proper providers immediately, and children who fit this specific presentation are provided with greater follow-up from providers to gather greater information on the symptoms present. In comparison, current screening procedures often result in broad referral of any children who score high on the M-CHAT-R for specialized ASD assessment, and often children who do not score above clinical cut-off are typically not referred.

A second consideration is that increasing parental knowledge about ASD seems to be a meaningful endeavor, however, this should not be viewed as the absolute way to increase screening accuracy. Knowledge plays a role in identifying more clear, overt cases of ASD, to ensure parents understand that the behaviors they are seeing are due to ASD. However, increasing knowledge may not help to solve gaps in identifying children who are experiencing subtle signs of ASD that may also appear with other developmental concerns.

A last consideration is to ensure that clinicians are routinely considering the characteristics of the parent who is filling out the M-CHAT-R or other screening measures. Given that parental characteristics were the only factor related to accuracy in rating the subtler, more difficult to identify, case of ASD in the present study, considering factors related to the parent such as their previous experiences in parenting or their overall self-confidence are important to consider, particularly when a parent provides a rating on the M-CHAT-R that falls on the milder end of the scale.

Conclusions

Overall, the majority of M-CHAT-R items appear easy for parents to accurately rate in a novel child based on a five-minute video stimulus. This provides support for the M-CHAT-R in

its current practice as a widespread screener for ASD. Generally speaking, parents struggled more in identifying symptoms of ASD in a child with milder ASD compared to a child with more severe ASD. This may suggest that parents are more easily able to identify symptoms when they are overt, and struggle to accurately identify symptoms when they are more subtle. Greater parental knowledge about child development and ASD was related to greater accuracy in identifying symptoms, but only when rating the child with more severe ASD. Parental characteristics (specifically lower education and greater self-efficacy), on the other hand, were related to accuracy in ratings of both children. Taken together, clearer presentations of ASD were easier for parents to identify accurately, particularly for parents who had higher levels of knowledge about ASD. However, it was more difficult for parents to identify symptoms of ASD that were subtler, and parental knowledge of ASD and child development did not aid in identifying these milder symptoms of ASD. Results highlight a potential divide in current screening practices, wherein severe and more clear cases of ASD may be easily detected, but milder cases of ASD may not be captured as well by current screening methods. Considering these two separate presentations of ASD, and differentiating screening of each, may help ensure that more clear cases of ASD are detected efficiently, and milder cases are more closely examined by providers.

References

Achenbach, T. M., & Edelbrock, C. S. (1981). Behavioral problems and competencies reported by parents of normal and disturbed children aged four through sixteen. *Monographs of the Society for Research in Child Development*, *46*(1), 1-82.

Alexander, K. L., Entwisle, D. R., & Bedinger, S. D. (1994). When expectations work: Race and socioeconomic differences in school performance. *Social Psychology Quarterly, 57*(4), 283-299.

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

Baio, J., Wiggins, L., Christensen, D. L., Maenner, M. J., Daniels, J., Warren, Z., ... & Durkin, M. S. (2018). Prevalence of autism spectrum disorder among children aged 8 years—Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2014. *MMWR Surveillance Summaries*, *67*(6), 1-23.

Baker, F. B., & Kim, S. H. (2017). *The Basics of Item Response Theory Using R*. Springer: Cham, Switzerland.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191-215.

Baranek, G. T. (1999). Autism during infancy: A retrospective video analysis of sensory-motor and social behaviors at 9–12 months of age. *Journal of Autism and Developmental Disorders*, *29*(3), 213-224.

Bartlett, J. D., Guzman, L., & Ramos-Olazagasti, M. A. (2018). *Parenting knowledge among first-time parents of young children*. Bethesda, MD: Child Trends.

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*(6), 1165-1174.

Bühler, E., Bachmann, C., Goyert, H., Heinzel-Gutenbrunner, M., & Kamp-Becker, I. (2011). Differential diagnosis of autism spectrum disorder and attention deficit hyperactivity disorder by means of inhibitory control and 'theory of mind'. *Journal of Autism and Developmental Disorders*, *41*(12), 1718-1726.

Canal-Bedia, R., García-Primo, P., Martín-Cilleros, M. V., Santos-Borbujo, J., Guisuraga-Fernández, Z., Herráez-García, L., & ... Posada-de la Paz, M. (2011). Modified checklist for autism in toddlers: Cross-cultural adaptation and validation in Spain. *Journal of Autism and Developmental Disorders*, *41*(10), 1342-1351.

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, *36*(5), 648-662.

Charman, T., & Baird, G. (2002). Practitioner review: Diagnosis of autism spectrum disorder in 2-and 3-year-old children. *Journal of Child Psychology and Psychiatry*, *43*(3), 289-305.

Chawarska, K., Paul, R., Klin, A., Hannigen, S., Dichtel, L. E., & Volkmar, F. (2007). Parental recognition of developmental problems in toddlers with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37*(1), 62-72.

Chlebowski, C., Robins, D. L., Barton, M. L., & Fein, D. (2013). Large-scale use of the Modified Checklist for Autism in low-risk toddlers. *Pediatrics*, *131*(4), e1121-e1127.

Chronis, A. M., Chacko, A., Fabiano, G. A., Wymbs, B. T., & Pelham, W. E. (2004). Enhancements to the behavioral parent training paradigm for families of children with ADHD: Review and future directions. *Clinical Child and Family Psychology Review*, *7*(1), 1-27.

Coleman, P. K., & Karraker, K. H. (1998). Self-efficacy and parenting quality: Findings and future applications. *Developmental Review*, *18*(1), 47-85.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston: Orlando, FL.

Daniels, A. M., Halladay, A. K., Shih, A., Elder, L. M., & Dawson, G. (2014). Approaches to enhancing the early detection of autism spectrum disorders: A systematic review of the literature. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(2), 141-152.

Daniels, A. M., & Mandell, D. S. (2014). Explaining differences in age at autism spectrum disorder diagnosis: A critical review. *Autism*, *18*(5), 583-597.

de Bildt, A., Mulder, E. J., Hoekstra, P. J., van Lang, N. D., Minderaa, R. B., & Hartman, C. A. (2009). Validity of the Children's Social Behavior Questionnaire (CSBQ) in children with intellectual disability: Comparing the CSBQ with ADI-R, ADOS, and clinical DSM-IV-TR classification. *Journal of Autism and Developmental Disorders*, *39*(10), 1464-1470.

De Giacomo, A., & Fombonne, E. (1998). Parental recognition of developmental abnormalities in autism. *European Child & Adolescent Psychiatry*, *7*(3), 131-136.

De Los Reyes, A. (2011). Introduction to the special section: More than measurement error: Discovering meaning behind informant discrepancies in clinical assessments of children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *40*(1), 1-9.

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858-900.

De Los Reyes, A., Henry, D. B., Tolan, P. H., & Wakschlag, L. S. (2009). Linking informant discrepancies to observed variations in young children's disruptive behavior. *Journal of Abnormal Child Psychology*, *37*(5), 637-652.

Dickerson-Mayes, S., Black, A., & Tierney, C. D. (2013). DSM-5 under-identifies PDDNOS: Diagnostic agreement between the DSM-5, DSM-IV, and Checklist for Autism Spectrum Disorder. *Research in Autism Spectrum Disorders*, *7*(2), 298-306.

Dosreis, S., Weiner, C. L., Johnson, L., & Newschaffer, C. J. (2006). Autism spectrum disorder screening and management practices among general pediatric providers. *Journal of Developmental and Behavioral Pediatrics*, *27*(2), S88-S94.

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 Research Reviews*, *11*(3), 253-262.

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology*, *84*(1), 5-17.

Ertem, I. O., Atay, G., Dogan, D. G., Bayhan, A., Bingoler, B. E., Gok, C. G., ... & Isikli, S. (2007). Mothers' knowledge of young child development in a developing country. *Child: Care, Health and Development*, *33*(6), 728-737.

Filipek, P. A., Accardo, P. J., Ashwal, S., Baranek, G. T., Cook, E. J., Dawson, G., & ... Volkmar, F. R. (2000). Practice parameter: Screening and diagnosis of autism: Report of the Quality Standards Subcommittee of the American Academy of Neurology and the Child Neurology Society. *Neurology*, *55*(4), 468-479.

Foley-Nicpon, M., Fosenburg, S. L., Wurster, K. G., & Assouline, S. G. (2017). Identifying high ability children with DSM-5 autism spectrum or social communication disorder:

performance on autism diagnostic instruments. *Journal of Autism and Developmental Disorders*, *47*(2), 460-471.

Fountain, C., King, M. D., & Bearman, P. S. (2011). Age of diagnosis for autism: individual and community factors across 10 birth cohorts. *Journal of Epidemiology & Community Health, 65*(6), 503-510.

Gabrielsen, T. P., Farley, M., Speer, L., Villalobos, M., Baker, C. N., & Miller, J. (2015). Identifying autism in a brief observation. *Pediatrics*, *135*(2), 330-338.

Garfield, C. F., & Isacco, A. (2006). Fathers and the well-child visit. *Pediatrics*, *117*(4), 637-645.

Geluk, C. A., Jansen, L., Vermeiren, R., Doreleijers, T. A., van Domburgh, L., de Bildt, A., ... & Hartman, C. A. (2012). Autistic symptoms in childhood arrestees: longitudinal association with delinquent behavior. *Journal of Child Psychology and Psychiatry*, *53*(2), 160-167.

Gura, G. F., Champagne, M. T., & Blood-Siegfried, J. E. (2011). Autism spectrum disorder screening in primary care. *Journal of Developmental & Behavioral Pediatrics*, *32*(1), 48-51.

Hampton, J., & Strand, P. S. (2015). A review of level 2 parent-report instruments used to screen children aged 1.5–5 for autism: A meta-analytic update. *Journal of Autism and Developmental Disorders*, *45*(8), 2519-2530.

Hansen, L. K. (2015). *Development and validation of a survey of knowledge of autism spectrum disorder* (Unpublished master's thesis). The University of Southern Mississippi, Hattiesburg, MS.

Hartman, C. A., Rhee, S. H., Willcutt, E. G., & Pennington, B. F. (2007). Modeling rater

disagreement for ADHD: Are parents or teachers biased?. *Journal of Abnormal Child Psychology*, *35*(4), 536-542.

Hassan, E. (2006). Recall bias can be a threat to retrospective and prospective research designs. *The Internet Journal of Epidemiology*, *3*(2), 339-412.

Herlihy, L. E., Brooks, B., Dumont-Mathieu, T., Barton, M. L., Fein, D., Chen, C., & Robins, D. L. (2014). Standardized screening facilitates timely diagnosis of autism spectrum disorders in a diverse sample of low-risk toddlers. *Journal of Developmental and Behavioral Pediatrics*, *35*(2), 85-92.

Hertz-Picciotto, I., & Delwiche, L. (2009). The rise in autism and the role of age at diagnosis. *Epidemiology, 20*(1), 84–90.

Howells Wrobel, N., & Lachar, D. (1998). Validity of self-and parent-report scales in screening students for behavioral and emotional problems in elementary school. *Psychology in the Schools*, *35*(1), 17-27.

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*(1), 330-336.

Jacob, T., & Johnson, S. L. (1997). Parent–child interaction among depressed fathers and mothers: Impact on child functioning. *Journal of Family Psychology*, *11*(4), 391-409.

Johnston, C., & Mash, E. J. (1989). A measure of parenting satisfaction and efficacy. *Journal of Clinical Child Psychology*, *18*(2), 167-175.

Johnson, C. P., & Myers, S. M. (2007). Identification and evaluation of children with autism spectrum disorders. *Pediatrics*, *120*(5), 1183-1215.

Jones, T. L., & Prinz, R. J. (2005). Potential roles of parental self-efficacy in parent and child

adjustment: A review. *Clinical Psychology Review*, *25*(3), 341-363.

Karp, S. M., Lutenbacher, M., & Wallston, K. A. (2015). Evaluation of the parenting sense of competence scale in mothers of infants. *Journal of Child and Family Studies*, *24*(11), 3474-3481.

Kazdin, A. E. (2009). *The Kazdin method for parenting the defiant child*. Houghton Mifflin Harcourt: Boston, MA.

Kim, S. H., Joseph, R. M., Frazier, J. A., O'Shea, T. M., Chawarska, K., Allred, E. N., & ... Kuban, K. K. (2016). Predictive validity of the Modified Checklist for Autism in Toddlers (M-CHAT) born very preterm. *The Journal of Pediatrics*, *178,* 101-107.

Kline, R.B. (2005), *Principles and Practice of Structural Equation Modeling* (2nd Edition ed.). New York: The Guilford Press.

Larsen, J. J., & Juhasz, A. M. (1986). The knowledge of child development inventory. *Adolescence, 21*(81), 39-54.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, *28*(4), 563-575.

Le Couteur, A., Haden, G., Hammal, D., & McConachie, H. (2008). Diagnosing autism spectrum disorders in pre-school children using two standardized assessment instruments: the ADI-R and the ADOS. *Journal of Autism and Developmental Disorders*, *38*(2), 362-372.

Leaper, C., & Smith, T. E. (2004). A meta-analytic review of gender variations in children's language use: talkativeness, affiliative speech, and assertive speech. *Developmental Psychology*, *40*(6), 993-1027.

Lee, J. Y. (2009). The relationships between parenting knowledge and parenting style of mothers with infants: The mediating effect of parenting efficacy. *Journal of the Korean Home*

Economics Association, *47*(5), 35-48.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936-949.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule, second edition (ADOS-2) manual: Modules 1–4. *Torrance, CA: Western Psychological Services*.

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*(8), 822-829.

Matson, J. L., & Boisjoli, J. A. (2007). Differential diagnosis of PDDNOS in children. *Research in Autism Spectrum Disorders*, *1*(1), 75-84.

Matson, M. L., Mahan, S., & Matson, J. L. (2009). Parent training: A review of methods for children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *3*(4), 868-875.

Mattsson, J. Y., Forsner, M., & Arman, M. (2011). Uncovering pain in critically ill non-verbal children: Nurses' clinical experiences in the paediatric intensive care unit. *Journal of Child Health Care, 15*(3), 187-198.

McPheeters, M. L., Weitlauf, A., Vehorn, A., Taylor, C., Sathe, N. A., Krishnaswami, S., ... & Warren, Z. E. (2016). *Screening for autism spectrum disorder in young children: A systematic evidence review for the U.S. Preventive Services Task Force* (AHRQ Publication No. 13-05185-EF-1). Rockville, MD: Agency for Healthcare Research and Quality.

Nijmeijer, J. S., Hoekstra, P. J., Minderaa, R. B., Buitelaar, J. K., Altink, M. E., Buschgens, C. J. M., ... & Hartman, C. A. (2009). PDD symptoms in ADHD, an independent familial trait?. *Journal of Abnormal Child Psychology*, *37*(3), 443-453.

Norris, M., & Lecavalier, L. (2010). Screening accuracy of level 2 autism spectrum disorder rating scales: A review of selected instruments. *Autism*, *14*(4), 263-284.

Ohan, J. L., Leung, D. W., & Johnston, C. (2000). The Parenting Sense of Competence scale: Evidence of a stable factor structure and validity. *Canadian Journal of Behavioral Science*, *32*(4), 251-261.

Øien, R. A., Schjølberg, S., Volkmar, F. R., Shic, F., Cicchetti, D. V., Nordahl-Hansen, A., ... & Ventola, P. (2018). Clinical features of children with autism who passed 18-month screening. *Pediatrics*, *141*(6), 1-10.

Ozonoff, S., Li, D., Deprey, L., Hanzel, E. P., & Iosif, A.-M. (2018). Reliability of parent recall of symptom onset and timing in autism spectrum disorder. *Autism*, *22*(7), 891–896.

Petrillo, J., Cano, S. J., McLeod, L. D., & Coon, C. D. (2015). Using classical test theory, item response theory, and Rasch measurement theory to evaluate patient-reported outcome measures: A comparison of worked examples. *Value in Health*, *18*(1), 25-34.

Rasch, G. (1960). Studies in mathematical psychology: I. Probabilistic models for some intelligence and attainment tests. Oxford, England: Nielsen & Lydiche.

Reis, J. (1988). Child-rearing expectations and developmental knowledge according to maternal age and parity. *Infant Mental Health Journal*, *9*(4), 287-304.

Reiner-Hess, C. R., Teti, D. M., & Hussey-Gardner, B. (2004). Self-efficacy and parenting of high-risk infants: The moderating role of parent knowledge of infant

development. *Journal of Applied Developmental Psychology*, *25*(4), 423-437.

Reise, S. P., & Rodriguez, A. (2016). Item response theory and the measurement of psychiatric constructs: some empirical and conceptual issues and challenges. *Psychological Medicine*, *46*(10), 2025-2039.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27-48.

Reisinger, L. M., Cornish, K. M., & Fombonne, É. (2011). Diagnostic differentiation of autism spectrum disorders and pragmatic language impairment. *Journal of Autism and Developmental Disorders*, *41*(12), 1694-1704.

Roane, H. S., Fisher, W. W., & Carr, J. E. (2016). Applied behavior analysis as treatment for autism spectrum disorder. *The Journal of Pediatrics*, *175*, 27-32.

Robins, D. L., Casagrande, K., Barton, M., Chen, C. M. A., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for autism in toddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics, 133*(1), 37-45.

Robbins, L. C. (1963). The accuracy of parental recall of aspects of child development and of child rearing practices. *The Journal of Abnormal and Social Psychology*, *66*(3), 261-270.

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more, version 0.5–12 (BETA). *Journal of Statistical Software*, *48*(2), 1-36.

Roth, B. M., Kralovic, S., Roizen, N. J., Spannagel, S. C., Minich, N., & Knapp, J. (2016). Impact of autism navigator on access to services. *Journal of Developmental & Behavioral Pediatrics*, *37*(3), 188-195.

Rutter, M., Le Couteur, A., & Lord, C. (2003). Autism diagnostic interview-revised. *Los*

Angeles, CA: Western Psychological Services.

Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., ... & McNerney, E. (2015). Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *45*(8), 2411-2428.

Serketich, W. J., & Dumas, J. E. (1996). The effectiveness of behavioral parent training to modify antisocial behavior in children: A meta-analysis. *Behavior Therapy*, *27*(2), 171-186.

Seung, H., Ji, J., Kim, S., Sung, I., Youn, Y., Hong, G., & ... Youm, H. K. (2015). Examination of the Korean Modified Checklist of Autism in Toddlers: Item response theory. *Journal of Autism and Developmental Disorders*, *45*(9), 2744-2757.

Siu, A. L., Bibbins-Domingo, K., Grossman, D. C., Baumann, L. C., Davidson, K. W., Ebell, M., ... & Krist, A. H. (2016). Screening for autism spectrum disorder in young children: US Preventive Services Task Force recommendation statement. *Jama*, *315*(7), 691-696.

Sturner, R., Howard, B., Bergmann, P., Morrel, T., Andon, L., Marks, D., & ... Landa, R. (2016). Autism screening with online decision support by primary care pediatricians aided by M-CHAT/F. *Pediatrics*, *138*(3), 1-11.

Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment, 18*(3), 291-307.

Toh, T., Tan, V. W., Lau, P. S., & Kiyu, A. (2018). Accuracy of Modified Checklist for Autism in Toddlers (M-CHAT) in detecting autism and other developmental disorders in community clinics. *Journal of Autism and Developmental Disorders*, *48*(1), 28-35.

U.S. Census Bureau (2017). *2015 National Content Test Results*. Retrieved from:

https://www2.census.gov/programs-surveys/decennial/2020/program-management/final-analysis-reports/2015nct-race-ethnicity-analysis.pdf.

Van der Valk, J. C., Van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (2001). Using parental ratings to study the etiology of 3-year-old twins' problem behaviors: different views or rater bias?. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *42*(7), 921-931.

Wallace, K. S., & Rogers, S. J. (2010). Intervening in infancy: implications for autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *51*(12), 1300-1320.

Wetherby, A. M., Brosnan-Maddox, S., Peace, V., & Newton, L. (2008). Validation of the Infant—Toddler Checklist as a broadband screener for autism spectrum disorders from 9 to 24 months of age. *Autism*, *12*(5), 487-511.

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*(1), 23-34.

Table 1. Bi-variate correlations between variables.

	2	3	4	5	6	7	8	9	10
1. KCDI – E	***.846	***.605	*-.133	***.487	.058	***.238	**-.151	**-.176	.079
2. KCDI – C	—	***.575	*-.116	***.459	.031	**.176	*-.123	*-.131	.077
3. ASK – K		—	.047	***.286	.096	.054	-.030	-.092	.062
4. ASK – C			—	**-.176	*.104	-.086	**.173	***.163	.042
5. PSOC- E				—	.096	**.158	**-.183	*-.137	***.204
6. Age					—	**-.157	***.205	.025	***.225
7. Sex						—	***-.300	**-.149	.049
8. Education							—	***.307	-.059
9. Income								—	.041
10. Children									—

Note. KCDI – E = Knowledge of Emotional Development. KCDI – C = Knowledge of Cognitive Development. ASK – K = Knowledge of ASD. ASK – C = Confidence in Knowledge of ASD. PSOC – E = Parental Self-Efficacy. Sex was coded as 0 = Male, 1 = Female. Education was coded as a categorical variable with increasing number corresponding to increased level of educational attainment. Children = the number of children the parent has raised who is older than 36 months old.

*** $p < .001$. ** $p < .01$. * $p < .05$.

Table 2. Sample Demographics

	Mean	***SD***
Parent Age	34	7.99
Household Income	$65,000**	--
Number of Children	2.347	1.391
Sex	***N***	**%**
Male	137	48
Female	151	52
Race	***N***	**%**
Caucasian/White	167	58
African-American/Black	58	20
Asian/Pacific Islander	32	11
Hispanic/Latino	15	5
Multiracial	6	2
Native American	3	1
Education	***N***	**%**
High School/GED	40	14
Some College/Associates	76	26
Technical School Degree	13	4.5
Bachelor's Degree	83	29
Master's Degree	64	22
Professional Degree	11	4
Prior Exposure to ASD	***N***	**%**
Have a Child w/ ASD	34	12
Have a Friend w/ASD	28	10
Have a Sibling w/ ASD	15	7
No Exposure	87	30

** Median is reported for Income due to violation of univariate normality.

Table 3. Results of IRT Rasch Model.

M-CHAT Item	Mild			Severe			Difference
	Percent	Difficulty	Probability	Percent	Difficulty	Probability	
Item 1	42%	0.324	.380	74%	-0.931	.803	9.543***
Item 2	71%	-0.806	.772	--			
Item 3	45%	0.179	.433	73%	-0.925	.802	8.436***
Item 4	70%	-.750	.757	--			
Item 5	71%	-0.818	.775	--			
Item 6	58%	-0.275	.602	64%	-0.531	.691	1.811
Item 7	54%	-0.124	.547	66%	-0.599	.712	3.447**
Item 8	--			--			
Item 9	42%	0.323	.380	67%	-0.636	.723	6.991***
Item 10	--			79%	-1.188	.858	
Item 11	--			77%	-1.089	.838	
Item 12	--			--			
Item 13	93%	-2.306	.970	93%	-2.268	.969	0.185
Item 14	48%	0.072	.473	78%	-1.161	.852	8.519***
Item 15	53%	-0.100	.538	72%	-0.867	.788	5.258***
Item 16	55%	-0.169	.564	76%	-1.048	.830	6.186***
Item 17	52%	-0.070	.526	76%	-1.057	.831	7.412***
Item 18	47%	0.125	.453	76%	-1.050	.830	8.588***
Item 19	48%	0.102	.461	72%	-0.857	.785	7.041***
Item 20	--			--			

Note. Percent = Percentage of sample that correctly rated the item. Difficulty = difficulty parameter (α) ranges from -2 to 2, with higher scores indicating greater difficulty. Probability = probability of a correct response from an individual with median ability level. Difference = paired samples *t*-test. * = $p < .05$, ** = $p < .01$, *** = $p < .001$.

Table 4. Factor Loading Table

M-CHAT-R (Mild Video)	Mild Factor Loading (λ)	Mild Factor Error (θ)		M-CHAT-R (Severe Video)	Severe Factor Loading (λ)	Severe Factor Error (θ)
Item 1	.830	.312		Item 1	.858	.264
Item 2	. 323	.895		Item 2	--	
Item 3	. 609	.629		Item 3	.796	.366
Item 4	.766	.413		Item 4	--	
Item 5	.640	.591		Item 5	--	
Item 6	.783	.387		Item 6	.599	.641
Item 7	.841	.293		Item 7	.744	.446
Item 8	--			Item 8	--	
Item 9	.763	.418		Item 9	.705	.503
Item 10	--			Item 10	.923	.149
Item 11	--			Item 11	.857	.266
Item 12	--			Item 12	--	
Item 13	**			Item 13	**	
Item 14	.720	.482		Item 14	.872	.298
Item 15	.540	.708		Item 15	.859	.263
Item 16	.733	.463		Item 16	.887	.212
Item 17	.827	.316		Item 17	.841	.293
Item 18	.734	.461		Item 18	.767	.412
Item 19	.686	.529		Item 19	.825	.320
Item 20	--			Item 20	--	

Note. All parameters reported are completely standardized solution parameters (i.e., "std.all" in Lavaan). ** = indicates that loading was removed.

Table 5. Structural equation modeling matrices

Psi (variances and covariances of endogenous latent variables)

	Mild	Severe
Mild	.635	
Severe	.309	.095

Phi (variances and covariances of exogenous latent variables)

	Knowledge	Characteristics
Parent Knowledge	1.000	
Parent Characteristics	0.549	1.000

Beta (path coefficients)

	Mild	Severe
Parent Knowledge	.076	.306
Parent Characteristics	.549	.686
ASK-ASD Confidence	-.107	-.330

Lambda-X (factor loadings onto exogenous latent variables)

	Knowledge	Characteristics
KCDI Emotional	.865	0
KCDI Cognitive	.829	0
ASK-ASD Knowledge	.780	0
PSOC	0	.528
Number of Children	0	.201
Parental Education	0	-.406

Note. Parameters reported are from completely standardized solution (std.all in Lavaan). Knowledge = Parent Knowledge factor. Characteristics = Parent Characteristics factor.

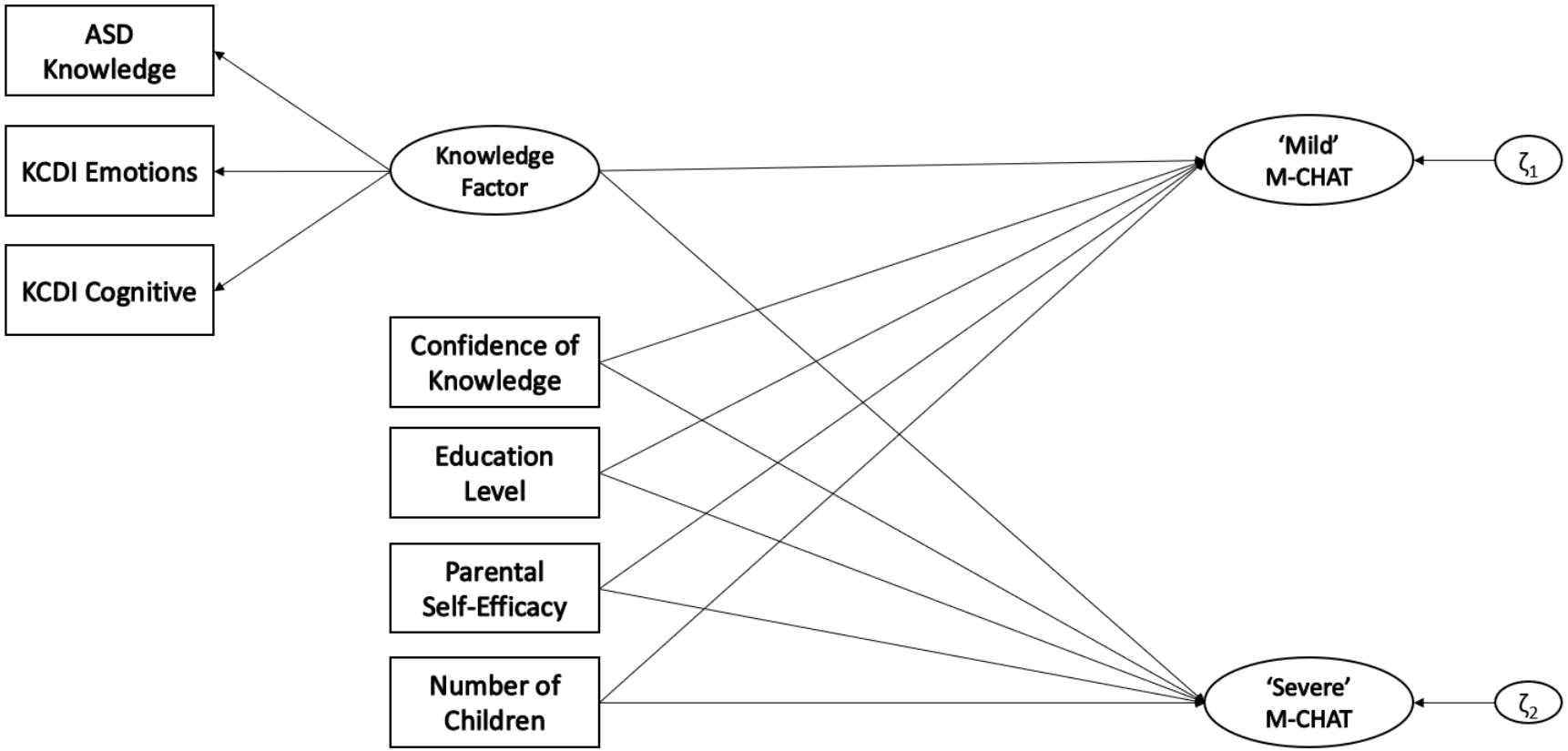

Figure 1. Full Proposed Model.

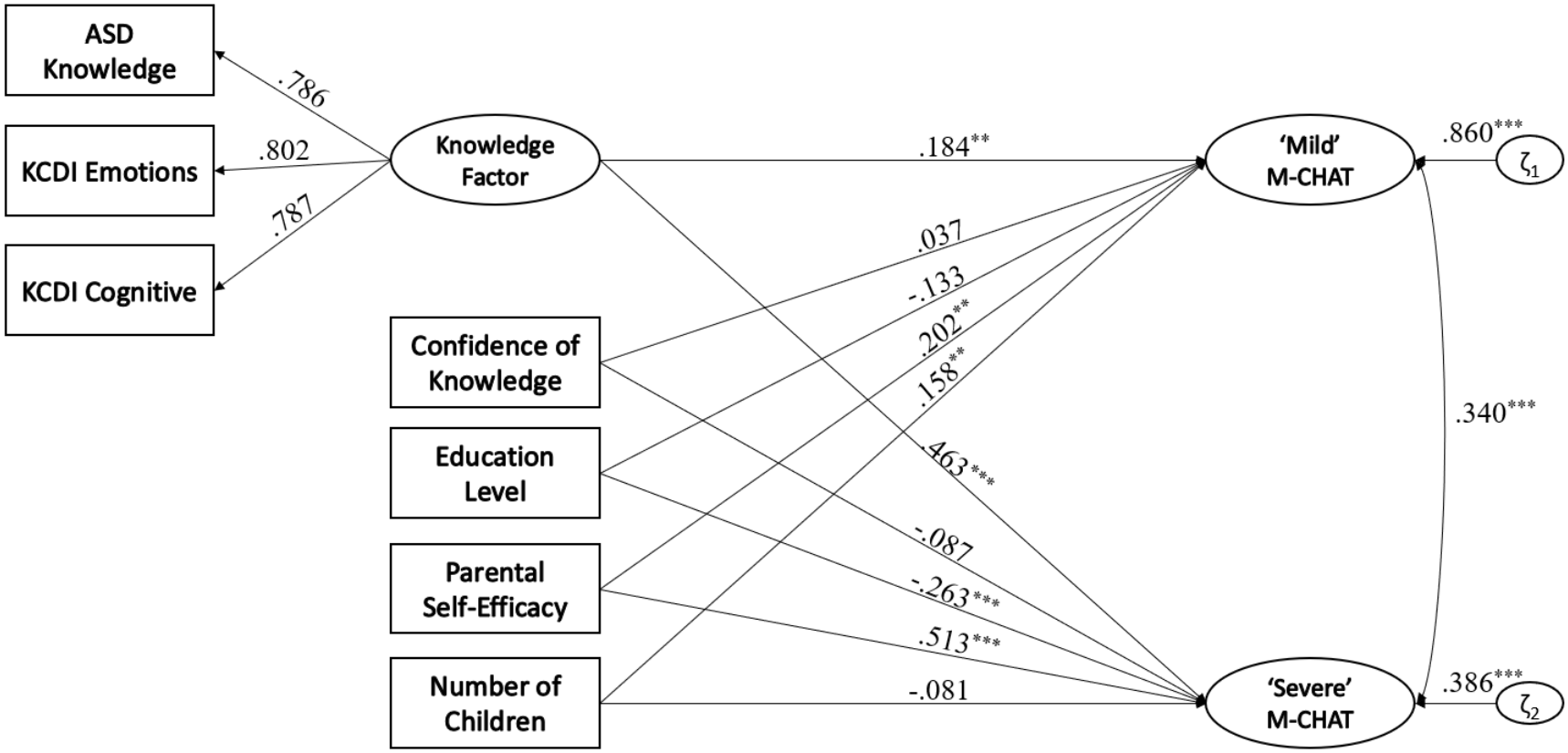

Figure 2. Final model with respective beta, lambda-x, phi, and psi reported. Lambda-y (loadings for endogenous variables) not illustrated for clarity; factor loadings can be found in Table 4. Model fit: χ^2 (514) = 786.006, $p < .001$; Robust CFI = .881; Robust RMSEA = .046.

* $p < .05$, ** $p < .01$, *** $p < .001$.

Appendix A

IRB Approval

THE UNIVERSITY OF ALABAMA | Office of the Vice President for Research & Economic Development
Office for Research Compliance

August 31, 2018

James Rankin
Department of Psychology
College of Arts & Sciences
Box 870348

Re: IRB Application #: 18-005 "Parental Knowledge in Screening for Autism: Development of an Intervention to Enhance Screening Accuracy"

Dear James Rankin:

The University of Alabama IRB has received the revisions requested by the full board on 6/21/18. The board has reviewed the revisions and your protocol is now approved for a one-year period. Please be advised that your protocol will expire one year from the date of approval, 6/21/18.

If your research will continue beyond this date, complete the IRB Continuing Review/Renewal Application within the e-Protocol system by the 15th of the month prior to project expiration. If you need to modify the study, please submit the Modification form within the e-Protocol system. Changes in this study cannot be initiated without IRB approval, except when necessary to eliminate apparent immediate hazards to participants. When the study closes, please complete the Final Report Form.

Please use reproductions of the IRB approved stamped consent forms to provide to your participants.

Should you need to submit any further correspondence regarding this proposal, please include the assigned IRB application number.

Good luck with your research.

Sincerely,

Joan M Barth

Joan Barth, PhD
Chair, Non-Medical Institutional Review Board

358 Rose Administration Building | Box 870127 | Tuscaloosa, AL 35487-0127
205-348-8461 | Fax 205-348-7189 | Toll Free 1-877-820-3066

Appendix A

Demographic Questionnaire

Age: _____

Gender:

— Male
— Female
— Transgender
— Other
— Prefer not to respond

Race:

— African American/Black
— Asian/Pacific Islander
— Hispanic/Latino
— Native American/American Indian
— White
— Multiracial
— Not Listed
— Prefer not to respond

Ethnicity

— Hispanic/Latino
— Not Hispanic/Latino

Household Annual Income: _____

Marital Status:

— Married
— Divorced/Separated, Currently Single
— Divorced/Separated, Living with Partner
— Never Married, Currently Single
— Never Married, Living with Partner
— Widowed, Currently Single
— Widowed, Living with Partner

Number of Children:

— 1
— 2
— 3
— 4
— 5
— 6
— 7

— 8
— 9
— 10+

Ages of Children:

Child 1: _____
Child 2: _____
Child 3: _____
Child 4: _____
Child 5: _____
Child 6: _____
Child 7: _____
Child 8: _____
Child 9: _____
Child 10: _____

Education Level:

— Some High School
— High School Graduate/GED
— Some College/Associates Degree
— Technical School Degree
— Bachelor's Degree
— Master's Degree (e.g., M.A., MPH, MBA)
— Professional Degree (e.g., JD, MD, PhD, PsyD)

We would like to know what, if any, exposure you have had to children with autism. Please choose any/all that apply to your situation

— I have a child with autism
— I have a niece, nephew, or cousin with autism
— I have a sibling with autism
— I have a friend with autism
— A friend has a child/family member with autism
— I work with children/adults with autism
— A friend or family member works with children/adults with autism
— Other relations to a child or adult with autism;
Please describe: _____

If you know someone who has children with autism, for how many years have you known this parent(s) of a child with autism? _____

What is the age range of their child(ren) with autism? _____

If you work with children or adults with autism, for how many years have you worked with children or adults with autism? _____

What is the age range of the children/adults with autism that you work with? _____