ASSESSING THEORY OF MIND IN SCHOOL-AGED PERSONS WITH AUTISM AND TYPICAL PEERS

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

BY

KENNETH WEHRHEIM

DR. MARIA HERNÁNDEZ-FINCH – ADVISOR

BALL STATE UNIVERSITY

MUNICE, INDIANA

JULY 2017

ACKNOWLEDGEMENTS

I would like to thank my committee members for their patience and guidance through this entire process. Your guidance has been invaluable and made possible for me what I could never have achieved on my own. I would like to thank my fellow graduate students and my work colleagues for their assistance in locating and testing the participants for my study. With no local connections to speak of, collecting the necessary data would have been impossible without you. I would like to thank my Fraser colleagues for their continued support as I have slowly completed this process, as well as for their role as my local research and statistics support team for those problems which required a live person to sit next to me and point out my mistakes in real-time. Finally, I would like to thank my wife for reasons that are too numerous to list here. While many other individuals have helped throughout this process, no one has sacrificed as much or dedicated as much energy to getting me over the finish line as you. You were a voice of encouragement in times of desperation, a source of reason in times of confusion, and a well-placed kick in times of laziness and procrastination.

TABLE OF CONTENTS

CHAPTER I

Introduction

In this chapter, autism spectrum disorder (ASD) and its predecessors are briefly discussed. Diagnostic criteria and developmental trajectory are also discussed, as well as Theory of Mind, a feature which has been found to be underdeveloped in individuals with ASD and hypothesized to contribute to some of the hallmark features of the disorder, specifically those related to social competence. The purpose of the present study is discussed, including research questions. The significance and limitations of the current study are also discussed.

Autism Spectrum Disorders

ASD and its predecessors, Autistic Disorder and Infantile Autism, are developmental disorders which are characterized by deficits in communication (verbal and non-verbal), social interaction, and repetitive and restricted interests, behaviors, and activities (American Psychiatric Association [APA], 1980; APA, 2000; APA, 2013). The most recent version of the Diagnostic and Statistical Manual (henceforth referred to as the DSM; APA, 2013) defines two core symptoms: deficits in social communications and interactions, including, but not limited to, non-verbal communication, metaphorical language, emotional regulation, understanding other's mental states (known as Theory of Mind, discussed in more detail in the following sections), and social competence; and restricted, repetitive behaviors and interests, such as repetitive motor movements, or a preoccupation with a particular topic beyond what is appropriate to the situation. Having long been thought of as a spectrum disorder with various levels of functioning, the DSM-5 is the first to recognize the disorder as such, with the previous version defining separate disorders under the overall category of "Pervasive Developmental Disorders."

The deficits associated with ASD are present throughout the lifespan. With early intervention, diagnosed individuals may learn coping strategies to compensate for the associated symptoms, but are unlikely to "grow out of it" (Kanner, 1971; Pandey, Wilson, Verbalis, & Fein, 2008; Rumsey, Rapoport, & Sceery, 1985). Difficulty with interpersonal relationships is one feature that has been shown to persist throughout the lifespan, making it difficult to establish or maintain relationships with others (Rutter & Lockyer, 1967; Mesibov, Schopler, Schaffer, & Michal, 1989; Piven, Harper, Palmer, & Arndt, 1996; Gilchrist, Green, Cox, Burton, Rutter, & Le Couteur, 2001; Seltzer, Krauss, Shattuck, Orsmond, Swe, & Lord, 2003; Howlin, Goode, Hutton, & Rutter, 2004). Early intervention to teach these skills may help these individuals learn coping strategies, and provide them with sufficient training to overcome the deficits and establish meaningful relationships with others. In order to develop and maintain meaningful interpersonal relationships, an individual must be able to understand the thoughts and feelings of others, which enables them to recognize and respond to nonverbal communication. This ability is known as "Theory of Mind."

Theory of Mind

Theory of Mind is the ability to understand a person's own thoughts, feelings, and emotions, as well as the ability to understand that other individuals also have their own thoughts, feelings, and emotions, which may differ from the individual's own (Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001). This ability to understand the differing minds of others is an essential component of social interaction, as well as social decision making. In order to successfully navigate a social interaction, a person must be able to accurately assess the intentions and expectations of those they are interacting with. Lacking such an ability would make it extremely difficult, if not impossible, for the individual to interact in a meaningful way.

This has been confirmed by multiple studies which show that Theory of Mind deficits in individuals with ASD are linked to deficits in social competency, communication, and interaction (Cutting & Dunn, 2006; Peterson, Garnett, Kelly, & Attwood, 2009; Watson, Nixon, & Capage, 1999).

Several methods of assessing an individual's Theory of Mind exist. The most common method is known as a false belief task. This task requires an individual to evaluate the incorrect belief of a character in a story. While the participant knows the truth of the situation, the character in the story is mistaken. The participant must infer what the character in the story will do based on their incorrect belief. This is known as a first order false belief. A more advanced task requires the participant to evaluate what a second character will do based on their belief of the first character's false belief, which is known as a second order false belief task. These tasks are generally able to be passed by typically developing four and six year old children, respectively (Cohen, Baron-Cohen, & Tager-Flusberg, 1994). While age-matched children with ASD are not able to pass these tasks, adolescents and adults with ASD are able to pass these tasks. While some have taken this to mean that, as these individuals age, they acquire Theory of Mind skills, others have disputed these accounts, pointing out that this means they have simply reached a point where they are capable of tasks that an average four or six year old can complete. Instead, it is simply likely that, as these individuals age, they are able to use cognitive strategies that allow them to complete these tasks, explaining why social difficulties persist, despite improved performance on these tasks (Scheeren, De Rosnay, Koot, & Begeer, 2013).

To more accurately measure Theory of Mind, several advanced tests of Theory of Mind have been developed which more closely approximate the demands of a real world social interaction. When presented with these more advanced measures, individuals with high

functioning autism showed impairment in their theory of mind, both compared to typical control subjects, and to control non-Theory of Mind tasks (Beaumont & Newcombe, 2006; Brent, Rios, Happé, & Charman, 2004; Happé, 1994; Jolliffe & Baron-Cohen, 1999; Kaland, Møller-Nielson, Smith, Mortensen, Callesen, & Gottlieb, 2005; Velloso, Duarte, & Schwartzman, 2013).

One advanced measure used to evaluate Theory of Mind skills can be found in the NEPSY-II. The NEPSY-II is a neuropsychological measure that includes a test of Theory of Mind. This subtest has been nationally normed, as well as tested with an ASD participant sample. The NEPSY-II Theory of Mind subtest measures participants' ability to accurately interpret the point of view of characters in pictures and stories, as well as correctly identify the appropriate emotion in a pictured situation based on contextual clues, both of which require Theory of Mind. Results have indicated that the NEPSY-II Theory of Mind subtest is a valid measure of these skills, and is able to detect deficits in this area with individuals with ASD (Korkman, Kirk, & Kemp, 2007; Narzisi, Muratori, Calderoni, Fabbro, & Urgesi, 2013).

Roberts-2

The Roberts-2 is a psychometric instrument used to evaluate an individual's social perception both adaptively and clinically. This allows the examiner to evaluate the participant's prosocial attitudes and behaviors about the world around them, such as their ability to identify support resources and general positive feelings about their natural environment, and social/emotional disturbances, such as an inability to identify a problematic situation, or an excessive degree of anxious content. In particular, the Problem Identification and Resolution scales require Theory of Mind skills to achieve a positive score. The Problem Identification scale requires participants to identify the feelings of pictured characters, and a plausible cause for those feelings. The Resolution scale requires the individual to adequately identify the emotions

of the pictured character in order to generate a prosocial solution which sufficiently addresses the feelings of the pictured characters, and resolves conflicts in a positive way (Roberts & Gruber, 2005). Both of these scales require the individual to be capable of correctly interpreting facial expressions and body language, taking the perspective of the characters in the scene, and understanding the emotions of those characters in order to generate a solution which would adequately address those emotions.

Conclusion

Individuals with ASD exhibit impairment in social communication and interaction throughout the lifespan. Multiple studies have concluded this to be due to a deficient Theory of Mind, which is the ability to understand that other people have different thoughts and feelings. The ability to understand that others do not feel the same way as the individual is a crucial component of social interaction. While some adolescents and adults with ASD are able to pass basic Theory of Mind tasks, the social deficits associated with ASD persist. More advanced Theory of Mind measures have been able to show that the deficits are still present, while the individuals have learned alternative problem solving strategies which allowed them to complete basic Theory of Mind tasks. Among these measures is the NEPSY-II, a neuropsychological measure containing a Theory of Mind subtest which has been shown to correctly identify Theory of Mind deficits in individuals with ASD. The Roberts-2 is a psychometric instrument used to measure social competency through tasks which require Theory of Mind skills. Despite the aspects of the Roberts-2 which require Theory of Mind abilities, it has not been validated as a measure for that purpose. Therefore, the current study will examine whether performance on the Problem Identification and Resolution scales of the Roberts-2 among children with ASD correlates with performance on the NEPSY-II Theory of Mind subtest.

Purpose of the Present Study

Assessment of Theory of Mind is generally limited to criterion-referenced measures, limiting the generalizability of the conclusions about an individual's performance in this domain, relative to what might be expected of a typically developing peer. In research settings, the purpose of assessing this skill is not to evaluate the overall functioning of the participant; therefore, criterion-based measurements are appropriate and informative. In clinical settings, however, there is a need to evaluate the performance of an individual within the context of what might be expected of an individual the same age in order to inform treatment goals and planning. To this end, the only existing norm-referenced measure of Theory of Mind is the NEPSY-II Theory of Mind subtest. While this measure has been tested with individuals diagnosed with ASD, the skill is tested in isolation, and does not provide additional information about the specific strengths and weaknesses the child displays in their social competency. The purpose of this study, therefore, is to determine if the Roberts-2 can be used as a valid test of Theory of Mind.

To answer this question, the Roberts-2 will be administered to children and adolescents with ASD and compared to results on the NEPSY-II Theory of Mind Subtest. The results will be analyzed to determine whether performance on the Problem Identification and Resolution scales correlates with performance on the NEPSY-II Theory of Mind subtest. The questions sought to be answered by this study are as follows:

1. Does performance on the Problem Identification and Resolution scales of the Roberts-2 correlate to Theory of Mind deficits in children with ASD on the NEPSY-II Theory of Mind subtest after controlling for cognitive abilities?

2. Does the performance of individuals with ASD on the Roberts-2 differ in a consistent pattern from typically developing peers, controlling for estimated cognitive abilities?

Significance of the Study

There are few studies that have used a nationally normed psychometric instrument in the intended fashion to evaluate Theory of Mind abilities in individuals with autism. Most studies of Theory of Mind in this population rely on criterion-referenced measures, such as the Strange Stories task, or Awkward Moments task (Brent et al., 2004; Happé, 1994; Jolliffe & Baron-Cohen, 1999; Kaland et al., 2005; Velloso et al., 2013). Others have used norm-referenced measures in a way that is not consistent with administration guidelines (Beaumont & Newcombe, 2006). Only one study could be located (Narzisi et al., 2013) that utilized a norm-referenced measure in the way intended by the test publishers to evaluate Theory of Mind abilities in individuals with ASD. Results showed that children with ASD showed a significantly different neuropsychological profile than typical peers. The authors found that Attention and Executive Functions (planning and other higher cognitive skills), Language, Learning and Memory, and Sensorimotor Processing were all impaired. Theory of Mind, which is measured in two portions, a verbal portion and a contextual portion, was found to be partially impaired. The contextual portion, which requires examinees to determine the appropriate emotion from contextual clues, was found to be average, suggesting that, with adequate contextual clues, children with ASD are capable of accurately evaluating emotions. The verbal portion, which requires examinees to evaluate a character's point of view, was found to be impaired.

The use of criterion-referenced assessments can be helpful in some situations. Criterion-referenced assessments are primarily used to evaluate, as the name implies, specific criteria. This is helpful when a particular skill or ability is being evaluated without regard to what might

be expected from a comparison individual. This can be useful in determining the presence of an ability or skill. For instance, if the goal is to simply determine whether or not an examinee is capable of performing a Theory of Mind task, a criterion-referenced assessment will provide the necessary information. Norm-referenced assessments are useful when attempting to make comparisons. Once it has been determined whether or not the child is capable of the task in the previous example, an examiner may wish to determine to extent to which the child is proficient in their use of the Theory of Mind skill relative to a typical peer. A criterion-referenced assessment can tell us whether or not the child possesses the skill we are interested in, while a norm-referenced assessment can provide additional information concerning what should be expected of child, relative to their chronological or mental age, and allow conclusions to be drawn based on expected developmental trajectory. Using a criterion-referenced assessment provides useful information about specific skills that the child may possess, but does not provide information about developmentally appropriate levels of performance. This makes it challenging to set appropriate goals, as well as to measure whether any set goals have been met in that area. Using a norm-referenced assessment in a fashion that it was not intended to be used renders the norming sample irrelevant, creating the same limitations. In addition, by using the measure in a non-standard fashion, the information that is intended to be collected by the instrument is not collected, further limiting the information yielded by the measure (Freeman & Miller, 2001).

The Roberts-2 is a measure that is used to evaluate an individual's social competency by evaluating the individual's ability to visually process the context clues of a social situation, and determine the emotions of those pictured, a plausible cause for those emotions, what the pictured individuals may plausibly do as a result of those emotions, and a plausible outcome that will adequately resolve the emotions of the pictured individuals. By evaluating the individual's

ability to do these things, reasonable conclusions may be drawn about whether the individual would be able to adequately detect the same context clues in a real world situation. An individual requires an intact Theory of Mind in order to make use of these skills in a real world setting, as being able to take the perspective of another individual and understand their thoughts and emotions is a necessary skill when attempting to communicate with another individual in a socially appropriate manner.

Because the Roberts-2 incorporates skills that require an intact Theory of Mind, it could be used to evaluate this skill, while also allowing the evaluator to draw conclusions about other areas of the child's social competence. In addition to providing a measurement of the child's Theory of Mind ability, the evaluation using the Roberts-2 will also provide the information that the measure was originally designed to measure, such as ability to identify support resources, and ability to generate realistic, positive, prosocial solutions to problem situations. The open-ended nature of responses to Roberts-2 stimuli also creates more opportunities for qualitative analysis of responses for particular themes and patterns of responding, which can provide more insight into why a child might be failing to accurately evaluate the pictured situations.

CHAPTER II

REVIEW OF LITERATURE

This chapter provides a review of the literature which is relevant to the current study. The first section discusses the relevant features of autism, assessment of autism, and provides a brief history of the condition. The first section will also contrast the development of individuals with autism to that of typically developing peers. The second section reviews Theory of Mind, and how it is affected in individuals with autism. The final section reviews measures which have been used to assess Theory of Mind, and why the Roberts-2 can provide a unique advantage over existing measures.

Autism Spectrum Disorders

History. Autism Spectrum Disorders (ASDs) are a class of psychiatric disorders characterized by deficits in social interactions, communication, and restricted interests. Autism was first described by Leo Kanner in his 1943 article, Autistic Disturbances of Affective Contact. Kanner described the children in his study as having an "innate inability to form the usual, biologically provided affective contact with people, just as other children come into the world with innate physical or intellectual handicaps," (p. 250). Kanner described in detail the style of interaction the children had with the people around them, first preferring to be left alone, but interacting more frequently as the children came to understand the value that other people hold in their ability to help the child access desired items, answer their questions, and teach them to do things. He noted that they eventually learn to comply with requests and instructions, in the interest of ending the interference of the other party sooner, and allowing the child to return to their preferred activities. Kanner believed that the children were of average intelligence, and that

the observed deficits were a result of a lack of motivation to interact with others on the part of the children (Kanner, 1943).

Despite being described by Kanner in 1943, the first two editions of the Diagnostic and Statistical Manual in 1952 and 1968 did not include autism as a separate diagnosis. Children displaying the symptoms of the diagnosis were given a label of childhood schizophrenia (American Psychiatric Association [APA], 1952, 1968; Goldstein & Ozonoff, 2009). With the third edition, the disorder was differentiated from schizophrenia, and was classified as a single diagnosis, infantile autism (APA, 1980). The 1987 revision, the DSM-III-R, removed the "infantile" descriptor, making the name "autistic disorder" (APA, 1987). The DSM-IV and DSM-IV-TR expanded the disorder to a class of disorders known as Pervasive Developmental Disorders, which included other diagnoses such as Rett's Disorder, Asperger Syndrome, and Childhood Disintegrative Disorder (APA, 1994, 2000). The currently accepted view of autism in the DSM-5 is that of a single spectrum that encompasses the previously distinct diagnoses, known as autism spectrum disorder (APA, 2013). Research on these disorders has suggested that high-functioning autistic disorder and Asperger Syndrome were phenotypically similar enough to justify combining the diagnoses into a single spectrum diagnosis.

Asperger Syndrome is a disorder from the DSM-IV and DSM-IV-TR characterized by the same social impairments and restricted interests found in children with Autism Spectrum Disorders; however, these individuals develop language at the same time and pace as their non-diagnosed peers. This has led to some debate regarding whether Asperger's Syndrome is a separate condition, or the highest end of the autism spectrum (Holaday, Moak, & Shipley, 2001; Howlin, 2003; Kurita, 1997; Macintosh & Disanayake, 2006; Schopler, 1996). While Asperger's Syndrome was incorporated into Autism Spectrum Disorder in the DSM-5 and is no

longer considered a separate diagnosis, it is still included in the International Classification of Diseases, 10th Edition, the diagnostic manual of the World Health Organization (World Health Organization [WHO], 1992), which is more commonly used by physicians. Though the upcoming 11th edition of the ICD proposes a move to a spectrum disorder which encompasses Asperger Syndrome as well. Much of the research on ASD has focused on higher functioning forms of ASD, including Asperger Syndrome.

Due to the relatively recent adaption of the DSM-5, there is comparatively little research that has been conducted using the most recent criteria for the disorder, compared to the DSM-IV-TR. The previous category of Pervasive Developmental Disorders used a three-factor model to diagnose these disorders. As greater understanding of these disorders was obtained, this was refined to a two-factor model, which was found to be a more accurate and valid system of diagnosis (Huerta, Bishop, Duncan, Hus, & Lord, 2012). Some concerns existed that the added criteria would lead to individuals who had previously been diagnosed no longer meeting the criteria for the diagnosis. This concern was disproven upon further analysis of the information collected. It was found that most children who qualified for a DSM-IV-TR diagnosis of autistic disorder would still qualify for the DSM-5 diagnosis of autism spectrum disorder (Huerta et al., 2012; Mayes et al., 2012). For this reason, research conducted under the older definition can still be considered valid and applicable, despite the revised diagnostic criteria. As more research is done, a greater understanding of the causes and expressions of the disorder is achieved, leading to more accurate diagnoses, and more effective treatments.

Diagnostic Criteria. To be diagnosed with an Autism Spectrum Disorder, an individual must have deficits in social communication that occur in all situations. These deficits are characterized by difficulty in social-emotional reciprocity, which is the ability to understand

another person's feelings and interests in order to have appropriate conversations. There are also difficulties understanding non-verbal communication, such as body language. Eye contact is also an area that is negatively impacted, in that individuals with an ASD often do not maintain appropriate eye contact when interacting. These deficits result can result in difficulty maintaining relationships with others, due to the inherent struggle with understanding the relationship, or establishing the relationship in the first place. These deficits can be severe, requiring step-by-step instruction to overcome, or can be less severe, only requiring some general guidance to overcome. The other major features of ASDs are restricted, repetitive behaviors, interests, and activities. This can take the form of repetitive movements, such as hand-flapping, overly rigid adherence to a schedule or routine, abnormally intense interest in a particular subject or object, or being under- or over-reactive to sensory stimulation, such as an oversensitivity to sounds that are considered to be at a normal volume, an abnormal aversion to a particular taste or texture, or, conversely, an abnormally high threshold for hot, cold, pain, or loud noise. If these qualities interfere with an individual's daily functioning, have been present, at least partially, since early development, and cannot be better explained by another disability, the individual meets the diagnostic criteria for Autism Spectrum Disorder. Intellectual disabilities also frequently accompany ASDs. While the overall categories of deficits in social communication and interaction and restricted and repetitive behaviors, interests, and activities must be present, every specific feature of these categories is not required to meet diagnostic criteria for ASD. ASD, as a result, can be expressed in many different ways. Individuals with ASD may struggle with non-verbal communication but not social-emotional reciprocity, and they may struggle with stereotyped movements, but have no difficulty with their sensitivity to sensation (APA, 2013).

Many of these features can also manifest in other disorders, which must be ruled out in order to make a diagnosis of ASD. Rett syndrome is a biologically-based disorder which affects girls typically between the ages of one and four, and has many of the same features of ASD. In the prior version of the DSM, the DSM-IV-TR (APA, 2000), Rett syndrome was included in the Pervasive Developmental Disorder category due to similarities between Rett syndrome and autism. The disorder was removed from the DSM-5 (APA, 2013) due to the discovery of a biological cause for the disorder. In contrast to ASD, children with Rett syndrome, after a regressive period between the ages of one and four, social communication skills improve, and autistic features become a less prominent area of concern (APA, 2013).

The DSM-IV-TR (APA, 2000) included development of language as a diagnostic criterion for autistic disorder. When the newest edition of the DSM was released, language development was no longer a consideration for a diagnosis of ASD, but was included under a related diagnosis, social communication disorder. Individuals with social communication disorder may exhibit impairment in social communication or interactions that resemble what is seen in ASD, but lack the restricted interests and repetitive behaviors seen in an ASD diagnosis. When communication deficits are observed in a child, an ASD diagnosis takes priority over a social communication disorder diagnosis, making it essential to determine the presence of restricted interests and repetitive behaviors seen in ASD before making either diagnosis (APA, 2013).

Several other disorders also have common features with ASD. Stereotypic movement disorder is characterized by the repetitive motor movements seen in ASD, and is only given when other diagnostic criteria for ASD are not met. Childhood onset schizophrenia can also closely resemble ASD. Individuals with schizophrenia may exhibit social impairments and

atypical interests and beliefs, and can resemble those seen in ASD. A key feature of schizophrenia is the presence of hallucinations (sensory perception of stimuli which are not actually present), which must be carefully evaluated in order to differentiate the two diagnoses. Care must be taken to word questions related to hallucinations in a way that would not lead an individual with ASD to misunderstand and affirm schizophrenic criteria (i.e. endorsing auditory hallucinations when actually referring to voices coming from the TV or radio) (APA, 2013)

Several features of autism have been present in each version of the DSM since its first recognition as a disorder in the 3rd edition. Language impairment has been one such feature that has persisted in each generation. The DSM-III required significant impairment in speech production, and if the child had language, displayed impairments such as echolalia (repeating the same sound, word, or phrase multiple times), metaphorical language, or pronominal reversal (use of incorrect pronouns) (APA, 1980). The fourth edition of the DSM increased the specificity of the impairment to include a significant delay or absence of spoken language, significant difficulty in starting or maintaining a conversation, or stereotyped and repetitive language (echolalia) or idiosyncratic language (APA, 1994). With the change in the DSM-5 to a spectrum view of autism, and resulting incorporation of the Asperger Syndrome diagnosis, the most recent edition of the DSM does not require deficits in the development of spoken language, but instead focuses on the social aspect of communication, requiring difficulty with social-emotional reciprocity, difficulty with nonverbal communication, and difficulty with interpersonal relationships (APA, 2013).

Children with autism have been observed to exhibit impairment in their social development, such as the development of eye contact, and recognizing social cues. Many also struggle with recognition of emotions, a skill that tends to develop around 5 years of age in

typically developing children (Cole, Dennis, Smith-Simon, & Cohen, 2009). This deficit has been found to persist throughout development and into adulthood, making interventions for this skill highly valuable. In a study evaluating facial recognition of specific features, accuracy of emotion identification, and developmental changes in emotion recognition, females with ASD struggled to differentiate happiness and anger from pictures of eyes, which reaches adult levels of competence around 10-11 years of age in typically developing children (Tonks, Williams, Frampton, Yates, & Slater, 2007). When images of faces showing blended emotions were presented as stimuli, ASD subjects more often identified negative emotions in ambiguous stimuli, and incorrectly identified anger, as opposed to other negative emotions. Subjects also had difficulty identifying neutral emotions. Recognition of emotion through the eyes improved with age, though they still did not reach the levels of typically developing peers. (Kuusikko et al., 2009)

Other diagnostic tools, including the ICD-10, the Autism Diagnostic Observation Schedule (ADOS) and its revision (ADOS-2), and Autism Diagnostic Interview – Revised (ADI-R), all use emotional competence as criteria. One of the skills necessary to understand others' emotions, and therefore to be capable of maintaining relationships with others, is Theory of Mind (Begeer, Koot, Rieffe, Meerum Terwogt, & Steege, 2008). The ability to understand complex emotions rests on the ability to understand the other individual's thoughts. This is because emotions are not observable in objective events. Therefore, the ability to understand that other individuals have thoughts and feelings that are unique and may differ from an individual's own thoughts and feelings is necessary to understand and comprehend the emotions of others, as emotions are not a product of a situation, but are instead a product of an individual's representation of that situation..

A study by Best and colleagues (2008) examined differences between children who met the diagnostic criteria for ASD and a comparison sample of children who exhibited limited symptoms but did not meet the criteria for diagnosis on three deficits commonly seen in ASD: weak central coherence, Theory of Mind deficits, and executive dysfunction. The purpose of the study was to identify whether these deficits were necessarily linked to the diagnosis itself, or if ASD is the extreme end of a spectrum, and these commonly observed indicators also exist in the typically developing population. Results indicated that individuals who did not meet the criteria for ASD, but exhibited high levels of ASD-consistent behavioral traits, were shown to have impaired Theory of Mind, and the superior local processing associated with the differences in central coherence observed in individuals with ASD (Frith & Happé, 1994), supporting the idea that autism is one extreme on a continuum of traits that extends into the typically developing population.

Developmental Trajectory. Autism is a lifelong disorder. A person who has been correctly diagnosed with autism is unlikely to "grow out of it," but may learn to cope with the various symptoms and function in a way that leads an unfamiliar observer to believe that they have (Kanner, 1971; Rumsey, Rapoport, & Sceery, 1985). Despite the prevailing theory that it is a lifelong condition, most research has focused on autism in childhood, while comparatively little research has been conducted on adolescence, and even less has examined autism in adulthood (Seltzer, Shattuck, Abbeduto, & Greenberg, 2004).

Many children with autism do not develop language skills. Of those who do, most have significantly delayed language skills compared to age norms, and also within the individual compared to the level of language development that would be expected, based on the individual's cognitive abilities. In addition to the spoken impairments observed, individuals also

frequently display impairments in non-verbal communication skills. Several studies have shown that these impairments improve to varying degrees throughout the lifespan (Rutter & Lockyer, 1967; Howlin, Mawhood, & Rutter, 2000; Mawhood, Howlin, & Rutter, 2000; Seltzer et al., 2003; Seltzer et al., 2004).

The social deficits present in autism have also been found to persist throughout the lifespan. While much is known about the presentation of these deficits in children with autism, less is known about their presentation in adulthood (Seltzer et al., 2004). Existing research has suggested that, as individuals mature and enter adulthood, significant difficulty with interpersonal relationships persists, making it difficult to establish or maintain relationships (Rutter & Lockyer, 1967; Mesibov et al., 1989; Piven et al., 1996; Gilchrist, Green, Cox, Burton, Rutter, & Le Couteur, 2001; Seltzer et al., 2003; Howlin et al., 2004).

In the domain of restricted, repetitive behaviors and interests, improvements are shown throughout the lifespan, though some studies have found that the improvements in this domain are not as significant as those in the other domains (Rutter & Lockyer, 1967; Piven et al., 1996; Howlin et al., 2004). Other studies have focused on the presence of these symptoms in adults with autism compared to other groups. In these studies, adults were found to have fewer symptoms in these areas than adolescents with autism (Seltzer et al., 2003), but more than other adults with cognitive impairments (Bodfish, Symons, Parker, & Lewis, 2000). In a large cross-sectional study, children with autism were found to have more adaptive behaviors than age-matched cognitively impaired children. However, the opposite was the case among adults (Jacobson & Ackerman, 1990).

Research on the outcomes for individuals with autism has not indicated a high degree of favorable outcomes. Kanner's follow-up to his original data found that the majority of his

subjects were still very dependent on their caregivers, living at home or in residential settings (Kanner, 1972). Similar findings were reported by Rutter and Lockyer, who examined the outcomes of 63 individuals that had been diagnosed in the 1950s and 1960s. Of the 63, only two had jobs, and the majority lived at home or in hospital settings. 61% were reported to have made poor social adjustments (Rutter & Lockyer, 1967; Lockyer & Rutter, 1969). Similar findings were reported by Kobayashi et al. (1992) in their follow-up study of 201 individuals diagnosed with autism in Japan. 5% were enrolled in school, 1/5 were employed, and out of 201, two were living independently. Kobayashi judged that 46% had unfavorable outcomes, nearly twice as many as had favorable outcomes (27%). These findings were confirmed by Ballaban-Gil and colleagues (1996), who found that, of 45 adults with autism, only five were employed, all of whom were employed in low-level, poorly paid positions.

A longitudinal study following 161 individuals with ASD over a 10 year period found very little educational or vocational improvement in the years following high school. While five percent of participants were deemed to have experienced substantial increases in vocational activities, nearly three times as many experienced substantial decreases in the same area, suggesting that achievement deficits at the end of high school were not temporary, but persevered for several years after high school – a period when typically developing peers would excel in the same area – and sometimes worsened (Taylor & Mailick, 2014).

Children with ASD often show improvements in behavior throughout secondary school, and experience significant gains. Upon graduation, however, the rate of improvement slows dramatically. Possibly contributing to the change, adult day activities are noticeably less intellectually stimulating than school activities, possibly resulting in the decreased pace of improvement following high school. Lower income was also identified as a correlate of a slower

rate of acquisition. This difference was only present after graduating high school; while still enrolled in high school, no significant differences were found between low and high SES families. This was suggestive of the idea that, while in high school, the Individuals with Disabilities Education Act (IDEA) creates a balance, but following graduation from high school, families with limited means do not the same degree of access to services as more affluent families, contributing to a slower rate of improvement (Taylor & Seltzer, 2010).

In addition to the behavioral changes observed in children with ASD following high school, employment and higher education are also negatively affected. A second study by Taylor and Seltzer (2011) found similar results. Individuals with ASD were found to have a low rate of employment following high school. Those who were employed had mundane jobs such as changing dirty glasses, and none were employed full time. Of those individuals who had no form of daytime activity to engage in, individuals with ASD and cognitive impairment were three times more likely than individuals without cognitive impairment to have formal daytime activities. This was taken to mean that there are inadequate services in place to help individuals with ASD but not cognitive impairment, and additional services must be put in place to help these individuals succeed.

Comorbid Features. In addition to the features of ASD necessary to make a diagnosis, several additional features are frequently present in individuals with ASD. These characteristics are not necessary to make a diagnosis of ASD, but contribute to the difficulties of individuals with ASD.

Cognitive Impairment. Cognitive impairment, while not required for a diagnosis of autism, is observed in many individuals who receive the diagnosis. When diagnosing cognitive impairment, several factors must be considered in order to accurately differentiate it from ASD.

Many behaviors included observed in ASD also occur in individuals with other forms of cognitive impairment. When an individual exhibits cognitive impairment, a comparison between their acquired skills must be made. The child's social communication and interaction skills must be compared to their other skills, such as fine motor skills, nonverbal problem solving, and processing speed. A diagnosis of ASD is appropriate when communication and interaction skills are significantly impaired in comparison to these other skills. Throughout the lifespan, cognitive abilities may also develop and change in differing ways. Howlin and colleagues (2004) found a decrease in performance IQ on the Wechsler Adult Intelligence Scale (3rd Edition) from children to adults, but an increase in verbal IQ. These findings were confirmed by Mawhood and colleagues (2000). An earlier study by Lockyer and Rutter (1969) also found a decrease in performance IQ from the age of 5 to the age of 15, but also found a decrease in verbal IQ. It is important to be mindful of the differences between ASD and cognitive impairment for the purposes of this study. Participants in the present study will display varying levels of cognitive ability. Because a diagnosis of ASD is dependent on the differential between communication and interaction skills and other skills, it will be essential to consider the performance of participants as they relate to the individual's other cognitive abilities. Theory of Mind skills that would be observed in a typically developing four year old would be qualitatively different when observed in an individual with a mental age equivalent of four years old versus an individual with a mental age of twelve, even if both individuals are the same chronological age.

Central Coherence. It has been pointed out that the "mindlessness" theory only accounts for impairments in the areas of socialization, communication, imaginative play, and repetitive interests. One theory suggests that this is due to executive functioning deficits in individuals with autism. Another theory is the weak central coherence theory. Central coherence is the

ability to see the whole, and use context clues to understand situations. Difficulty in seeing the whole, as opposed to the component parts, has been suggested to account for autistic individuals' superior performance on tasks that require attention to detail, such as the Block Design subtest of the WISC. While weak central coherence may produce benefits in cases where processing of parts instead of wholes is beneficial, it creates deficits in situations where gestalt-type processing is needed, such as facial recognition tasks. In a task requiring individuals with ASD to select from several homographs to complete a sentence, participants in the study were found to have difficulty in making use of contextual clues, and selecting the correct homograph. Another study is noted wherein an individual with savant drawing ability started with a small detail and expanded from that detail contiguously, supporting the superior attention to detail. It is suggested that it is this deficit in consolidating information into a cohesive whole that explains why autistic individuals can pass 2nd order Theory of Mind tasks in isolation, but struggle when presented in a story format or other applied setting (Frith, 1994).

Central coherence has been suggested to be related to Theory of Mind, in that individuals must be able to make use of available information, and combine it into a centralized picture, in order to accurately use Theory of Mind abilities. Some individuals on the autism spectrum possess a strong ability to perceive details. This attention to detail is hypothesized to be indicative of the perception of each part of a situation to be an individual factor, rather than a part of a whole. This can lead to the individual to become distressed when small details are changed, because in their view, the entire situation is different. This has been termed "weak central coherence". Individuals with weak central coherence have an inability to integrate information into a cohesive whole, which, in turn, impairs their ability to make use of context. This may also explain the observed difficulty in generalizing skills to new situations. Research

in the area has shown that individuals with ASD have superior local processing abilities, but evidence that this advantage comes at the cost of global processing abilities has been mixed, leading to an emphasis on the former. These individuals have been shown to be able to integrate information when given explicit instructions (Happé, 2006).

These deficits are believed to be due to weak central coherence. A 1997 study by Happé used a task that required individuals with autism to read a sentence containing a homophone. Data were collected regarding whether or not the correct homophone was used, and whether the context of the sentence came before or after the decision needed to be made. It was found that the individuals with autism, compared to the control group, despite having what appeared to be superior reading skills, did not make use of the sentence context in order to select the correct homophone (Happé, 1997).

Comprehension in communication requires integration of all the information in a conversation, or global processing. The Weak Central Coherence theory of autism states that these individuals only process locally, leading to difficulty understanding conversational subtleties. A study examining children with ASD administered the *Wechsler Preschool and Primary Scale of Intelligence – 3rd edition* (WPPSI-III) Block Design subtest in order to evaluate their ability to synthesize information. Selected stories from the *Clinical Evaluation of Language Fundamentals -4th edition* (CELF-4) were used to measure verbal comprehension. A task was created where several short stories were read to the participants, and 3 questions were asked: a factual question to verify that the child had been paying attention, a script question to measure the child's ability to take the environmental context of the story into consideration when answering, and a propositional question to measure the participant's ability to draw logical conclusions from provided premises. The group with ASD was found to perform significantly

poorer on the global processing task, the script question, and equally on the local processing task, the propositional question. It was concluded that the children with autism failed to employ global processing to utilize context clues, indicating weak central coherence (Nuske, 2011).

Executive Dysfunction. Executive functioning is a term used to refer to a set of complex mental processes, including planning, working memory, inhibition, and mental flexibility. Children with autism have been found to experience executive dysfunction, which has been suggested to result in the restricted interests and repetitive behaviors found in ASD – features not explained by the weak central coherence or Theory of Mind deficit hypotheses. Mental flexibility, in particular, has been linked to stereotyped behavior. The Wisconsin Card Sorting Task and similar tasks have been shown to detect deficits in children with ASD. In these tasks, children with ASD struggle to identify higher numbers of categories, and commit higher numbers of errors on the tasks. This deficit was found to persist over time and across cultures (Ozonoff & Jensen, 1999; Hill, 2006). Additionally, a set-shifting task on the Cambridge Neuropschological Test Automated Battery has been used to evaluate mental flexibility in ASD. Results from this test indicated that "stuck-in-set perseveration," which is a fixation on a particular thought or topic, was only found in individuals with ASD, suggesting that, in contrast to other forms of disabilities, individuals with ASD experience a type of perseveration exclusive to the disorder (Hughes, Russell, & Robins, 1994).

Additional findings have suggested a relationship between executive dysfunction and other areas of impairment as well. A relationship has been shown between the ability to "self-start," or initiate actions independently, and adaptive functioning. This finding may also indicate that the measured deficit actually related to the ability to generate new information, or creativity. The latter explanation would provide some insight into the social difficulties of children with

ASD (Gilotty, Kenworthy, Sirian, Black, & Wagner, 2002). Some disagreement exists regarding the connection between executive dysfunction and other areas of functioning, such as adaptive skills or language ability (Corbett, Carmean, & Fein, 2009)

Executive dysfunction in children with ASD can have a number of effects on their learning in the classroom. Children experiencing executive dysfunction exhibit difficulty with response inhibition, which can result in impulsivity. Working memory is also negatively impacted, creating challenges in responding to the curriculum. A child with autism may be capable of these skills, but must deliberately activate them, and may not be able to do so when stressed, overwhelmed, or confused (Attwood, 2008).

Joint Attention. Joint attention is the ability for an individual to direct their attention to a stimulus to which another individual is also attending. Research by Tomasello (2003) suggests that joint attention provides the framework in which children and adults establish common ground, and the child learns to infer the adult's intentions. The affective theory of joint attention complements Tomasello, suggesting that joint attention requires sharing of attention – awareness – rather than simply looking at the same object as someone else at the same time (Hobson, 2005). Each of these theories support the idea that joint attention is a significant factor in Theory of Mind.

Joint attention is a skill that is typically impaired in children with ASD, and is essential for the development of social and language skills. This results in a lack of desirable behaviors, such as following gazes, social pointing, and gesturing. Each of these behaviors are features of an emerging awareness of others' minds – known as Theory of Mind - and ways to engage them. Studies of typically developing children have shown a high correlation between joint attention activities between children and parents and children's vocabulary. This has also been the case in

studies of children with ASD, finding that joint attention was predictive of language development, suggesting that joint attention plays a role in language development. This may indicate that the literature on language development can offer some ideas regarding language deficits in children with ASD. It also emphasizes the importance of interventions which target joint attention (Twatchman-Cullen, 2008).

Theory of Mind

Theory of Mind refers to an individual's understanding of their own thoughts, feelings, and emotions. It also encompasses the individual's ability to understand that other individuals may have their own unique set of thoughts, feelings, and emotions, which differ from their own (Baron-Cohen, 2001). The ability to understand the thoughts and feelings of others, as well as the ability to understand that those feelings may differ from the individual's own thoughts and feelings is an important component of social interaction.

Theory of Mind is also an important skill for social decision making. Appropriate responding in a social situation requires an individual to accurately assess the attitudes and expectations of those they are interacting with, again illustrating the importance of Theory of Mind for social interactions. A study by Leipold et al. (2013) examined the attitudes of participants toward other players in a game. As opposed to other measures of Theory of Mind, which require participants to observe a stimulus of some kind and comment on it, the Public Goods Game required participants to actually engage in the situation, allowing researchers to directly observe their interactional styles. Players were classified as having one of two styles of play: cooperative, or selfish. Players who were classified as "cooperative" were found to have a superior Theory of Mind to individuals who were classified as "selfish." The cooperative style of play could be described as more social, suggesting that individuals who are more able to

accurately judge the intentions and emotions of others are also more likely to be social (Leipold, Vetter, Dittrich, Lehmann-Waffenschmidt, & Kliegel, 2013)

In a study of individuals diagnosed with autism, those individuals who performed poorly on Theory of Mind tasks were also found to perform poorly on tasks designed to measure their everyday social competency (Peterson, Garnett, Kelly, & Attwood, 2009). Theory of Mind has also been shown to be essential for the development of social communication and interaction skills (Cutting & Dunn, 2006; Watson, Nixon, Wilson, & Capage, 1999), both areas which are impaired in individuals with autism (APA, 2013). Communication is rarely entirely explicit, and generally requires some level of inference on the part of the listener to decipher both the content and pragmatics of the language. Without the ability to infer the nonverbal content of a spoken or written message, it would be difficult to understand the intended meaning of the message, instead interpreting it on a literal level, possibly causing the intended message to be missed. Such a deficit has been observed in individuals with autism, which has been attributed to their failure to notice figurative speech. Theory of Mind is also required to engage in pretend play, another skill which requires the ability to take the perspective of another individual (Happé, 1995).

A 1991 study by Loveland and Kelley examined the relationship between Theory of Mind deficits and real world social interactions using a modified version of the Vineland Adaptive Behavior Scale, which is a psychometric instrument designed to evaluate an individual's ability to meet their own needs in real-world settings. Findings indicated that Theory of Mind deficits correlated with social dysfunction, as measured by the Communication, Daily Living, and Socialization scales of the Vineland. The observed deficits in daily functioning of individuals with autism is consistent with what might be expected of individuals who struggle

to understand the intentions and feelings of other individuals, confirming the idea that these deficits translate into actual difficulty in the individual's daily life.

While most individuals diagnosed with autism display Theory of Mind deficits, it is not considered to be a specific diagnostic criterion for autism spectrum disorder. This is due to multiple factors. Among these factors is the lack of specificity to autism. Children with intellectual disabilities, but not autism, may also fail these tasks (Yirmiya, Solomonica-Levi, Shulman, & Pilowski, 1996). Deaf and blind children have also been shown to experience delays in their Theory of Mind development (Peterson & Siegal, 1995).

Previous research has suggested that Theory of Mind deficits in individuals with mental or neurological disabilities are specific to individuals with autism, and that individuals with other disabilities, such as Down syndrome, develop Theory of Mind at the same rate as typical individuals matched for mental age. Additional research, however, showed mixed results, challenging the "similar structure hypothesis" – the idea that mental abilities such as Theory of Mind develop in line with mental age. The results of the study found that older adults with ASD performed similarly to mental age-matched children, but younger adults and adolescents with ASD did not. This suggests that, as a person ages, they may acquire experiences and knowledge that allow them to compensate for their lack of a Theory of Mind (Charman, Campbell, & Edwards, 1998).

The modularity view of Theory of Mind suggests that it is independent of other deficits found in ASDs. A module is defined as a system that operates independently from other systems and serves the purpose of creating information for the higher level systems that create perception. Vision, hearing, and the other senses are also defined as modules. Theory of Mind, therefore, is independent of other areas that may or may not be impaired, consistent with a

spectrum view of ASDs. While other, comorbid deficits may be observed in other areas of functioning, these occurred independently from Theory of Mind deficits, and likewise, Theory of Mind deficits can exist in absence of other deficits. This ability for deficits to occur independently of each other is used to explain the existence of higher functioning forms of ASD, and the associated patterns of strengths and weaknesses (Adams, 2011).

In a study evaluating the effects of interaction styles on emotional competence in typically developing preschool-aged children (three to five years old), results have indicated that it is the way mothers talk to their children, not their quality of attachment, which contributes to their emotional understanding. It was found that when mothers were trained to interact with their children by asking Wh- questions (who, what, where, etc.), providing detailed descriptions, and discussing emotions, their children showed increased understanding of emotional knowledge than control children (Van Bergen, Salmon, Dadds, & Allen, 2009). Additional studies confirmed the finding that it was the discussion of mental states and emotions that occurred between mother and child, not the quality of the attachment itself, which enhanced mental state understanding. Typically developing children who engaged in family conversations about emotions and their causes were found to enhance the children's understanding of others' feelings after three years compared to children from families that did not have the same conversations (Dunn, Brown, Slomkowski, Tesla, & Youngblade, 1991). Another major predictive factor is the degree to which conversations are connected. If more turns of a conversation flow from previous statements, children develop better understanding of mental states (Ensor & Hughes, 2008). When responding to children's misbehavior, consequences were found to negatively impact emotional understanding, whereas affective discussion contributed to it (how do you think that made them feel? etc.) (Rohrer, Cicchetti, Rogosch, Toth, & Maughen, 2011).

Assessment of Theory of Mind.

Several methods of evaluating Theory of Mind have been identified in the literature. The first of these, "eyes", presets photographs of a person's eyes, from which the individual must infer the mental state of the pictured individual. The eyes task measures primarily affective Theory of Mind. Affective Theory of Mind is also known as "hot" Theory of Mind, and refers to the ability to understand affective states, emotions, and feelings of others (Henry, Phillips, Ruffman, & Bailey, 2013). "False belief" tasks present the individual with a story featuring a protagonist with a false belief. The individual is required to demonstrate an understanding that the protagonist is not aware of the correct information, and what the protagonist will do based on the information they have. False belief tasks are used to evaluate cognitive Theory of Mind. Cognitive Theory of Mind is also known as "cold" Theory of Mind, and refers to the ability to understand thoughts, cognitive states, beliefs, and intentions of others (Henry et al., 2013). The third task, "stories", require an individual to explain the behavior of a character in a story by describing their underlying mental state. This task measures both affective and cognitive Theory of Mind. The "videos" task, as the name suggests, presents video clips of situations involving people, and requires the individual to infer both emotions and cognitions, though the task is more commonly used to evaluate affective Theory of Mind. The final type of task, "faux pas," evaluates the individual's ability to recognize socially inappropriate behavior. This type of task is used to measure both cognitive and affective Theory of Mind (Henry et al., 2013).

Because most Theory of Mind tasks are verbal in nature, most research regarding Theory of Mind with ASD has focused on High Functioning Autism (HFA) and Asperger's Syndrome (AS). To address this gap in research as well as to determine the role of language in

development of a Theory of Mind, Colle and colleagues (2007) gave a non-verbal Theory of Mind task to children with low-functioning autism (LFA). In order to be certain that any deficits that may be found were due to Theory of Mind deficits, not language ability, the LFA group was compared to a group with language impairment, which had similar levels of language ability, but were hypothesized to lack the Theory of Mind deficits found in ASD. It was found that, compared to a control group of students with Specific Language Impairment, children with low-functioning autism were found to have an impairment in their Theory of Mind. This is consistent with studies that have found Theory of Mind deficits in older children and adults with AS and HFA, suggesting that these deficits are present throughout development (Baron-Cohen & Hammer, 1997). The consistency across functioning levels also contradicted the idea that language development is central to the development of a Theory of Mind (Colle et al., 2007).

Measures for Adolescents and Adults. Discussion of Theory of Mind, particularly false belief tasks, most often refers to two levels, or orders, of beliefs (though orders beyond the second are possible). The first level, first order beliefs, refer to the beliefs of an individual regarding whether or not other individuals also have thoughts or beliefs similar to their own. The second level, second order beliefs, refers to the beliefs of a second individual regarding the first individual's belief (Obiols & Berrios, 2009).

Adolescents and adults have been found in several studies to perform as well or better than controls on Theory of Mind tasks, contradicting the idea that individuals with an autism spectrum disorder experienced deficits in their Theory of Mind, yet it is known that these individuals continue to have difficulty with social interactions, despite their improved performance. The tasks, generally first and second order false belief tasks, used to determine that there were no Theory of Mind deficits in adolescents and adults with an autism spectrum

disorder were considered not to be sufficiently sensitive to pick up deficits in individuals with a mental age above six. The tasks used in these studies were first-order and second-order Theory of Mind tasks, which typically developing four year olds and six year olds are capable of passing, respectively. These tasks, therefore, did not accurately evaluate whether a Theory of Mind deficit existed, but simply that the individuals tested were capable of passing tasks that an average six year old could pass (Baron-Cohen & Jolliffe, 1997; Beaumont & Newcombe, 2006; Happé, 1994). This has been speculated to be due to the difference in what is required to successfully complete a laboratory Theory of Mind task compared to the complexity of a real world social situation (Scheeren et al., 2013).

The Eyes Task involves viewing pictures of a person's eyes, and making a forced choice between mental state terms, some of which are simple, and some of which are complex. This was determined to be an appropriate measure due to the usage of mental state words that went beyond simple emotional recognition, and required more complex processing, such as "scheming", or "reflective". When compared to mental age-matched controls, adolescents and adults with HFA and AS displayed Theory of Mind deficits (Baron-Cohen & Jolliffe, 1997).

In order to more accurately measure Theory of Mind skills, several other measures have been developed that more closely approximate the demands of a real world social interaction. By creating measures which evaluate skills in a way that is closer to the demands of a real world social interaction, conclusions can be generalized to real world settings with more confidence, in contrast to previous tasks, which tested skills out of context, limiting the ability to draw conclusions about the real world implications of the results. The Strange Stories Task and Stories from Everyday Life tasks require participants to explain non-literal utterances, and the Reading the Mind in the Eyes Test and Reading the Mind in the Voice tasks require test-takers to

determine mental states from pictures of eyes, or voice tones, respectively. These tasks test skills in isolation, however, and do not require various stimuli to be integrated. This led to the development of the Awkward Moments Test which required participants to view TV commercials and answer Theory of Mind questions. Adolescents and adults with HFA and AS performed significantly poorer on these tasks when compared to control tasks which did not require Theory of Mind skills (Beaumont & Newcombe, 2006).

The Theory of Mind deficit hypothesis has also been supported by findings with the same age group using a more advanced Theory of Mind task, the Strange Stories task. This more advanced task required participants to evaluate skits or stories about situations involving people saying things that are not meant literally. Participants are required to fully describe the main character's mental state in order to receive full credit, rather than simply label what the person may be thinking, as would be required in Scheeren's second-order Theory of Mind tasks. Results indicated that (despite being able to pass traditional Theory of Mind tasks) participants with ASD struggled to successfully complete the task, supporting the Theory of Mind deficit hypothesis (Brent et al., 2004; Happé, 1994; Jolliffe & Baron-Cohen, 1999; Kaland, Møller-Nielson, Smith, Mortensen, Callesen, & Gottlieb, 2005; Velloso et al., 2013).

To mask Theory of Mind deficits, individuals with ASD may learn coping skills as they develop. It has been speculated that, due to this acquisition of coping mechanisms, individuals with ASD are able to pass the first and second order Theory of Mind tasks which had commonly been used to measure this skill, despite not actually having the skills purported to be measured by the tasks. The development of more advanced measures of Theory of Mind has allowed this skill to be measured in a way which more closely approximates the demands of a real world social situation. Using these more advanced measures has shown that, despite the coping skills

learned throughout development, deficits in this area are still present during adolescence and adulthood.

NEPSY-II. The NEPSY-II (Korkman, Kirk, & Kemp, 2007) is a neuropsychological assessment designed to evaluate children's performance in the areas of: Attention and Executive Functioning, Language, Memory and Learning, Sensorimotor, Social Perception, and Visuospatial Processing. Included in the Social Perception domain is the Theory of Mind subtest. The NEPSY-II Theory of Mind subtest assesses the ability of children ages 3-16 to understand mental functions such as belief, intention, deception, emotion, imagination, and pretending, as well as to understand the differing beliefs, thoughts, and ideas of others. The test is divided into two tasks, the first of which, the verbal portion, requires the child to assess another individual's point of view to correct answer questions. This requires the child to be able to understand differing perspectives, and correctly infer solutions based on the different perspective that a character has. The second portion of the test, the contextual portion, requires children to evaluate a picture depicting a social situation and correctly select a photograph of a face depicting the most appropriate emotion for the pictured situation. This requires the child to correctly evaluate the contextual cues of a pictured situation, and reasonably infer the feelings of an individual experiencing that situation. The child must then select a picture of a face that would accurately represent that emotion or feeling (Korkman et al., 2007).

The NEPSY-II has been validated for use with individuals with the DSM-IV diagnoses Autistic Disorder and Asperger Syndrome. It was found that individuals with Autistic Disorder or Asperger Syndrome performed worse on the Theory of Mind subtest when compared to typically developing control participants, though the difference was significantly more pronounced in the group diagnosed with Autistic Disorder (Korkman et al., 2007). An

independent study by Narzisi et al. confirmed these findings, showing that individuals with high functioning autism performed significantly worse than controls on the verbal portion of the Theory of Mind subtest. There was no significant difference in performance on the contextual portion, suggesting that, given adequate context clues, individuals with HFA are able to compensate for a lack of Theory of Mind (Narzisi, Muratori, Calderoni, Fabbro, & Urgesi, 2013).

Projective/Storytelling Assessments with ASD. The use of projective and story-telling assessments may be criticized due to the necessarily subjective nature of scoring. While these measures do, by their nature, have low test-retest reliabilities, the open-ended style of response is one of the greatest assets of these measures: allowing the personality of the examinee to be evidenced in their responses. Story-telling measures also have the advantage of directly measuring the target skill using stimuli that are identical to the skill being tested. Each of the story-telling measures, the TAT, CAT, and Roberts-2 measure the ability to generate positive outcomes to a situation by requiring the examinee to do exactly that, providing excellent face validity.

Thematic Apperception Test. The Thematic Apperception Test (Murray, 1943; TAT) was created for use in assessing social functioning and cognition in the general population between the ages of 5 and 79, defined as understanding the thoughts and emotions of other people. Administration of the TAT is conducted by presenting the examinee with a series of 20 picture cards. The examinee is then directed to explain the content of each picture card by telling a story, including the current scene, preceding factors, and a conclusion. The Emotional Investment and Understanding of Social Causality scales of the SCORS scoring protocol for the TAT (Westen, 1985) were determined to be the most useful predictors of autistic symptomology

in individuals between the ages of 19 and 54. Previous research has suggested that individuals with high-functioning autism respond to questions on the TAT with an equivalent number of mental state words, but significantly fewer causal mental state explanations, meaning that while the same number of words were being used by both groups, the HFA group made fewer attributions to the mental state of the individuals in the picture. This study hypothesized that the amount of ASD phenomena would negatively correlate to scores on the EI and SOC scales. This hypothesis was confirmed, suggesting more difficulty in social-emotional insight in individuals with autism. It was also found that participants with ASD tended to focus on one detail of the picture and construct a story around it, failing to see the "whole" of the picture, supporting the Weak Central Coherence hypothesis (Eurelings-Bontekoe, Zwinkels, Schaap-Jonker, & Edrisi 2011). While this has been shown to result in superior local processing abilities (Happé & Frith, 2006), it has been suggested to relate to Theory of Mind deficits by preventing individuals with ASD from incorporating relevant context clues in social situations, thereby limiting their ability to understand the thoughts and emotions of others (Happé, 1997; Nuske & Bavin, 2011).

Childhood Apperception Test. The Childhood Apperception Test (Bellick & Bellick, 1949) is similar to the TAT in its delivery of picture cards, and was developed with the idea that children, ages three through ten years old, would respond better to depictions of animals, rather than humans, allowing for more accurate results. Research did not support this conclusion, finding that there was often no difference between the TAT and CAT, and when there was, results favored the TAT. A set of picture cards depicting humans in identical situations to the animal cards was developed (termed the CAT-H) to determine whether children would respond to otherwise identical stimuli in differing patterns. Results again showed that there was generally not a difference between the two, but when present, favored the pictures depicting

humans. Despite a lack of support for the superiority of the CAT, the "pull" of each card for specific themes has led to its continued use among clinicians. Additionally, the use of the animal cards has also continued for multicultural purposes, due to the inability to determine the race, gender, or age of the pictured animals (Teglasi, 2010).

Administration of the CAT requires the presentation of ten picture cards with verbal directions to tell a complete story to explain the picture, including preceding factors and a conclusion. The content of the story is then scored on ten variables: main character, main theme, needs of the main character, environmental concept, interpersonal relationships, significant conflicts, nature of anxieties, defenses, adequacy of superego, and integration of ego (Teglasi, 2010).

Rorschach. The Rorschach (Rorschach, 1927) is a projective test used in the psychological assessment of individuals who have been referred for an evaluation of several possible conditions. In a study by Holaday and colleagues (2001) examining responses to the Rorschach test from individuals with HFA and AS, results supported the hypothesis that Rorschach responses can reflect the first diagnostic criterion: impairment in social functioning. This was evidenced by underreporting of human content or movement, and cooperative movement in humans or animals. Furthermore, the participants in the study also showed a tendency to have a deficit in positive coping strategies, which was interpreted to indicate unrewarding social relationships, social ineptness, and the inability to form and direct responses – a main criterion for a diagnosis of ASD under the DSM-5 (APA, 2013). While several other variables were not found to distinguish participants with Asperger syndrome from other clinical non-ASD participants, the responses of the participants were found to differ significantly from normative data in the areas of displaying or experiencing emotion, ability to establish and

maintain intimacy and closeness, and perseverations with cognitive inflexibility, all of which are associated features of ASD. The findings of this study support the idea that projective assessments may be able to detect the cognitive differences and emotional deficits of individuals with ASD.

Roberts-2. The Roberts-2 (Roberts & Gruber, 2005) is the second edition of the 1982 Roberts Apperception Test for Children. The measure is described by the authors as an instrument to assess social perception in two overall areas in children from ages 6-18: developmental adaptive function, and clinical function. The former evaluates prosocial behaviors and attitudes about the world around them. The latter evaluates social and emotional problems through the use of more unusual and atypical elements when expressing social understanding than typical peers. Similar to the Thematic Apperception Test, Childhood Apperception Test, and Rorschach, The Roberts-2 requires participants to view an ambiguous stimulus, and generate an explanation for what is pictured. Responses are then evaluated for the thematic content, and scored accordingly. The Roberts-2 is stated to be used for children or adolescents that have been referred for social or emotional adjustment problems, as well as in research settings where measures of social competence are required (Roberts & Gruber, 2005).

Administration of the Roberts-2 follows a similar procedure to the TAT and CAT, requiring participants to view picture cards of social interactions, and provide a narrative to explain the pictures. The advantage of story-telling measures such as the Roberts-2, TAT, and CAT is its depiction of humans (or anthropomorphic animals, in some cases with the CAT) as stimuli, instead of the abstract figures of the Rorschach. Particularly for children with a disorder such as ASD which often includes deficits in abstract reasoning ability, the ability to respond to

stimuli which do not require abstraction allows for information which is more relevant to the participant is valuable.

In addition to the advantages of the story-telling method of assessment, the Roberts-2 confers additional advantages over other storytelling methods. The SCORS protocol, discussed above, was developed after the development of the TAT, and was designed to be used in conjunction with the TAT. The Roberts-2 protocol, however, was designed by the test manufacturers with test development in mind. The CAT also relies on psychological constructs which are not measurable, and lack scientific validation, such as the ego and superego. In order to score the corresponding scale on the CAT, speculations must be made about internal processes which cannot be observed or quantified. While the nature of some scales on the Roberts-2 are necessarily subjective, the same materials can be experienced by a second observer who can then also assign a score. This is not possible with scoring criteria that require speculation regarding internal processes. The Roberts-2 is also the most recently updated measure of the three, thereby having the most contemporarily dressed characters in stimulus cards, maximizing the degree to which examinees can identify with the pictured characters.

Many of the scales of the Roberts-2 can be used to evaluate various dimensions of social competence and functioning (Roberts & Gruber, 2005). The Problem Identification and Resolution scales, in particular, require the individual to evaluate the internal states and processes of the pictured characters in order to achieve positive scores. The Problem Identification scale requires the child to provide a description of the problem present in the scene, including the feeling or behavior, reason for the feeling or behavior, preceding events, and reasons behind the feelings or behaviors of the pictured characters. The Resolution scale requires the child to provide a solution to the problem which adequately resolves the problem, as

well as addresses the concerns and emotions of the characters (Roberts & Gruber, 2005). To achieve a score on either of these scales, a child would require the ability to accurately interpret the pictured situation based only on the facial expressions and body language of the characters, as well as take the perspective of the characters to understand the situation and generate a prosocial outcome. Because scores on either of these scales require the ability to take another person's perspective and understand their thoughts, emotions, or intentions, an intact Theory of Mind is necessary.

At the time of writing, no studies employing the Roberts-2 with participants diagnosed with ASD could be found. Information on the Roberts-2 was obtained from the administration manual. The prior version, the RATC, has been a part of several studies which will be discussed.

Roberts Apperception Test for Children. The first version of the RATC (1982) has been used to evaluate multiple different personal factors. It has been found to be useful in evaluating depression (Joiner, 1996; McClelland, Koestner, & Weinberger, 1989; Shedler, Mayman, & Manis, 1993; Worchel, Rae, Olson, & Crowley, 1992), and pathology in sexually abused children (Babiker & Herbert, 1998). It has also been used to evaluate general emotional difficulties as part of a comprehensive emotional assessment (Handwerk & Marshall, 1998).

The Roberts-2 is a nationally normed instrument for use in evaluating social competence and awareness. Similar to the Strange Stories task, the Roberts-2 requires individuals to process pictured social situations and assign thoughts, feelings, intentions, and emotions to the pictured characters. While the Strange Stories task has experienced success in evaluating Theory of Mind deficits in individuals with autism, it is not a nationally normed measure, making it difficult to say with confidence how an individual's Theory of Mind compares to what might be expected of a typically developing peer in measurable terms. By using the newer version of the RATC, the

Roberts-2, to assess Theory of Mind abilities, it becomes possible to identify areas of weakness that might not be as readily apparent without a national comparison sample. The use of the Roberts-2 would also provide valuable opportunities to evaluate other areas of the individual's functioning that are assessed by the Roberts-2, including the ability to recognize available support resources, and the degree of depressed, anxious, aggressive, or rejected emotions being experienced by the individual.

CHAPTER III

METHODS

Participants

The present study included fourteen individuals with high-functioning autism spectrum disorder and fifteen control participants with no diagnosis of autism spectrum disorder. Each group's participants were from a sample of convenience. Clinical participants were recruited from a summer camp for children with autism spectrum disorders. To be enrolled in the summer camp, children must have received a diagnosis of autistic disorder or autism spectrum disorder from the DSM-IV-TR or DSM-5, respectively. Participants who have met the criteria for inclusion in the summer camp will be considered to have met the criteria for inclusion in the present study. Comparison participants were recruited informally through personal referrals of workplace colleagues. The age range of the participants was six to sixteen years old. All participants were administered a brief intelligence measure to determine their cognitive functioning level. Any individuals, clinical or nonclinical, with an IQ below 70 were not included in the present study.

Of the fourteen clinical participants evaluated as part of the study, five were not able to complete the Roberts-2 due to insufficient verbal abilities or behavioral issues which interfered with administration. Despite these participants possessing the necessary skills to complete the KBIT-II and score above an estimated IQ of 70, the KBIT-2 verbal tasks can be completed by receptively identifying pictures and providing single-word answers. Some participants lacked the more complex verbal skills to provide a narrative of Roberts 2 stories. In addition to the verbal weaknesses present in select participants, other participants lacked the behavioral inhibition necessary to remain seated and engaged throughout the evaluation. These participants

engaged in behaviors such as attempting to press emergency alert buttons in the testing room and threatening to destroy testing materials. In response to these behaviors, the testing session was ended. As a result, these five participants were not included, and the clinical group was reduced to nine.

Participants were tested after obtaining parental consent, participant assent, and a demographic survey. Per parent response to the demographic survey, 26 of 29 participants were identified as Caucasian. The remaining three were identified as Hispanic/Latino.

For the clinical sample, data was collected in a small Midwestern town from fourteen participants with Autistic Disorder or Autism Spectrum Disorder. They ranged in age from 6 years, 5 months to 11 years, 8 months. Of these individuals, five participants lacked the verbal ability necessary to complete the Roberts-2 and were not included for analyses. Eleven of the 14 children received some form of special education services. All 14 participants were male.

Data from typically developing peers was collected in a large upper-Midwestern city from 15 participants. They ranged in age from 6 years, 5 months to 16 years, 0 months. All 15 participants were appropriate for inclusion in analysis. Six participants were male and the remaining 9 were female.

Measures

***Kaufman Brief Intelligence Test – 2nd Edition* (KBIT-II)**

The KBIT-II (Kaufman & Kaufman, 2004) will be administered to screen for verbal and nonverbal ability level to ensure that participants will be able to understand the required task. The KBIT-II is a brief intelligence test that was adapted from the *Kaufman Assessment Battery for Children, 2nd Edition* (KABC-II; Kaufman & Kaufman, 2004a) for use with individuals from age 4 years through 90 years, 11 months. The KBIT-II was standardized on a group of 2,120

participants between the ages of four and 90 years of age. Participants in the standardization sample were English-speaking, noninstitutionalized, and did not have physical, perceptual, and psychological impairments that would interfere with their ability to complete the tasks. The norming sample was designed to match the US population on education level of the examinee if older than 26, or of the examinee's parents if younger than 26, race/ethnicity, and geographic region (Kaufman & Kaufman, 2004, p.24). The KBIT-II yields standard scores, which have a mean of 100, and a standard deviation of 15. This means that a score within 15 points of 100 (85-115) is considered average.

The KBIT-II consists of three subtests, two of which are verbal subtests, while the remaining is non-verbal. The first verbal subtest, Verbal Knowledge, requires the examinee to select a picture from an array of six which most closely represents the meaning of the word or answer to the question provided by the examiner. The second verbal subtest, Riddles, requires the examinee to employ verbal comprehension, reasoning, and vocabulary to determine the answer to a riddle based on two clues in the early portion of the test, and three clues in the later portion. The nonverbal subtest, Matrices, require the examinee to select the visual stimuli which best completes a visual pattern or analogy in a 2x2 or 3x3 grid. The KBIT-II is used to screen for intellectual difficulties in assessment procedures where intelligence is not the primary concern. Verbal items require one word answers, and visual items require the individual to point, making it appropriate for use by any professional with adequate training, as opposed to full-scale intelligence tests which requires a trained psychologist to use (Kaufman & Kaufman, 2004, p. 4).

Reliability. Internal consistency data on the verbal, non-verbal, and IQ composite scores was collected for the KBIT-II in the norming sample (Kaufman & Kaufman, 2004, p. 51-52).

Participants in the norming sample were divided into one group per chronological age year between ages 4 and 16, and larger groups for participants between 16 and 90. Mean Verbal split-half reliability coefficient was .91, and mean Composite IQ split-half reliability was .93. Nonverbal split-half reliability coefficients were in the .80s and .90s for groups older than 5. 4 and 5 year old groups had a split-half reliability of .78. Mean test-retest reliability for the Verbal composite fell at .91, .89 for the Nonverbal composite, and .90 for the Composite IQ.

Validity. Raw scores on Verbal (crystallized) tasks increased throughout the lifespan. Raw scores on Nonverbal (fluid) subtests peaked in middle age, and declined after, matching developmental trends. Comparison validity studies yielded correlations of .47 (nonverbal), .77 (verbal), and .80 (IQ Composite) with the KBIT (First edition). The KBIT-II correlates with other existing measures of cognitive abilities in the moderate to high range. Correlations with the Wechsler Abbreviated Scale of Intelligence ranged from .80-.86 (Verbal), .62-.80 (Nonverbal), and .81-.90 (IQ Composite/FSIQ). Correlations with the Wechsler Intelligence Scale for Children, 3rd and 4th editions were, respectively, .83 & .79 (Verbal), .53 &.56 (Nonverbal), and .76 & .77 (IQ Composite/FSIQ). The WISC-IV GAI, which excludes the processing speed and working memory indices correlated to the KBIT-II IQ Composite at .84 (Kaufman & Kaufman, 2004). The correlation to existing measures of cognitive abilities provides strong evidence of construct validity (Madle, 2007).

NEPSY-II: Theory of Mind Subtest

Overview. The NEPSY-II (Korkman, Kirk, & Kemp, 2007) was developed from a foundation of developmental and neuropsychology theory and practice. It contains subtests designed to measure a number of areas of functioning, such as attention, executive functioning, language, memory, learning, sensorimotor, social perception, and visuospatial processing. The

Social Perception domain is composed of tasks that measure Affect Recognition and Theory of Mind (Korkman et al., 2007).

The NEPSY-II (Korkman et al., 2007) was standardized on a group of 1,200 participants between the ages of three and 16 years of age. The norming sample was designed to match the US population on race/ethnicity, and geographic region. Participants were divided into four categories of parent education (0-11, 12, 13-15, or 16+ years of education). The NEPSY-II Theory of Mind subtest yields percentile ranks. Because providing a specific percentile rank was believed to be misleading, percentile rank ranges are provided (<2, 2-5, 6-10, 11-25, 26-50, 51-75, and >75). Percentile ranks indicate the percentage of participants who were at or below the obtained score. For example, a percentile rank of 50 would indicate that 50% of participants obtained scores that were equal to or lower than the obtained score. (Korkman et al., 2007)

The Theory of Mind task is divided in two parts: a verbal portion, and a contextual portion. The verbal portion of the *NEPSY-II* Theory of Mind subtest measures the participant's ability to "understand mental functions, such as belief, intention, deception, emotion, imagination, and pretending, as well as the ability to understand that others have their own thoughts, ideas, and feelings that may be different from one's own." The contextual portion requires participants to select a photograph of a face that they feel accurately represents a social situation in which they cannot see the face of the target individual. This measures the child's "ability to recognize facial affect and to understand how emotion relates to social context and to recognize the appropriate affect given various social cues," (Korkman et al., 2007, p. 22).

Reliability. Internal consistency data for the NEPSY-II Theory of Mind subtest indicated reliability coefficients that range from .76 in four and five year olds to .84 in six and seven year

olds. In the clinical sample, reliability coefficients range from .84 in five year olds to .86 in six year olds. Test-retest reliability for the Theory of Mind subtest was .77/.84

Validity. Subtest intercorrelation between the Theory of Mind subtest and other NEPSY-II subtests were generally low, ranging from .03 to .41. Correlation within the Social Perception domain, which contains the Theory of Mind subtest, was small (r = .21). Correlation with the Differential Ability Scale for Children, 2nd edition (Elliot, 2012) clusters ranged from .23 to .38. Correlation with the overall composites was in the moderate range, particularly with the special nonverbal composite (.53). Correlational analyses with other cognitive measures did not include the Theory of Mind subtest due to a lack of an analogous measure.

Roberts-2

Overview. The Roberts-2 (Roberts & Gruber, 2005) is a measure of social cognitive skills for use with children and adolescents aged 6 to 18 years old that have been referred for "social or emotional adjustment problems". To assess these areas, the Roberts-2 contains scales to evaluate a child's ability to tell a complete story, identify available resources in their environment, recognize and appropriately identify conflicts, and generate prosocial solutions to problems. Additionally, the Roberts-2 contains scales which note negative or bizarre content, negative solutions to problems, and negative emotions present in the stories provided. The Roberts-2 manual states that it is appropriate for use in research related to social competence. Social cognitive skills are assessed through an evaluation of the child's use of expressive language. The Roberts-2 is used for two major assessment purposes. The first is to evaluate children's changes as they "grow older and more socially experienced," and the second is to evaluate social and emotional problems through the use of "unusual and atypical elements when expressing social understanding" when compared to typically functioning peers. These areas are

known as “developmental adaptive function” and “clinical function”, respectively. Both functions evaluate the child’s response in a hierarchy of quality of social understanding. The Roberts-2 was standardized on a group of 1,060 participants between the ages of six and 19 years of age. Participants in the standardization sample were selected to match the US population on parental education level, race/ethnicity, and geographic region (Roberts & Gruber, 2005, p. 127). The Roberts-2 yields T-scores, which have a mean of 50, and a standard deviation of 10. This means that a score within 10 points of 50 (40-60) is considered average. For developmental/adaptive scales, a score below 40 or above 60 is considered clinically significant. A score below 40 indicates that the participant’s answers contained the targeted feature or theme at a lower rate than would be typical for a peer, while a score above 60 indicates a higher rate. For the Problem Identification and Resolution scales, scores above 60 on lower-ranked subscales may indicate that the participant is providing answers that are more simplistic than what would be expected of a peer. For clinical scales, a score above 60 is considered clinically significant, indicating that the participant’s answers involve negative emotions (on the Emotions subscales) or negative outcomes (on the Outcomes subscale) at a higher rate than what would be expected of a peer. Validity evidence is presented that 27 of 28 scales are able to differentiate across ages, as well as between typical and clinical samples. The only scale that was not shown to be able to differentiate across ages or clinical status was the Depression scale. (Medway, 2010).

The Roberts-2 uses 16 picture cards showing different examples of situations the child or adolescent is likely to encounter on a regular basis, such as the behavior of peer groups, or family interactions. The child is asked to provide a narrative explaining what is happening in the scene in a complete story which has a beginning, middle, and end, and has some reference to

how the characters are feeling. The transcript of this story is then scored on several scales to reflect the child's social cognitive competence.

The Theme Overview scales are composed of the Popular Pull (POP) and Complete Meaning (MEAN) Scales. The POP scale is defined as how the majority of non-referred participants perceive the picture. The MEAN scale is determined by the participant's incorporation of a backstory that led to the scene, description of the scene, feelings and interactions of the featured characters, and a resolution of the scene. To receive credit for MEAN, the narrative must also receive credit for POP, and Problem Identification and/or Resolution credit of 2 or higher. There must also be no atypical content.

The Available Resources scales are composed of the Support Self—Feeling (SUPS-F), Support Self—Advocacy (SUPS-A), Support Other—Feeling (SUPO-F), Support Other—Help (SUPO-H), Reliance on Other (REL), and Limit Setting (LIM) scales. The SUPS-F scale applies to cards in which the character experiences positive feelings such as happiness, love, pride, or pleasure. The SUPS-A scale is scored when characters support themselves, learn from an experience, or generally solve a problem independently. The SUPO-F scale applies to cards in which the environment or other characters respond supportively to the main character, as opposed to the SUPS-F scales which are focused on the individual. The SUPO-H scale is scored when there is a helpful interaction between two or more people. REL relates to the character seeking help from others. LIM is related to consequences or punishment. Any time a character is disciplined, or has rules enforced on them in some way, LIM is scored.

The Emotion scales address four categories: Anxiety (ANX), Aggression (AGG), Depression (DEP), and Rejection (REJ). The ANX scale reflects fearful or anxious responses, commonly on the themes of death, illness, guilt, or worry. AGG is scored when aggressive

content is present, both feelings and actions. Any content related to angry/mad feelings, verbal aggression, or physical aggression is scored for AGG. DEP is scored when content contains emotional responses such as sadness, depression, crying; physical symptoms such as tiredness or apathy; or hopelessness or a failure to resolve the conflict. REJ is a measure of social isolation. The central theme of REJ is disturbances in attachment, such as physical separation, or a divorce.

The Outcome scales deal with the manner in which the situation is resolved, and can be scored out of four possibilities: Unresolved Outcome (OUT1-UNRS), Nonadaptive Outcome (OUT2-NON), Maladaptive Outcome (OUT3-MAL), and Unrealistic Outcome (OUT4-UNRL). Unresolved outcomes are those that leave the story in the present, and do not provide a conclusion. Nonadaptive outcomes provide a conclusion that does not resolve the problem presented in the narrative. Maladaptive outcomes are those that make the situation worse, rely on violence or death to resolve the conflict, or rely on antisocial behavior such as manipulation. Unrealistic outcomes are a result of fantasy or wishful thinking. Unrealistic outcomes are generally positive, but are not plausible solutions, such as magic, or sudden good fortune.

Unusual or Atypical responses fall into three categories: Refusal, No Score, and Antisocial. The Refusal scale reflects resistance to cooperation or cognitive delays. The No Score scale is used when the participant is unable to produce any scoreable content. The Antisocial scale reflects actions that are against the accepted rules of a situation, such as truancy. The Atypical category identifies content that significantly deviates from nonreferred children. These include illogical thought (ATYP1), misidentification of the theme (ATYP2), misidentification of a person (ATYP3), violence or aggression (ATYP4), physical abuse, sexual abuse, or neglect (ATYP5), imaginary content such as monsters (ATYP6), death of a main figure

(ATYP7), any kind of sexual content (ATYP8), or any other unusual content or clinically significant responses (ATYP9)

Problem Identification scale. The Problem Identification scales form a hierarchy wherein Problem Identification 1 (PID1) reflects a vague recognition of a problem, while Problem Identification 5 (PID5) reflects a fully developed identification of the problem with all features fully articulated, as well as the feelings of the characters, and the cognitive states of the characters.

In the first level, Problem Identification 1, the participant makes a nonspecific reference to a behavior or feeling with no explanation of the preceding factors, and no specific problem referenced. The reasons for the behavior or feeling are very poorly explained, if at all. The next level, Problem Identification 2, increases the degree to which the present situation is described, but does not describe the preceding factors. The situation is described and explained, and the characters' feelings are described. Problem Identification 3 includes a limited description of the preceding factors, and may include descriptions of the characters' cognitive processes, which may include cognitive terms relating to the character thinking, understanding, or wondering. Problem Identification 4 and Problem Identification 5 include the same components as the previous three levels, but have an increased level of detail assigned to each component.

Problem Identification is stated to require participants to demonstrate accurate perception and insight to identify the causes of emotional responses, which in turn contributes to the ability to solve problems. Lower level scores on this scale are considered to reflect an inability to develop insight. Both cognitive and emotional Theory of Mind are necessary abilities to demonstrate adequate insight in a social situation. Therefore, because the Problem Identification

scale is stated to depend on insight to achieve higher level scores, the Problem Identification scale was determined to be appropriate for consideration as a measure of Theory of Mind.

Resolution scale. The Resolution scales also form a hierarchy, reflecting the increasing complexity of problem-solving skills. Resolution 1 (RES1) refers to endings that are simple and do not explain how the ending came about and have no continuity from the story. Resolution 5 (RES5) scores reflect a fully elaborated resolution, addressing both feelings and the situation.

The lowest Resolution level, Resolution 1, reflects a simple, easy ending to the pictured conflict. No explanation is offered for how the resolution came about. The story concludes by abruptly switching to a poorly defined outcome, with no realistic continuity. The next level, Resolution 2, also provides no description of how the ending occurred, but does bear some relation to the content of the problem. Resolution 3 adds some degree of explanation to the outcome, but does not address the feelings of the characters, unless the problem was defined only in terms of those feelings. As with Problem Identification 4 and 5, Resolution 4 and 5 contain all the elements of a positive resolution, with Resolution 5 reflecting a higher level of detail and complexity.

Higher levels of Resolution scores are attributed to increased insight. The higher scores on this scale attend to the feelings of the characters in addition to the resolution itself, which taps into emotional Theory of Mind. While recognition of the characters' feelings will facilitate a higher score at all levels, it is a requirement of the fourth and fifth levels of Resolution that feelings be addressed. Because an understanding of emotions is required to achieve higher levels on the Resolution scale, it is appropriate to evaluate its use as a measure of Theory of Mind, in addition to the Problem Identification scale.

Reliability. In a study of agreement between multiple scorers, interscorer reliability was .8 or higher with the exception of the SUPS-F (.78), PID2 (.64), and RES1 (.79) scales. Test-retest reliability estimates were found to be "reliable." No additional reliability information was provided by the publisher, and independent studies confirming the reliability of the Roberts-2 could not be located.

Validity. A MANOVA was performed using all of the Roberts-2 subscales to verify its two primary functions. "The developmental factor was established by age in years, producing a test with 12 levels (6 years through 17+ year-olds separately)." "The clinical factor was established at two levels that distinguished the standardization sample from the referred clinical sample," (Roberts & Gruber, 2005, p. 135).

Scale validity indicated that both clinical and developmental trends were present for POP and MEAN, though the trends were different from each other. The four Support scales show a mild to moderate developmental trend on the SUPS-F, SUPS-A, and SUPO-F scales. A second trend indicates lower usage of this scale in the referred group. The REL and LIM scales showed no developmental or clinical trends. The PID scales show a developmental trend, with older participants being more likely to provide an answer that falls at a higher level of the hierarchy. PID2 was the most commonly scored scale. The youngest group scored PID1 4 times as frequently as PID4, with the inverse being true of the oldest group. In the referred group, the use of PID2 was consistent with the standardization sample, while all other PID responses were lower, indicating a lower likelihood for referred individuals to identify problems. A similar trend was observed in the RES scales, with RES1 and RES3 showing reciprocal changes with age. The developmental trend on the Emotion scale was found to be small to weak, while no clinical trend was found. OUT1-UNRS, OUT3-MAL, and OUT4-UNRL were observed significantly

more often in the clinical group, indicating a strong clinical trend. Differences between clinical and nonclinical children on the Unusual and Atypical responses indicate a very strong effect size, suggesting that these scales strongly reflect differences between the groups.

Procedures

Clinical participants (Group One) and typically developing peer participants (Group Two) were tested individually by the doctoral candidate with the assistance of graduate students trained in the administration of psychological assessment measures and under the supervision of the dissertation chair. Total testing time per participant was approximately 35-80 minutes. This wide range of administration times was due to some participants struggling with measures and requiring longer breaks than others, occasionally up to fifteen minutes. Five minute breaks between assessments were offered.

Participants were first administered the KBIT-II to confirm their cognitive functioning levels and to screen for verbal and nonverbal ability level to ensure that participants would be able to understand and complete the required tasks. The KBIT-II requires approximately 20-40 minutes to complete per participant depending on cognitive abilities. The KBIT-II was immediately scored to determine the individual's appropriateness for inclusion in the study when possible. When immediate scoring was not feasible, the examiner qualitatively analyzed performance to determine whether the participant was appropriate for inclusion. Participants were then administered both the NEPSY-II Theory of Mind subtest, and the Roberts-2. Half of the participants received the NEPSY-II first, while the other half received the Roberts-2 first to control for order effects. Roberts-2 story narratives were audio recorded for later scoring to ensure accuracy. Once responses were scored, participant recordings were destroyed.

The Roberts-2 was administered to evaluate the child's ability to incorporate elements of the picture to provide a narrative of the events, including the thoughts or feelings of the characters in the story. The Roberts-2 required approximately 10-20 minutes to administer. The NEPSY-II is a neuropsychological measure that includes an established Theory of Mind task, and this subtest was used to validate the results of the Roberts-2 in regards to the Roberts-2 being useful for discerning Theory of Mind development. The NEPSY-II Theory of Mind subtest requires approximately 5 to 10 minutes to complete.

Both the Roberts-2 and NEPSY-II have limited scoring with children between six and eight years old. The NEPSY-II offers a converted score for the Verbal task, but an overall Theory of Mind score cannot be obtained. The Roberts-2 does not provide converted scores for responses that achieve a rating of 4 or 5 on the Problem Identification or Resolution scales. Because a total score could not be obtained for all participants, and due to findings that the verbal portion of the NEPSY-II Theory of Mind test is more useful in detecting theory of mind deficits in children with ASD (Narzisi, Muratori, Calderoni, Fabbro, & Urgesi, 2013), only the Verbal score was used in analysis. No participants across groups provided responses that would have achieved scores in the 4 or 5 range on either scale of the Roberts-2, therefore these scales were excluded from analysis as well. A glossary of analyses and statistical terms used in the present study may be found in Figure 3.1.

Figure 3.1: Glossary of Statistical Terms	
Elastic Net Regularization	A regularized least squares regression method utilizing two forms of analysis (such as lasso and ridge regression). It is designed to identify only truly salient independent variables.
Theil-Sen Estimator	A non-parametric alternative to linear regression that is less sensitive to outliers. It estimates model parameters using ordered pairs of observations.

CHAPTER IV

RESULTS

Overview

The purpose of the present study was to investigate the potential utility of the Roberts-2 as a Theory of Mind measure for children and adolescents on the autism spectrum. Children and adolescents with ASD were administered a brief cognitive screening measure, the KBIT-II to ensure the appropriateness of their inclusion in the study, the NEPSY-II Theory of Mind subtest, a validated Theory of Mind measure, and the Roberts-2, a projective measure of social functioning. The same measures were administered to a group of typically developing peers for comparison purposes. This chapter presents the data collected and the results of the statistical analyses performed on this data.

Descriptive Statistics

Test scores were collected, and initial descriptive statistics were calculated for the KBIT-II, NEPSY-II, and Roberts-2. Descriptive statistics were calculated using the Statistical Package for the Social Sciences (SPSS) 22.0 program.

KBIT-II

The means, standard deviations, and ranges of the KBIT-II Verbal composite, Nonverbal composite, and IQ composite are presented in Table 4.1. The composite scores have a population standardized mean of 100 and standard deviation of 15. Average scores range from 85 to 115 (Kaufman & Kaufman, 2004). The mean of the IQ composite for the children used in the present study (n = 24) was largely average (M = 102.33, SD = 19.27). This was also true of the Verbal composite (M = 100.13, SD = 20.32) and Nonverbal composite (M = 103.13, SD = 17.07). Therefore, the means for the KBIT-II composite scores for this sample fell within the average

range, which was expected given the a priori inclusion criteria of participants with the verbal ability necessary to describe a picture verbally.

Table 4.1

Means, Standard Deviations, and Ranges of observed scores of the total sample on the KBIT-II

Composites	Mean	SD	Range
IQ	102.33	19.27	60 – 131
Verbal	100.13	20.32	56 – 132
Nonverbal	103.13	17.07	72 – 128

Means, standard deviations, and ranges were also calculated for each research group and are presented in Table 4.2. KBIT-II effect sizes are presented in Table 4.3. The mean of the IQ composite for the clinical group (n = 9) was in the lower average range (M = 87.89, SD = 17.53). Mean IQ composite for the comparison group (n = 15) was average (M = 109.93, SD = 15.06). The Verbal IQ composite mean for the clinical group was low average (M = 83.78, SD = 17.58). Verbal IQ composite mean for the comparison group was average (M = 109.93, SD = 15.06). The clinical group's mean score on the Nonverbal composite was average (M = 93.89, SD = 18.38), as was the comparison group's mean (M = 108.67, SD = 14.06). Both groups' mean scores were in the average range for all three composites, with the exception of the clinical group's Verbal composite, which was low average. Analysis of effect size using Cohen's D found a large effect size, suggesting that the difference in performance between the two groups was very significant. This is consistent with previous research which has found lower verbal abilities in individuals with ASD (Nader, Courchesne, Dawson, & Soulières, 2016; Nader, Jelenic, & Soulières, 2015; Rutter & Lockyer, 1967).

Table 4.2

Means, Standard Deviations, and Ranges of observed scores of each group on the KBIT-II

Composites	Mean		SD		Range	
	Clinical	Comparison	Clinical	Comparison	Clinical	Comparison
IQ	87.89	111.00	17.525	14.823	60 – 113	77 – 121
Verbal	83.78	109.93	17.584	15.059	56 – 106	68 – 132
Nonverbal	93.89	108.67	18.381	14.064	72 – 122	77 – 128

Table 4.3

KBIT-II effect sizes

	Cohen's d
IQ	-1.4238777
Verbal	-1.5974093
Nonverbal	-0.9031209

NEPSY-II

The mean and range of the NEPSY-II Theory of Mind Verbal task were calculated. The NEPSY-II yields scores in the form of percentile ranges and does not provide more precise scores (Korkman, Kirk, & Kemp, 2007). Because the percentile ranges provided by the NEPSY-II are not uniform, it was not possible to treat the obtained scores as continuous variables. Therefore, for the purpose of analysis, NEPSY-II scores were treated as categorical variables. The possible scores on the Theory of Mind task and their corresponding categorical assignments are detailed in Table 4.4. For simplicity, percentile ranges will be referenced by their numbered category for the remainder of the present study. The mean score of the participants in this study was in category 3. Observed scores had a minimum of category 1 and a maximum of category 6.

Table 4.4

NEPSY-II percentile ranges and categorical assignments for analysis

Percentile Range	Categorical Assignment
<2	0
2 – 5	1
6 – 10	2
11 – 25	3
26 – 50	4
51 – 75	5
>75	6

The means and ranges were also calculated for each research group and are presented in Table 4.5. The mean score for the clinical group was in category 2. Mean score for the comparison group was in category 5. The clinical group's scores had a minimum of category 0 and a maximum of category 5. The comparison group's scores had a minimum of category 3 and a maximum of category 6. Both groups' mean scores were in the average range for all three composites, with the exception of the clinical group's Verbal task score, which was low average. The majority of clinical participants (n = 6) fell at a comparison score of 1, while the majority of comparison participants (n = 7) fell at a comparison score of 7. No comparison participants obtained a score corresponding to a categorical assignment lower than 4. The relative frequency of each response quality is presented by group in Table 4.6.

Table 4.5

Means and Ranges (in categorical assignments) of observed scores of each group on the NEPSY-II

Group	Mean	Range
Clinical	2	1 – 5
Comparison	3	3 – 6

Table 4.6

Frequency of categorical score assignments of each group on the NEPSY-II

Score	Clinical	Comparison
1	6	0
2	0	0
3	1	0
4	0	2
5	0	2
6	2	4
7	0	7

To evaluate the difference between the two groups, Fisher's Exact Test was conducted, which is a method for analyzing the relationship between categorical variables when the assumption of expected frequencies of 5 or more in 75% of the table cells, which underlies the standard Chi-square test of association, is violated. The results of this analysis indicated that there was a significant difference in performance between the two groups (p = .000). Visual inspection of the frequency distribution of the data indicates that this analysis was significant due to the comparison participants scoring higher on average than clinical participants.

Roberts-2

The Roberts-2 Problem Identification and Resolution categories assign scores based on the complexity and thoroughness of the respondent's story with regard to these features. Scores between 1 and 5 can be achieved, with 1 being the most simplistic and 5 the most complex. Due to the rarity of scores of 4 or 5 in children under 8 in the norming sample, a T score cannot be obtained for these scales in children under 8. Therefore, Problem Identification 4, Problem Identification 5, Resolution 4, and Resolution 5 were not included for analysis. The means,

standard deviations, and ranges of the Roberts-2 Problem Identification 1, Problem Identification 2, Problem Identification 3, Resolution 1, Resolution 2, and Resolution 3 scales are presented in Tables 4.5 and 4.6. The population scale scores have a standardized mean of 50 and standard deviation of 10. Average population scores range from 40 to 60 (Roberts & Gruber, 2005). For the present sample, Problem Identification 1 had a high average mean (M = 63.54, SD = 13.66). Problem Identification 2 had a low average mean (M = 34.83, SD = 5.78). Problem Identification 3 was average (M = 42.79, SD = 5.49). Resolution 1 had an average mean score (M = 55.71, SD = 13.43). Resolution 2 fell in the low average range (M = 37.50, SD = 4.46). Resolution 3 was average (M = 45.42, SD = 4.13). The means for the Roberts-2 scale scores were low average to average when compared to population expectations.

Means, standard deviations, and ranges were also calculated for each research group and are presented in Tables 4.7, 4.8, and 4.9. Effect sizes of the Roberts-2 scales are presented in Table 4.10. The mean Problem Identification 1 score for the clinical group was (M = 59.00, SD = 14.18). The comparison group mean was slightly higher (M = 66.27, SD = 13.05). The mean Problem Identification 2 score for the clinical group was low average (M = 34.67, SD = 5.24), as was the comparison group mean (M = 34.93, SD = 6.26). The mean Problem Identification 3 score was average for both the clinical (M = 44.44, SD = 6.54) and comparison groups (M = 41.80, SD = 4.71). Mean Resolution 1 scores were average for the clinical group (M = 49.56, SD = 11.67) and the comparison group (M = 59.40, SD = 13.39). Mean Resolution 2 score was low average for the clinical group (M = 39.11, SD = 4.31) and the comparison group (M = 36.53, SD = 4.41). Mean Resolution 3 score was average for the clinical group (M = 45.33, SD = 2.35) and the comparison group (M = 45.47, SD = 4.98). Both groups' mean scores were in the average

range for all three composites when compared to population expectations, with the exception of the clinical group's Verbal composite, which was low average.

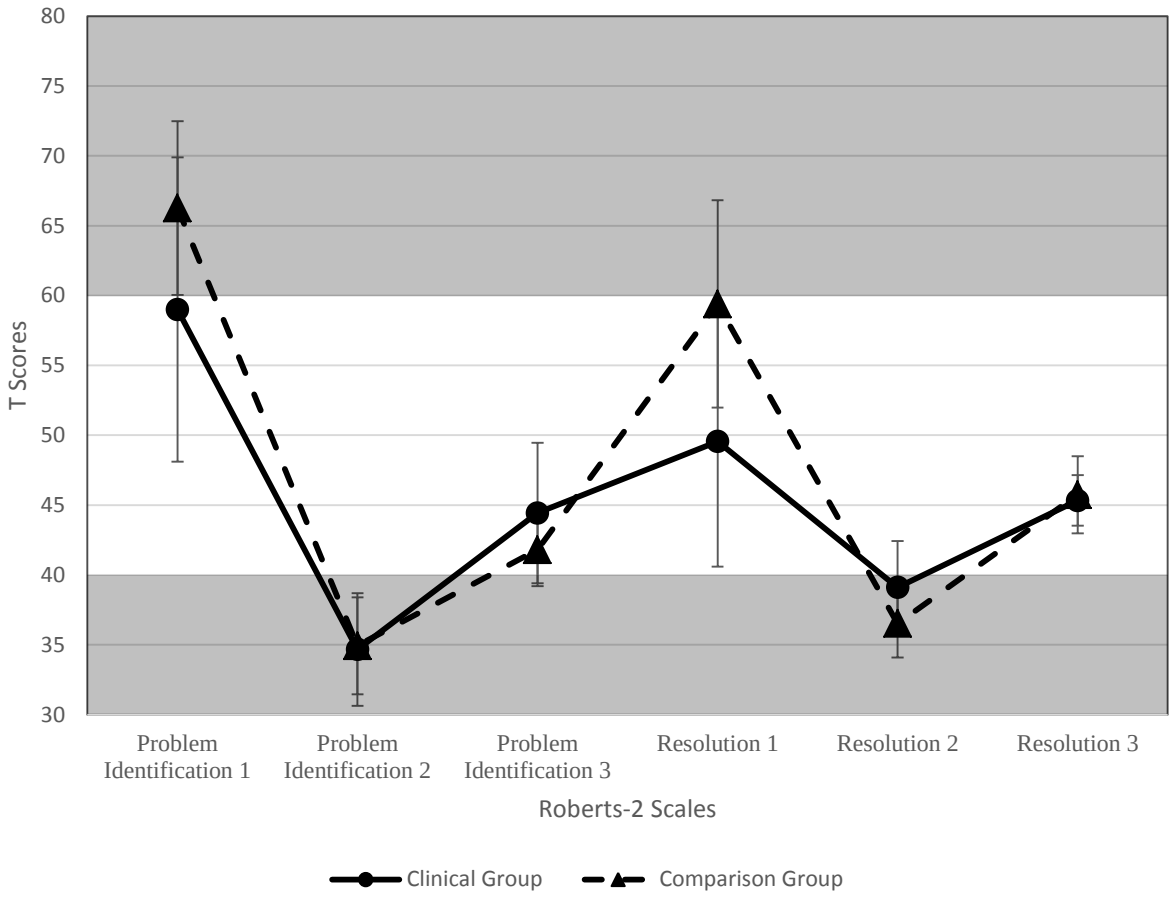

Figure 4.1 Profile Plot of Roberts-2 T Scores Means for clinical (ASD) and comparison (non-ASD) groups

Note. Clinically significant ranges (Above 60 or Below 40) shaded

Table 4.7

Means (Standard Deviations) and Ranges of observed T Scores of the total sample on the Roberts-2

Composites	Total Sample (*n = 29*)		
	M (SD)	95% CI	Range
Problem Identification 1 – Recognition	63.54 (13.66)	57.77 – 69.31	42 – 80
Problem Identification 2 – Description	34.83 (5.78)	32.39 – 37.28	30 – 53
Problem Identification 3 – Clarification	42.79 (5.49)	40.48 – 45.11	36 – 57
Resolution 1 – Simple Closure or Easy Outcome	55.71 (13.43)	50.04 – 61.38	38 – 80
Resolution 2 – Easy and Realistically Positive Outcome	37.50 (4.46)	35.62 – 39.38	32 – 49
Resolution 3 – Process Described in Constructive Resolution	45.42 (4.13)	43.67 – 47.16	39 – 61

Note. T Scores have a mean of 50 and a standard deviation of 10. Scores below 40 or above 60 are considered clinically significant; CI = Confidence Interval

Table 4.8

Means (Standard Deviations) and Ranges of observed T Scores of the clinical group on the Roberts-2

Composites	Clinical Group (*n* = 9)		
	M (SD)	95% CI	Range
Problem Identification 1 – Recognition	59.00 (14.18)	48.10 – 69.90	43 – 48
Problem Identification 2 – Description	34.67 (5.24)	30.64 – 38.70	30 – 47
Problem Identification 3 – Clarification	44.44 (6.54)	39.42 – 49.47	37 – 57
Resolution 1 – Simple Closure or Easy Outcome	49.56 (11.67)	40.58 – 58.53	38 – 73
Resolution 2 – Easy and Realistically Positive Outcome	39.11 (4.31)	35.80 – 42.43	35 – 49
Resolution 3 – Process Described in Constructive Resolution	45.33 (2.35)	43.53 – 47.14	43 – 48

Note. T Scores have a mean of 50 and a standard deviation of 10. Scores below 40 or above 60 are considered clinically significant; CI = Confidence Interval

Table 4.9

Means (Standard Deviations) and Ranges of observed T Scores of the comparison group on the Roberts-2

Composites	Comparison Group (*n = 15*)		
	M (SD)	95% CI	Range
Problem Identification 1 – Recognition	66.27 (13.05)	59.04 – 73.49	39 – 61
Problem Identification 2 – Description	34.93 (6.26)	31.47 – 38.40	30 – 53
Problem Identification 3 – Clarification	41.80 (4.71)	39.19 – 44.41	36 – 52
Resolution 1 – Simple Closure or Easy Outcome	59.40 (13.29)	51.98 – 66.82	41 – 80
Resolution 2 – Easy and Realistically Positive Outcome	36.53 (4.41)	34.09 – 38.97	32 – 46
Resolution 3 – Process Described in Constructive Resolution	45.47 (4.98)	42.71 – 48.23	39 – 61

Note. T Scores have a mean of 50 and a standard deviation of 10. Scores below 40 or above 60 are considered clinically significant; CI = Confidence Interval

Table 4.10

Roberts-2 effect sizes

	Cohen's d
Problem Identification 1 – Recognition	-0.5335107
Problem Identification 2 – Description	-0.0450405
Problem Identification 3 – Clarification	0.4632445
Resolution 1 – Simple Closure or Easy Outcome	-0.7868060
Resolution 2 – Easy and Realistically Positive Outcome	0.5917042
Resolution 3 – Process Described in Constructive Resolution	-0.0359548

Question 1

To address the first question, "Does performance on the Problem Identification and Resolution scales on the Roberts-2 correlate to Theory of Mind deficits in children with ASD on

the NEPSY-II Theory of Mind subtest after controlling for cognitive abilities?," regression analyses were conducted using the NEPSY-II Theory of Mind verbal task score as a dependent variable, and estimated overall intelligence (KBIT-II IQ Composite) and relevant Roberts-2 scales as independent variables. Regression analysis to answer Question One was conducted only on clinical participants. A multiple regression was performed using all six relevant Roberts-2 scales as the independent variables. For the clinical group, the regression model was nonsignificant (p = .561). Assumptions for the linear multiple regression were analyzed. An examination of QQ plots of NEPSY-II Theory of Mind Verbal task scores found that the assumption of normality was not met. Examination of the residuals plot indicated that the assumptions of homoscedasticity and independence were met. Visual inspection of the standardized residuals does not suggest a linear relationship between residual values and predicted values, though due to the small sample size it is not feasible to fit a nonlinear model. This model violated the assumption of a lack of multicollinearity, and therefore the regression was conducted using the elastic net estimator, which is a regularized least squares regression method suitable for use with small sample sizes (Shen, Han, & Braverman, 2016). Table 4.11 lists the Tolerance and Variance Inflation Factor for each scale. Tolerance values under .1 and Variance Inflation Factors (VIF) over 10 indicate possible problems with multicollinearity. With the exception of Resolution 3, all scales yielded a tolerance less than .2. Four of the six scales also yielded a VIF over 10. To address this problem with multicollinearity, separate models were analyzed using Problem Identification scales and Resolution scales independent of the other.

Table 4.11

Tolerance and Variance Inflation Factor for Roberts-2 scales

Scale	Tolerance	VIF
Problem Identification 1	.167	5.974
Problem Identification 2	.026	39.154
Problem Identification 3	.061	16.319
Resolution 1	.066	15.104
Resolution 2	.074	13.558
Resolution 3	.224	4.464

Analysis was conducted using the NEPSY-II Theory of Mind Verbal task score as an outcome variable. Predictors were Roberts-2 Problem Identification 3, Problem Identification 2, and Problem Identification 1 scales in the first model analyzed. The second model analyzed used the Resolution 3, Resolution 2, and Resolution 1 scales as predictor variables. Tables 4.12 and 4.13 list *R*, adjusted R^2, and regression coefficients for the Problem Identification and Resolution scales, respectively. KBIT-II IQ Composite scores were included in both models to determine the extent to which observed differences in NEPSY-II scores were accounted for by differences in estimated IQ. For the clinical group, the regression model was nonsignificant in both models ($p = .457$; $p = .373$).

Table 4.12

Multiple Regression Analyses of Relationships Between NEPSY-2, KBIT-II, and Roberts-2 Problem Identification scores

	R	AR^2	*B (S.E.)*	β	*t*	*p*
Variables	.459	-.082				
KBIT2 IQ Composite			.057 (.063)	.461	.905	.417
Problem Identification 3			.087 (.207)	.260	.418	.697
Problem Identification 2			.102 (.222)	.245	.459	.670
Problem Identification 1			.090 (.072)	.583	1.251	.279

Table 4.13

Multiple Regression Analyses of Relationships Between NEPSY-2, KBIT-II, and Roberts-2 Resolution scores

	R	AR^2	*B (S.E.)*	β	*t*	*p*
Variables	.586	.171				
KBIT2 IQ Composite			.009 (.071)	.076	.133	.901
Resolution 3			.248 (.519)	.267	.478	.658
Resolution 2			.434 (.291)	.859	1.491	.210
Resolution 1			-.028 (.071)	-.150	-.394	.714

Question 2

To address the second question, "Does the performance of individuals with ASD on the Roberts-2 differ in a consistent pattern from typically developing peers after controlling for estimated cognitive abilities?", a multivariate analysis of covariance (MANCOVA) was conducted. Prior to analysis, several assumptions were checked. Because data was collected from different participants, independence is presumed. Regarding multivariate normality, Problem Identification 2 possessed a high skewness and Resolution 3 had both high skewness and high kurtosis, suggesting that performance on these scales was not normally distributed. Overall multivariate normality was assessed using Mardia's Test (Kankainen, Taskinen, & Oja, 2004) and was found to be nonsignificant ($p = .53$). Box's Test was nonsignificant ($p = .13$), confirming the assumption of homogeneity of covariance matrices. Dependent variables were the six Roberts-2 scales. The independent variable used was the research group (clinical or comparison). KBIT-II IQ Composite score was entered as a covariate in order to control for differences between groups that resulted from differences in estimated IQ. No significant differences were found between groups $F(6, 16) = 1.06$, $p = .42$, suggesting that both groups performed relatively consistently across the six Roberts-2 scales.

Further evaluation of the present data was conducted using alternate methods of analysis suited to small sample sizes. Significance values could not be calculated due to limitations created by sample size. Elastic net regularization (Shen, Han, & Braverman, 2016) was used to obtain the sample intercept and coefficient values for each of the scales. Coefficients were then estimated using a nonparametric method of analysis known as the Theil-Sen estimator (Wilcox, 2001). Confidence intervals were obtained using bootstrapping. The coefficients, nonparametric estimated coefficients, standard errors, and confidence intervals of the intercept and seven scales are listed in Table 4.14. The relative magnitude of the slope coefficients suggests that each scale predicts comparatively little difference between participant scores compared to the intercept value. For example, the predictor with the largest coefficient, Resolution 3, represents a .336 unit increase in scores on the Resolution 3 scale for each 1 unit increase in the predictor. In contrast, the intercept representing the initial values without any predictors was -15.803. Further, analysis of the Thiel-Sen confidence intervals of each scale suggests that the greatest difference between groups occurred on the Resolution 2 scale.

Table 4.14

Regression Coefficients, Bootstrapped Coefficients, Standard Errors, and Confidence Intervals of KBIT-II, Roberts-2 Problem Identification, and Roberts-2 Resolution Scales

	B	*B (Thiel-Sen)*	*(S.E.)*	*C.I.*
Intercept	-15.803	-0.145	1.115	-0.612 – 0.553
KBIT2 IQ Composite	-0.011	0.053	1.667	-0.629 – 0.469
Problem Identification 1	-0.017	0.384	1.153	-0.250 – 0.674
Problem Identification 2	0.079	1.283	1.033	-0.404 – 0.792
Problem Identification 3	-0.109	-0.171	1.444	-0.283 – 0.643
Resolution 1	-0.059	-1.374	1.315	-1.452 – 0.196
Resolution 2	0.201	0.764	0.769	0.000 – 1.161
Resolution 3	0.336	0.722	0.758	-0.378 – 0.513

CHAPTER V

DISCUSSION

Theory of Mind is an essential component of social interaction, as it enables an individual to take the perspective of others and understand that the other individual may not have the same thoughts and feelings as they do. This ability is often found to be deficient in individuals with Autism Spectrum Disorder. The Roberts-2 is a measure of social functioning which requires the examinee to view an ambiguous social scene and provide a narrative of the events pictured, including a description of a possible problem and a constructive resolution to that problem. This task requires the examinee to take the perspective of the pictured individuals, which is a skill that uses Theory of Mind. The purpose of the present study is to investigate the Roberts-2's possible use as a Theory of Mind measure, as there are few existing norm-referenced Theory of Mind measures, and the Roberts-2 could also provide additional information about a child's social functioning.

One of the few norm-referenced measures of Theory of Mind is the NEPSY-II Theory of Mind subtest. Recent research has confirmed that the performance of children with ASD differ significantly in their performance on the NEPSY-II Theory of Mind subtest (Narzisi, Muratori, Calderoni, Fabbro, & Urgesi, 2013). Similar to the results found by Narzisi et al., the performance of children with ASD in the present study was significantly lower than the performance of comparison participants on the NEPSY-II Theory of Mind task. The performance of children on the Roberts-2 was compared to their performance on this subtest to examine the possible connection between Roberts-2 performance and observed Theory of Mind deficits. Individuals with average cognitive abilities can often learn the skills needed to complete Theory of Mind tasks, therefore a measure of estimated IQ was also administered.

Analysis of effect size indicated that there was a significant difference in the estimated IQ between clinical and comparison participants, therefore these scores were controlled for in order to ensure that observed differences in performance were not due to cognitive abilities. The present study found no significant correlation between Theory of Mind deficits, as measured by the NEPSY-II Theory of Mind verbal task, and performance on the Roberts-2 Problem Identification and Resolution scales. This suggests that performance on Problem Identification and Resolution scales is not dependent on Theory of Mind abilities, and that these scales do not provide meaningful insight into the development of this skill in children with ASD. While these findings do not provide preliminary support for the use of the Roberts-2 as a Theory of Mind measure, they do suggest that this measure is valid for its intended use with children with ASD. Additionally, findings suggest that clinicians using the Roberts-2 with children with ASD may not require special modifications or adaptations to the administration procedure.

In addition to possible correlations to Theory of Mind deficits, this study also sought to determine if the performance of children with ASD on the Roberts-2 differed in a consistent way from the performance of their typically developing peers. In order to demonstrate the validity of the Roberts-2 as a measure of social skill deficits in individuals with ASD, the performance of children with ASD on the Roberts-2 was compared to the performance of typically developing children and adolescents. Analysis indicated that there were no significant differences in performance on the Roberts-2 between children with ASD and typically developing peers. Previous research has demonstrated the validity of storytelling/projective measures to detect social and cognitive deficits in individuals with ASD (Holaday, Moak, & Shipley, 2001). This suggests that the Roberts-2 is not a valid measure for differentiating children with ASD from their typically developing peers based on their ability to identify and resolve a hypothetical

problem. Results do suggest that the Roberts-2 can be used with individuals on the spectrum for its intended purpose of evaluating social functioning. This should be explored further with an expanded and more representative sample.

Limitations and Directions for Future Research

The present study has several limitations. Limitations and recommendations related to the participants will be discussed first, followed by a discussion of limitations related to research design and measurement tools.

Participants

Participants for the present study were recruited from a sample of convenience. Random sampling was not conducted, which therefore limited the generalizability of the results obtained. In the absence of random sampling, it is difficult to establish whether the results obtained are truly representative of the larger population the samples represent, or are confounded by characteristics shared among the participants due to non-random sampling.

Both the clinical and comparison groups were relatively small. Small sample sizes limit the generalizability of research findings, and analyses of small sample sizes have less power, making them more susceptible to Type II error. The present data yielded a power of .419, suggesting relatively low power. It is possible that a larger sample size would have yielded significant findings. Larger sample sizes are more robust, allowing for more confidence in results.

While representative of the local population from which participants were drawn, the racial and ethnic composition of the present study was not representative of the American population as a whole. 92% of the participants in the present study identified as Caucasian, with the remaining 8% (n = 2) identified as Hispanic/Latino. There was no representation of African

American, Asian American, Arab American, Native American, or other ethnic groups in the present study, limiting the generalizability of the present study primarily to Caucasians.

Clinical and comparison group participants had several significant differences between them that may have introduced several other variables that could not be controlled. First, participants in each group were collected in different states. While the symptoms of ASD are relatively stable across cultures and regions, the differing education systems, special education supports, and mental health services available in different states may have an impact on the skill development of children. Second, the clinical group was made up of exclusively male participants, while the comparison group was 60% female. This introduces a variable (gender) which is not present in the clinical group, and limits the extent to which results across groups can be directly compared. Mean age was also significantly different between the two groups ($p < .05$). The mean age of the clinical group was 9 years, 5 months, while the mean comparison age was 12 years, 2 months. As Theory of Mind is a skill that develops with age, observed differences in NEPSY-II performance may be partially due to age at the time of testing.

Clinical participants were recruited through their involvement in a summer camp which provides therapeutic supports to individuals with ASD. This makes it likely that all individuals in the clinical group had received some degree of social and behavioral intervention before participation in the study, which may limit the generalizability of findings to those individuals who have received similar intervention services.

Future research on this topic should include a substantially higher number of participants. A higher number of participants will ideally result in a more ethnically representative sample, as well as provide closer mean group ages. Recruitment would ideally occur across several states for both clinical and comparison participants. If multi-state recruitment is not practically viable,

both groups should be drawn from the same locale at a minimum. Clinical participants should also be recruited from a variety of sources to allow for a more heterogeneous sample of previous emotional and behavioral interventions.

Measures

The KBIT-II is an abbreviated version of the KABC-II and is used as a screening tool to estimate cognitive abilities. This measure was chosen in order to reduce fatigue associated with the testing process, but is not as reliable of a measure of verbal and non-verbal reasoning abilities as a full cognitive battery would be. Individuals who were close to the cutoff point may have been included in this study who may not have been otherwise included with a more thorough evaluation.

Comparison across measures was somewhat limited due to the scoring system used by the NEPSY-II. The NEPSY-II provides percentile ranks as indicators of performance, while the Roberts-2 utilizes T scores. Percentile ranks do not utilize a ratio scale of measurement (i.e., the difference between the 1st and 2nd percentile is not the same significance as the difference between the 21st and 22nd percentile) and do not have standard deviations, in contrast to T scores, which use a ratio scale of measurement (i.e., the difference between a T score of 20 and 21 is the same as the difference between a T score of 30 and 31) and have a standard deviation of 10. In addition to the different scoring systems, the NEPSY-II does not provide a specific percentile rank which could be converted to a T score; instead, the NEPSY-II provides a percentile *range*, which is an imprecise estimate of performance. Additionally, percentile ranges are not evenly distributed (e.g., percentile ranks below the 25th are divided into 4 different ranges, while percentile ranks above the 75th occupy a single range). These substantial differences in scoring limit the usefulness of a direct comparison of scores.

Both the NEPSY-II and Roberts-2 had limitations at lower ages. While the minimum age for the NEPSY-II is four years old, the measure does not provide normative data for the Theory of Mind Total score for children under the age of eight. For children in the four to eight year old range, only the Verbal task score can be obtained. Similarly, while the minimum age for the Roberts-2 is six years old, due to a lack of normative data, converted scores cannot be obtained for the Problem Identification 4, Problem Identification 5, Resolution 4, or Resolution 5 scales if the child is in the six to eight year old range. While no children in the present study provided responses that would have achieved scores in the (4) or (5) range on the Problem Identification or Resolution scales, the lack of a NEPSY-II Theory of Mind Total score effectively excludes the participant's performance on the Contextual task. Research has suggested that the Verbal task is the more useful measure for children with ASD; however, the similarity of the Contextual task to the skills required for the Roberts-2 would allow for more direct comparison between the two measures.

Interpretation of the Roberts-2 Problem Identification and Resolution scales is more complicated than most psychometric measures. Overall composite Problem Identification and Resolution scores are not provided. Instead, each complexity level within each scale receives an individual T score, which limits the ability to compare a consolidated "Problem Identification" or "Resolution" level of performance. Each qualitative level of performance must be individually examined. While none of these comparisons were significant in the present study, if such comparisons had been significant for some qualitative levels of Problem Identification or Resolution but not all, the meaning of these findings would be unclear.

Finally, qualitative analysis of Roberts-2 responses suggests that administration guidelines do not adequately support thorough responses to test stimuli. The instructor is

directed to provide as many prompts as necessary to achieve a complete response on the first two pictures. From the third picture through the end of the measure, the instructor is not to provide prompts of any kind, and is only to ask for additional information if it is not clear to whom the examinee is referring in their narrative (i.e., saying "he" without pointing when the card shows more than one male character). Although most participants were able to successfully utilize the prompts provided on the first two items to achieve complete narratives, the complexity of the story substantially decreased for nearly all participants once these prompts were no longer provided per administration guidelines. Most participants provided simple descriptions of the picture that did not meet criteria for even the lowest level of complexity on these scales. For example, participants would provide responses such as "the boy and the girl are talking," or "he's doing his homework," before indicating that their narrative had concluded and they were ready to move to the next picture. Therefore, it is unclear if the lack of significant findings in the present study were the result of a confirmation of the null hypothesis, or if administration guidelines simply did not provide the structure necessary to elicit responses from participants that demonstrate the skills of interest.

In addition to the changes to subject recruitment and participation recommended above, future research should also further restrict the age range of participants in studies which make use of these measures. Participants should be a minimum of eight years old in order to make use of all possible scores on both the NEPSY-II and Roberts-2. Researchers may also wish to consider the use of a full cognitive battery in place of a screening tool such as the KBIT-II, though substantial modification to administration procedures would be required due to the significantly increased total testing time that a full cognitive battery would create. Future researchers may also wish to consider establishing an additional standardized administration

procedure for the Roberts-2 in which, after administering the Roberts-2 according to administration guidelines, examiners return to pictures which received minimal responses and provide the same prompts allowed for the first two items. This increased structure may elicit the desired skills (or deficits), and provide the first steps toward an administration protocol that can be used specifically for examinees with suspected deficits in their Theory of Mind.

Conclusion

In summary, the Roberts-2 did not detect Theory of Mind deficits in children with Autism Spectrum Disorder, nor did it differentiate children with ASD from their typically developing peers. Though results were not significant, the results obtained are influenced by several limitations. Future research which makes use of a larger, more representative sample, utilizes more robust measures of abilities, and implements additional procedures to ensure a more accurate representation of true ability may produce different results. As this is the first study to examine the use of the Roberts-2 to evaluate the skills and abilities of children with Autism Spectrum Disorder, this was an important step in identifying a wider breadth of assessment tools.

REFERENCES

Adams, M. (2011). Explaining the Theory of Mind deficit in autism spectrum disorder. Philosophical Studies, 163(1), 233-249.

American Psychiatric Association. (1952). Diagnostic and statistical manual of mental disorders (1st ed.). Washington,: American Psychiatric Association.

American Psychiatric Association. (1968). Diagnostic and statistical manual of mental disorders (2nd ed.). Washington,: American Psychiatric Association.

American Psychiatric Association. (1980). Diagnostic and statistical manual of mental disorders (3rd ed.). Washington,: American Psychiatric Association.

American Psychiatric Association. (1987). Diagnostic and statistical manual of mental disorders (3rd ed., text rev.). Washington,: American Psychiatric Association.

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders (4th ed.). Washington,: American Psychiatric Association.

American Psychiatric Association. (2000). Diagnostic and statistical manual of mental disorders (4th ed., text rev.). Washington,: American Psychiatric Association.

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

Attwood, T. (2008). An Overview of Autism Spectrum Disorders. In K.D. Buron & P. Wolfberg, (Eds.), *Learners on the Autism Spectrum*. Shawnee Mission, KS: Autism Asperger Publishing Co.

Babiker, G., & Herbert, M. (1998). Critical issues in the assessment of child sexual abuse. *Clinical Child and Family Psychology Review*, *1*(4), 231–252.

Ballaban-Gil, K., Rapin, I., Tuchman, R., & Shinnar, S. (1996). Longitudinal examination of the behavioral, language, and social changes in a population of adolescents and young adults with autistic disorder. *Pediatric neurology, 15*(3), 217-223.

Baron-Cohen, S., & Hammer, J. (1997). Parents of children with Asperger syndrome: What is the cognitive phenotype?. *Journal of Cognitive Neuroscience, 9*(4), 548-554.

Baron-Cohen, S., & Jolliffe, T. (1997). Another Advanced Test of Theory of Mind: Evidence from Very High Functioning Adults with Autism or Asperger Syndrome. *Journal Of Child Psychology & Psychiatry & Allied Disciplines, 38*(7), 813-822.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of child psychology and psychiatry, 42*(2), 241-251.

Beaumont, R., & Newcombe, P. (2006). Theory of mind and central coherence in adults with high-functioning autism or Asperger syndrome. *Autism: The International Journal Of Research & Practice, 10*(4), 365-382.

Begeer, S., Koot, H. M., Rieffe, C., Meerum Terwogt, M., & Stegge, H. (2008). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. Developmental Review, 28(3), 342-369.

Best, C. S., Moffat, V. J., Power, M. J., Owens, D. C., & Johnstone, E. C. (2008). The Boundaries of the Cognitive Phenotype of Autism: Theory of Mind, Central Coherence and Ambiguous Figure Perception in Young People with Autistic Traits. *Journal Of Autism & Developmental Disorders, 38*(5), 840-847.

Begeer, S., Koot, H. M., Rieffe, C., Meerum Terwogt, M., & Stegge, H. (2008). Emotional competence in children with autism: Diagnostic criteria and empirical evidence. *Developmental Review, 28*(3), 342-369.

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

Brent, E., Rios, P., Happé, F., & Charman, T. (2004). Performance of children with autism spectrum disorder on advanced theory of mind tasks. *Autism, 8*(3), 283-299.

Charman, T., Campbell, A., & Edwards, L. S. (1998). Theory of Mind performance in children, adolescents, and adults with a mental handicap. *Cognitive development, 13*(3), 307-322.

Cole, P. M., Dennis, T. A., Smith-Simon, K. E., & Cohen, L. H. (2009). Preschoolers' Emotion Regulation Strategy Understanding: Relations with Emotion Socialization and Child Self-regulation. *Social Development, 18*(2), 324-352.

Cohen, D. J., Baron-Cohen, S., & Tager-Flusberg, H. (Eds.). (1994). *Understanding Other Minds: Perspectives from Autism.* Oxford University Press.

Colle, L., Baron-Cohen, S., & Hill, J. (2007). Do Children with Autism have a Theory of Mind? A Non-verbal Test of Autism vs. Specific Language Impairment. *Journal Of Autism & Developmental Disorders, 37*(4), 716-723.

Corbett, B., Carmean, V., & Fein, D. (2009). Neuropsychological Functioning. In S. Goldstein, J. A. Naglieri & S. Ozonoff (Eds.), *Assessment of Autism Spectrum Disorders*. New York, NY: The Guilford Press.

Cutting, A. L., & Dunn, J. (2006). Conversations with siblings and with friends: Links between relationship quality and social understanding. *British Journal of Developmental Psychology, 24*(1), 73-87.

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child development, 62*(6), 1352-1366.

Ensor, R., & Hughes, C. (2008). Content or connectedness? Mother–child talk and early social understanding. *Child development, 79*(1), 201-216.

Eurelings-Bontekoe, E., Zwinkels, K., Schaap-Jonker, H., & Edrisi, M. (2011). Formal Characteristics of Thematic Apperception Test Narratives of Adult Patients with an Autism Spectrum Disorder. A Preliminary Study. *Psychology, 2*, 687.

Freeman, L., & Miller, A. (2001). Norm-referenced, Criterion-referenced, and Dynamic Assessment: what exactly is the point?. *Educational Psychology in Practice, 17*(1), 3-16

Frith, U., & Happé, F. (1994). Autism: Beyond 'Theory of Mind.'. *Cognition, 50*(1-3), 115-132.

Gilchrist, A., Green, J., Cox, A., Burton, D., Rutter, M., & Le Couteur, A. (2001). Development and current functioning in adolescents with Asperger syndrome: A comparative study. *Journal of Child Psychology and Psychiatry, 42*(2), 227-240.

Gilotty, L., Kenworthy, L., Sirian, L., Black, D. O., & Wagner, A. E. (2002). Adaptive Skills and Executive Function in Autism Spectrum Disorders. *Child Neuropsychology, 8*(4), 241.

Goldstein, S., & Ozonoff, S. (2009). Historical Perspective and Overview. In S. Goldstein, J. A. Naglieri & S. Ozonoff (Eds.), *Assessment of Autism Spectrum Disorders*. New York, NY: The Guilford Press.

Handwerk, M. L., & Marshall, R. M. (1998). Behavioral and emotional problems of students with learning disabilities, serious emotional disturbance, or both conditions. *Journal of Learning Disabilities, 31*(4), 327–338.

Happé, F. E. (1994). An Advanced Test of Theory of Mind: Understanding by story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal Of Autism & Developmental Disorders, 24*(2), 129-154.

Happé, F. E. (1995). Understanding Minds and Metaphors: Insights from the Study of Figurative Language in Autism. *Metaphor And Symbol, 10*(4), 275-295.

Happé, F. E. (1997). Central coherence and Theory of Mind in autism: Reading homographs in context. *British Journal Of Developmental Psychology, 15*(Pt 1), 1-12.

Happé, F., & Frith, U. (2006). The Weak Coherence Account: Detail-Focused Cognitive Style in Autism Spectrum Disorders. *Journal Of Autism And Developmental Disorders, 36*(1), 5-25.

Henry, J. D., Phillips, L. H., Ruffman, T., & Bailey, P. E. (2013). A meta-analytic review of age differences in Theory of Mind. *Psychology And Aging, 28*(3), 826-839.

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in cognitive sciences, 8*(1), 26-32.

Hobson, R. P. (2005). What puts the jointness into joint attention? In N. Eilan, C. Hoerl, T. McCormack, & J, Roessler (Eds.), *Joint attention: Communication and other minds: Issues in Philosophy and Psychology* (pp. 185-204). New York: Oxford University Press

Holaday, M., Moak, J., & Shipley, M. A. (2001). Rorschach Protocols From Children and Adolescents With Asperger syndrome. *Journal Of Personality Assessment, 76*(3), 482-495.

Howlin, P., (2003). Outcome in High-Functioning Adults with Autism With and Without Early Language Delays: Implications for the Differentiation Between Autism and Asperger Syndrome. *Journal of Autism and Developmental Disorders, 33*(1), 3-13.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. Journal Of Child Psychology & Psychiatry, 45(2), 212-229.

Howlin, P., Mawhood, L., & Rutter, M. (2000). Autism and developmental receptive language disorder—a follow-up comparison in early adult life. II: Social, behavioural, and psychiatric outcomes. *Journal of Child Psychology and Psychiatry, 41*(5), 561-578.

Huerta, M., Bishop, S. L., Duncan, A., Hus, V., & Lord, C. (2012). Application of DSM-5 Criteria for Autism Spectrum Disorder to Three Samples of Children With DSM-IV Diagnoses of Pervasive Developmental Disorders. *American Journal Of Psychiatry, 169*(10), 1056-1064.

Hughes, C., Russell, J., & Robbins, T. W. (1994). Evidence for executive dysfunction in autism. *Neuropsychologia, 32*(4), 477-492.

Jacobson, J. W., & Ackerman, L. J. (1990). Differences in adaptive functioning among people with autism or mental retardation. *Journal of Autism and Developmental Disorders, 20*(2), 205-219.

Joiner, T. C. (1996). The relations of thematic and non-thematic childhood depression measures to defensiveness and gender. *Journal of Abnormal Psychology, 24*(12), 803–813.

Jolliffe, T., & Baron-Cohen, S. (1999). A test of central coherence theory: linguistic processing in high-functioning adults with autism or Asperger syndrome: is local coherence impaired?. *Cognition, 71*(2), 149-185.

Kaland, N., Møller-Nielsen, A., Smith, L., Mortensen, E. L., Callesen, K., & Gottlieb, D. (2005). The strange stories test. *European child & adolescent psychiatry, 14*(2), 73-82.

Kankainen, A., Taskinen, S., & Oja, H. (2004). On Mardia's Tests of Multinormality. *Theory and Applications of Recent Robust Methods*, 153-164.

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child, 2*(3), 217-250.

Kanner, L. (1971). Follow-up study of eleven autistic children originally reported in 1943. *Journal of autism and childhood schizophrenia, 1*(2), 119-145.

Kaufman, A., & Kaufman, N. (2004). *Kaufman Brief Intelligence Test, Second Edition*. Circle Pines, MN: AGS Publishing.

Kobayashi, R., Murata, T., & Yoshinaga, K. (1992). A follow-up study of 201 children with autism in Kyushu and Yamaguchi areas, Japan. *Journal of autism and developmental disorders, 22*(3), 395-411.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: Clinical and interpretive manual.* San Antonio, TX: The Psychological Corporation.

Kurita, H. (1997). A comparative study of Asperger syndrome with high-functioning atypical autism. *Psychiatry and clinical neurosciences, 51*(2), 67-70

Kuusikko, S., Haapsamo, H., Jansson-Verkasalo, E., Hurtig, T., Mattila, M., Ebeling, H., & ... Moilanen, I. (2009). Emotion Recognition in Children and Adolescents with Autism Spectrum Disorders. *Journal Of Autism & Developmental Disorders, 39*(6), 938-945

Leipold, K., Vetter, N. C., Dittrich, M., Lehmann-Waffenschmidt, M., & Kliegel, M. (2013). Individual and developmental differences in the relationship between preferences and theory of mind. *Journal of Neuroscience, Psychology, and Economics, 6*(4), 236.

Lockyer, L., & Rutter, M. (1969). A five-to fifteen-year follow-up study of infantile psychosis III. Psychological aspects. *The British Journal of Psychiatry*, *115*(525), 865-882.

Loveland, K. A., & Kelley, M. L. (1991). Development of adaptive behavior in preschoolers with autism or Down syndrome. *American Journal on Mental Retardation.*

Macintosh, K., & Dissanayake, C. (2006). Social skills and problem behaviours in school aged children with high-functioning autism and Asperger's disorder. *Journal of autism and developmental disorders*, *36*(8), 1065-1076.

Madle, R. A. (2007). Review of the Kaufman Brief Intelligence Test, Second Edition. In K.F. Geisinger (Ed.), *The seventeenth mental measurements yearbook (456-459)*. University of Nebraska-Lincoln: Buros Institute of Mental Measurements.

Mawhood, L., Howlin, P., & Rutter, M. (2000). Autism and Developmental Receptive Language Disorder—a Comparative Follow-up in Early Adult Life. I: Cognitive and Language Outcomes. *Journal of Child Psychology and Psychiatry*, *41*(5), 547-559.

Mayes, S. D., 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.

McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ?. *Psychological Review*, *96*(4), 690–702.

Medway, F. J. (2010). Review of the Roberts-2. In R. A. Spies & J. F. Carlson (Eds.), *The eighteenth mental measurements yearbook (490-492)*. University of Nebraska-Lincoln: Buros Institute of Mental Measurements.

Mesibov, G. B., Schopler, E., Schaffer, B., & Michal, N. (1989). Use of the childhood autism rating scale with autistic adolescents and adults. *Journal of the American Academy of Child & Adolescent Psychiatry, 28*(4), 538-541.

Nader, A. M., Courchesne, V., Dawson, M., & Soulières, I. (2016). *Does WISC-IV underestimate the intelligence of autistic children?.* Journal of autism and developmental disorders, 46(5), 1582-1589.

Nader, A. M., Jelenic, P., & Soulières, I. (2015). Discrepancy between WISC-III and WISC-IV Cognitive Profile in Autism Spectrum: What Does It Reveal about Autistic Cognition?. PloS one, 10(12), e0144645.

Narzisi, A., Muratori, F., Calderoni, S., Fabbro, F., & Urgesi, C. (2013). Neuropsychological profile in high functioning autism spectrum disorders. *Journal of autism and developmental disorders, 43*(8), 1895-1909.

Nuske, H., & Bavin, E. L. (2011). Narrative comprehension in 4-7-year-old children with autism: testing the Weak Central Coherence account. *International Journal Of Language & Communication Disorders, 46*(1), 108-119.

Obiols, J., & Berrios, G. (2009). The historical roots of Theory of Mind: the work of James Mark Baldwin. *History of Psychiatry, 20*(79 Pt 3), 377-392.

Ozonoff, S., & Jensen, J. (1999). Specific executive function profiles in three neurodevelopmental disorders. *Journal Of Autism And Developmental Disorders, 29*(2), 171-177.

Pandey, J., Wilson, L., Verbalis, A., & Fein, D. (2008). Can Autism Resolve? In B. K. Shapiro & P. J. Accardo (Eds.), *Autism Frontiers* (191-205). Baltimore, MD: Paul H. Brookes Publishing Co.

Peterson, C. C., Garnett, M., Kelly, A., & Attwood, T. (2009). Everyday social and conversation applications of theory-of-mind understanding by children with autism-spectrum disorders or typical development. *European child & adolescent psychiatry, 18*(2), 105-115.

Peterson, C. C., & Siegal, M. (1995). Deafness, conversation and theory of mind. *Journal of child Psychology and Psychiatry, 36*(3), 459-474.

Piven, J., Harper, J., Palmer, P. A. T., & Arndt, S. (1996). Course of behavioral change in autism: A retrospective study of high-IQ adolescents and adults. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*(4), 523-529.

Roberts, G. E., & Gruber, C. (2005). Roberts-2. Los Angeles: Western Psychological Services.

Rohrer, L. M., Cicchetti, D., Rogosch, F. A., Toth, S. L., & Maughan, A. (2011). Effects of maternal negativity and of early and recent recurrent depressive disorder on children's false belief understanding. *Developmental psychology, 47*(1), 170.

Rumsey, J. M., Rapoport, J. L., & Sceery, W. R. (1985). Autistic children as adults: Psychiatric, social, and behavioral outcomes. *Journal of the American Academy of Child Psychiatry, 24*(4), 465-473.

Rutter, M., & Lockyer, L. (1967). A five to fifteen year follow-up study of infantile psychosis I. Description of sample. *The British Journal of Psychiatry, 113*(504), 1169-1182.

Scheeren, A. M., De Rosnay, M., Koot, H. M., & Begeer, S. (2013). Rethinking Theory of Mind in high-functioning autism spectrum disorder. *Journal Of Child Psychology & Psychiatry, 54*(6), 628-635.

Schopler, E. (1996). Are autism and Asperger syndrome (AS) different labels or different disabilities?—VZ. *Journal of Autism and Developmental Disorders, 26*(1), 109-110.

Seltzer, M. M., Krauss, M. W., Shattuck, P. T., Orsmond, G., Swe, A., & Lord, C. (2003). The symptoms of autism spectrum disorders in adolescence and adulthood. *Journal of autism and developmental disorders, 33*(6), 565-581.

Seltzer, M., Shattuck, P., Abbeduto, L., & Greenberg, J. S. (2004). Trajectory of development in adolescents and adults with autism. *Mental Retardation & Developmental Disabilities Research Reviews, 10*(4), 234-2.

Shedler, J., Mayman, M., & Manis, M. (1993). The illusion of mental health. *American Psychologist, 48*(4), 1117–1131.

Shen, Y., Han, B., & Braverman, E. (2016). Stability of the elastic net estimator. *Journal of Complexity, 32*(1), 20-39.

Taylor, J. L., & Mailick, M. R. (2014). A Longitudinal Examination of 10-Year Change in Vocational and Educational Activities for Adults With Autism Spectrum Disorders. *Developmental Psychology, 50*(3), 699-708.

Taylor, J. L., & Seltzer, M. M. (2010). Changes in the Autism Behavioral Phenotype During the Transition to Adulthood. *Journal Of Autism & Developmental Disorders, 40*(12), 1431-1446.

Taylor, J. L., & Seltzer, M. M. (2011). Employment and post-secondary educational activities for young adults with autism spectrum disorders during the transition to adulthood. *Journal Of Autism And Developmental Disorders, 41*(5), 566-574.

Teglasi, H. (2010). *Essentials of TAT and other storytelling assessments* (Vol. 64). John Wiley & Sons.

Tomasello, M. (2003). *Constructing a language: A wage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Tonks, J., Williams, W. H., Frampton, I., Yates, P., & Slater, A. (2007). Assessing emotion recognition in 9-15-years olds: preliminary analysis of abilities in reading emotion from faces, voices and eyes. *Brain Injury, 21*(6), 623-629.

Twatchman-Cullen, D. (2008). Symbolic Communication: Common Pathways and Points of Departure. In K.D. Buron & P. Wolfberg, (Eds.), *Learners on the Autism Spectrum.* Shawnee Mission, KS: Autism Asperger Publishing Co.

Van Bergen, P., Salmon, K., Dadds, M. R., & Allen, J. (2009). The effect of mother training in emotion-rich, elaborative reminiscing on children's shared recall and emotion knowledge. *Journal of Cognition and Development, 10,* 162–187.

Velloso, R. D. L., Duarte, C. P., & Schwartzman, J. S. (2013). Evaluation of the Theory of Mind in autism spectrum disorders with the Strange Stories test. *Arquivos de neuro-psiquiatria, 71*(11), 871-876.

Watson, A. C., Nixon, C. L., Wilson, A., & Capage, L. (1999). Social interaction skills and theory of mind in young children. *Developmental psychology, 35*(2), 386.

Wilcox, R. R. (2001). Theil-Sen estimator. *Fundamentals of modern statistical methods. New York: Springer-Verlag*, 207-210.

Worchel, F. F., Rae, W. A., Olson, T. K., & Crowley, S. L. (1992). Selective responsiveness of chronically ill children to assessments of depression. *Journal of Personality Assessment, 59*(3), 605–615.

World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines.

Yirmiya, N., Solomonica-Levi, D., Shulman, C., & Pilowsky, T. (1996). Theory of mind abilities in individuals with autism, Down syndrome, and mental retardation of unknown etiology:

The role of age and intelligence. *Journal of Child Psychology and Psychiatry, 37*(8), 1003-1014.