The Relationship between a Norm Referenced Measure of Theory of Mind and Preschoolers' Social Skills in the Classroom

Sarah Ellen Birch
Graduate Center, City University of New York

Recommended Citation
Birch, Sarah Ellen, "The Relationship between a Norm Referenced Measure of Theory of Mind and Preschoolers' Social Skills in the Classroom" (2015). CUNY Academic Works.
https://academicworks.cuny.edu/gc_etds/863
The Relationship between a Norm Referenced Measure of Theory of Mind and Preschoolers’ Social Skills in the Classroom

Sarah E. Birch

CUNY: Graduate Center

A dissertation submitted to the Graduate Faculty in Educational Psychology in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2015
This manuscript has been read and accepted by the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Georgiana Shick Tryon, Ph. D

_________________  ____________________________
Date                  Chair of Examining Committee

Bruce D. Homer Ph. D.

__________
Date

Acting Executive Officer

Helen L. Johnson, Ph. D.

__________

David Rindskopf, Ph. D.

Supervisory Committee
Abstract

THE RELATIONSHIP BETWEEN A NORM REFERENCED MEASURE OF THEORY OF MIND AND PRESCHOOLERS’ SOCIAL SKILLS IN THE CLASSROOM

By

Sarah E. Birch

Adviser: Georgiana Shick Tryon, Ph. D.

This study sought to examine the effect of Theory of Mind (ToM) performance as measured by a norm-referenced assessment on the social skills of a multicultural group of preschool students. This study extends the current research in this area. Most previous studies examined the relationship among ToM, other social cognitive skills and social skills using participants from predominately Caucasian middle class families (Slaughter et al., 2015). Additionally, earlier research relied primarily on homemade measures of such skills (Cassidy et al., 2003; Disendruck & Ben-Eliyahu, 2006; McAlister & Peterson, 2013; Watson et al., 1999; Walker, 2005).

Participants included 67 children between 39 and 50 months of age enrolled in publicly funded preschool programs. Assessment using the NEPSY-II (Korkman, Kirk & Kemp, 2007) determined participants ToM and AR skills. Social skills were assessed through behavior observation and teacher report.

Study findings indicate that demographic variables play a role in how teachers rate student both pro-social and anti-social behaviors. NEPSY-II reliability coefficients this sample were extremely low and the validity of the measure is discussed (van de Vijver & Leung, 1997). Qualitative analysis of the NEPSY-II ToM scale is presented and possible explanations for low Coefficient alpha are discussed.
Acknowledgements

Many people provided support and assistance throughout this project. I would like to thank Lisa Gilday for allowing me to collect data in two of the schools within her wonderful agency. I would also like to thank principals Janice Goldman and Deborah Thomas, as well as their staff members for making me welcome in each of those programs. It was a great pleasure to work with your talented teachers and remarkable students. My research assistant Michal Barnea spent almost as many hours in the schools as I did and provided invaluable help with data collection.

The faculty and staff of the Graduate Center have provided me with a great deal of support and guidance during the course of both my tenure as a student and with this dissertation specifically. Special thanks to Kay Powell, the former Administrator of the Institutional Review Board, worked diligently to help get this project approved. There are not enough words to fully thank Dr. Georgiana Tryon for her extraordinary patience and unconditional supportive during the course of this project. Dr. Tryon, counseled me when I was discouraged, actively supported me in meeting challenges and served as a much needed expert critical eye throughout the course of this project. It would have been impossible to complete this dissertation without her guidance, support and assistance.

My family and friends have served as role models and cheerleaders. Deborah Jaspen, Elayne Livote and Theresa Bhoopsingh each provided me with support and guidance during the course of the project. Karena Rush and Bruce Mortenson supported me as I returned to school and provided guidance and even the occasional break as I worked on this dissertation. Lastly, my mother Margaret Hertzig has always been my role model. Her intelligence, grace and dedication to her family and profession inspire
me daily. This dissertation would not have been possible without her support, guidance and generous assistance.

**TABLE OF CONTENTS**

Chapter 1: Introduction 1

Social-emotional skills 1

Theory of Mind 2

Development of Theory of Mind skills 3

Relationship between Theory of Mind and Social skills 4

Current Study 6

Chapter 2: Literature Review 7

Description of Theory of Mind 7

Theories of ToM Development 11

Modular Theories 11

Theory Theory 14

Factors Associated with the Development of Theory of Mind 17

Skills associated with Theory of Mind at Different Ages 23

Relationship between Theory of Mind and Social Skills 26

Failures of Theory of Mind 29

Autism 29

Relationship between Theory of Mind and Disruptive Behavior Disorders 30

The Current Study 34

Chapter 3: Methodology 37

Research Participants 37
<table>
<thead>
<tr>
<th>Measures</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>NEPSY-II</td>
<td>41</td>
</tr>
<tr>
<td>Coefficient Alphas</td>
<td>44</td>
</tr>
<tr>
<td>Behavior Assessment System for Children, Second Edition</td>
<td>45</td>
</tr>
<tr>
<td>Fluharty Preschool Speech and Language Screening Test - 2</td>
<td>47</td>
</tr>
<tr>
<td>Behavioral Observation</td>
<td>49</td>
</tr>
<tr>
<td>Research Assistants and Their Training</td>
<td>50</td>
</tr>
<tr>
<td>Procedure</td>
<td>52</td>
</tr>
<tr>
<td>Research Design</td>
<td>54</td>
</tr>
<tr>
<td>Chapter 4: Results</td>
<td>56</td>
</tr>
<tr>
<td>Descriptive Statistics</td>
<td>56</td>
</tr>
<tr>
<td>Communication Skills</td>
<td>56</td>
</tr>
<tr>
<td>Social Perception Skills</td>
<td>57</td>
</tr>
<tr>
<td>Teacher Ratings of Social Skills</td>
<td>58</td>
</tr>
<tr>
<td>Observations of Social Skills Behavior</td>
<td>59</td>
</tr>
<tr>
<td>Bivariate Correlations</td>
<td>60</td>
</tr>
<tr>
<td>Results of Research Question Analyses and Hypotheses Testing</td>
<td>64</td>
</tr>
<tr>
<td>Preliminary Analyses</td>
<td>65</td>
</tr>
<tr>
<td>Hypotheses Testing</td>
<td>68</td>
</tr>
<tr>
<td>Hypothesis 1</td>
<td>68</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>70</td>
</tr>
<tr>
<td>Hypothesis 3</td>
<td>71</td>
</tr>
<tr>
<td>Hypothesis 4</td>
<td>75</td>
</tr>
<tr>
<td>Summary of Findings from Hypotheses Testing</td>
<td>78</td>
</tr>
<tr>
<td>Supplemental Analyses</td>
<td>80</td>
</tr>
</tbody>
</table>
Chapter 5: Discussion 85

Key Findings 85

Theory of Mind Affect Recognition and Social Skills 86
Theory of Mind and Teachers Ratings of Preschooler's Social Emotional Skills 88
Theory of Mind, Language Skills, and Social Behaviors 90
Theory of Mind measurement in the Current Study 91
Implications for Practice 92
Study Limitations 94
Directions for Future Research 95
Conclusions 97

Appendix A: Behavior Observation Data Form 99
Appendix B: Script for Partial Interval Instruction 101
Appendix C: Practice Data Collection Sheet (Butterfly Effects) 103
Appendix D: Practice Data Collection Sheet (McIntyre) 104
Appendix E: Information about the Study 105
Appendix F: Consent Letter 107
Appendix G: Operational Definitions of Target Behaviors 108
References 109
List of Tables

Table 1: Demographic Information for Participants ................................................. 40
Table 2: Communication Skills as Measured by the Fluharty-2 ................................ 56
Table 3: Social Perception Skills as Measured by the NEPSY-II ............................. 57
Table 4: Teacher Ratings of Social Skills as Measured by the BASC-2 ..................... 58
Table 5: Participant Observations of Social Skill Behaviors .................................. 59
Table 6: Bivariate Correlations of Study Variables ................................................. 60
Table 7: Multiple Regression: Summary of Demographic Variables and Participants’ Social Skills ................................................................. 66
Table 8: Multiple Regression: Summary of Hypothesis One Variables and Participants’ Social Skills ................................................................. 67
Table 9: Multiple Regression: Summary of Hypothesis Two Variables and Participants’ Social Skills ................................................................. 69
Table 10: Multiple Regression: Summary of Hypothesis Three Variables and Participants’ Social Skills ................................................................. 71
Table 11: Multiple Regression: Summary of Hypothesis Four Variables and Participants’ Social Skills ................................................................. 72
Table 12: Summary of Findings from Hypotheses .................................................... 77
Table 13: Percentages of Student Correct and Incorrect Answers to ToM Items ......... 80
Table 14: Summary of ToM Scale Questions the Majority of Participants Answered
Incorrectly 81
CHAPTER I

Introduction


Pro-social skills such as cooperation, sharing, and empathy have been found to predict academic success. Such findings have been demonstrated in pre, elementary, and middle school students (Caprara et al., 2000; Konold et al., 2010; Rhoades et al., 2008; Wentzel & Caldwell, 1997). Such positive social-emotional skills have been suggested to have a direct impact on academic achievement (Denham & Brown, 2010). Pro-social behavior has also been suggested to influence variables such as teachers’ perceptions of students that, in turn, have direct effects on student achievement (Konold et al., 2010)

Negative behavior patterns include antisocial behaviors such as aggression, fighting, and poor relationships with adults and peers (Happé & Frith, 1996; Hughes, Dunn, & White, 1998; Kazdin, Siegel, & Bass, 1992; Sutton, Reeve, & Keogh, 2000). Children who exhibit such
disruptive behavior patterns experience difficulty making and keeping friends (Bryant et al., 1999; Frick, 2001; Lochman, Coie, Underwood, & Terry, 1993), getting along with family members (Frick, 2001; Kazdin et al., 1992), and performing successfully in school (Bryant et al., 1999; Frick, 2001; Kazdin et al., 1992; Webster-Stratton et al., 2004).

The impact that social-emotional skill performance has on student achievement illustrates the importance of this topic. Understanding what factors relate to positive social emotional skills and which variables contribute to anti-social behavior may relate to the development of effective tools to promote pro-social skills and address anti-social behaviors in students.

Research indicates that information processing skills, the manner in which an individual encodes, organizes, and interprets stimuli, impacts the way he or she interacts with others (Denham & Brown, 2010). Dodge (1980) found that children who exhibit disruptive behavior problems demonstrate different patterns of cognitive attribution in response to ambiguous stimuli than their typically developing peers. One paradigm that may be useful in conceptualizing the exhibition of both positive and negative behavior patterns is Theory of Mind (ToM).

**Theory of Mind**

ToM refers to individuals’ ability to recognize both that they themselves construct internal representations, known as mental states, which are used to organize and process environmental stimuli, and that others also demonstrate this skill. Individuals with intact ToM skills understand that the mental states of others are not identical to the mental states that they themselves experience. They then use this understanding to make predictions regarding the behavior of others (Astington & Gopnik, 1991).

There are two general theoretical schools that describe the development of ToM skills. The Modular theory describes ToM as a cognitive module dedicated to encoding, interpreting,
and storing specific types of environmental stimuli (Garfield, Peterson, & Perry, 2001).

Simulation theory is a subset of Modular theory which hypothesizes that a person’s ability to understand and act in accordance with the mental states of others is based upon each person’s innate skill in reproducing the emotion, intention, or desire experienced by the person with whom they are interacting within their own mind (Gallese & Goldman, 1998). The Theory of Mind theory posits that an individual’s understanding regarding the mind is an informal self-developed framework similar to a scientific theory (Flavell, 1999, 2004; Gopnik & Meltzoff, 1997). According to this theory, individuals are born with some innate capacities that relate to the development of ToM; however, their ability to understand and interpret mental events develops primarily from environmental interactions (Flavell, 2004).

**The Development of Theory of Mind Skills**

The development of ToM is related to by a number of influences. Humans are born with some innate skills in this area such as emotion expression and the ability to differentiate people from other environmental stimuli (Gopnik, Meltzoff, & Kuhl, 1999; Izard, Huebner, Risser, McGinnes, & Dougherty, 1980). Additionally, humans, in comparison to other species, demonstrate both strong learning skills and an innate drive to learn new skills (Klippert, Emde, Butterfield, & Campos, 1986; Vaish & Striano, 2004). Lastly, children receive informal instruction in such skills from other people in their environment (Gopnik, Meltzoff, & Kuhl, 1999).

Key skills associated with the development of ToM have been found across children from a variety of different cultures (Shahaeian, Peterson, Slaughter & Wellman, 2011). The development of these skills has been shown to proceed according to a specific sequence common to typically developing children across several cultures (Astington, 1993; Callaghan, 2005;
Wellman, Cross, & Watson, 2001). However other research has found that children from different cultures develop such skills in a slightly different order (Viden, 1996). Infants demonstrate distinct emotional states and independent desires (Astington & Gopnik, 1991; Izard et al., 1980). By the age of one year, babies begin to demonstrate joint attention skills (Gopnik, 1999). Two year olds predict how others will behave based upon the expressed desires of those individuals (Astington, 1993). As they reach the age of three, youngsters consistently differentiate between real and pretend (Astington, & Gopnik, 1991). At about the age of four, children begin to demonstrate the ability to answer first order false belief questions correctly (Astington, 1993). As children approach age six, they develop second order belief attributions (Coull, Leekam, & Bennett, 2006). The development of these skills allows the child to determine the motivations of those around her and to act in accordance with those determinations (Astington, 1998; Hollerandse, van Hout, & Hendriks, 2014).

**The Relationship between Theory of Mind and Social Skills**

Knowledge about both how they and others think provides a basis that allows individuals to explain their thoughts and to identify common assumptions as well as misunderstandings with others. These skills allow an individual to determine the motivations of those around them and to act in accordance with those determinations. Such skills are foundation for successful social interactions (Astington, 1998).

Research (Capage & Watson, 2001; Cassidy et al., 2003; Hughes & Ensor, 2006) provides support for the relationship between ToM skills and pro-social behavior in children. Strong ToM skills relate to the social interactions of young children. Cassidy and others (2003) examined the relationship between children’s understanding of belief, their understanding of
emotion, and their social competency. Peer ratings of child likability related to performance in ToM skills where children who were rated as more likeable performed better on these tasks.

The relationship between ToM skills and antisocial behaviors appears to be complex and not well understood. There is evidence linking antisocial behavior to failures in executive functioning (Dodge, 1980, Hughes, Cutting, & Dunn, 2006). Dodge (1980) found that children who exhibit antisocial behaviors are more likely to make hostile attributions regarding ambiguous initiations from others. Such findings might indicate that these children demonstrate poor abilities in understanding the perspectives of others. Others suggest that perhaps individuals who exhibit specific anti-social behaviors such as bullying have not only adequate but advanced ToM skills. They theorize that in order to successfully intimidate others one must be skillful at understanding the desires, emotions, and beliefs expressed by another person and predicting what responses that person will make in reaction to an intentionally provocative statement on his or her part (Sutton, Smith, & Swettenham, 1999).

The Current Study

The purpose of this study is to further examine the relationship between a culturally diverse group of preschool aged children’s understanding of ToM and their levels of pro-social and anti-social behavior in in the classroom. Research indicates that it is during the preschool period, between the ages of 3 and 5, that young children begin to consistently use the skills associated with ToM (Astington, 1993; Slaughter, Bennis, & Pritchard, 2002; Wellman, 1990). Additionally, research in the area of early intervention indicates that both the positive and negative experiences of young children impact their future performance (Herrod, 2007). Examining the differences in preschoolers’ social-cognitive skills will provide additional insight
into one factor that is thought to play a large role in the social interactions children have with the others in their environment (Hughes, Dunn, & White, 1998).

While previous research in this area has relied on “homemade” measures of ToM skills, my dissertation will utilize a norm-referenced measure, the NEPSY-2, to examine ToM ability. This will allow me to compare performance within my sample and to provide comparisons to the general population. Notably, most studies to date have examined the relationship between social cognitive skills such as ToM and the preschoolers’ social skills in relatively uniform samples of students (Slaughter, Imuta, Peterson & Henry, 2015). This study seeks to expand the research in this area by focusing on a multicultural group of preschoolers attending publicly supported educational programs.

Sixty seven (67) preschool students between the ages of 3 years and 5 years of age were recruited to participate in this study. A correlational research design was used to examine the relationship between preschoolers’ social skills and their ToM skills. Descriptive data including gender and age of participants were obtained. Multiple regression analyses were used to analyze the data collected. Supplementary analyses using partial correlations and ANOVAs were conducted as necessary.

Specifically, this study examines if preschool children who receive higher scores in ToM and affect recognition skills, as measured by the NEPSY-II, will exhibit higher rates of pro-social behavior and lower rates of antisocial behavior than preschool students who receive lower scores on ToM and affect recognition skills. Additionally, I hypothesize that preschool children who receive lower scores in ToM and affect recognition skills, as measured by the NEPSY-II, will exhibit lower rates of pro-social behavior and higher rates of antisocial behavior than preschool students who receive higher scores on ToM and affect recognition skills.
CHAPTER II

Literature Review

This chapter examines the research concerning ToM. The predominate models of ToM functioning are examined, a developmental trajectory of ToM skills is presented, and findings regarding failures of ToM are examined. Additionally, research on the impact of ToM skills on social-emotional behavior is discussed. The chapter concludes with a discussion of the rationale and research questions for this study.

Theory of Mind

Description of Theory of Mind (ToM). Human behavior can be characterized in two distinct ways. First, people can be thought of as objects that interact with other objects mechanically, much in the way atoms and molecules act upon one another. Second, humans can be thought of as psychological entities whose actions come from internal mental states (Wellman & Gellman, 1992).

Jean Piaget (Carpendale, Lews, Müller, & Racine, 2005; Flavell, 2000; Lillard, 1998) conducted the initial work on how people begin to understand their own thoughts and the thoughts of others as well as the world around them. Piaget’s research in this domain examined several facets of the growth of children’s’ mental understanding. Specific research questions within this domain focused on understanding cause and effect relationships and discriminating between external (environmental) and internal (mental activities). He proposed that children follow a specific sequence in developing such skills (Lillard, 1998; Wellman & Estes, 1986; Wellman & Gellman, 1992).

According to Piaget (1928 reprinted in 2009), young children go through a series of developmental changes in their understanding of the social world. Initially, young children
experience the environment around them and, through their experience, form perceptions regarding environmental stimuli. However, they do not understand that their perceptions in turn shape their experience with the environment. As they progress through a series of developmental stages, youngsters begin to understand that their thoughts as well as the ways in which they interact with the world around them influence their understanding of the environment. Two of the steps within this progression set the stage for future of socio-cognitive research (Flavell, 2004). Young children demonstrate what Piaget refers to as a Magico-Phenomenist concept of the mind (Piaget, 1928 reprinted in 2009). Children within this stage of social-cognitive development are Egocentric; they are unable to differentiate their individual thoughts and actions from the world that surrounds them. By approximately the age of three years, Piaget states that children enter the stage of Moral Explanation. It is within this stage of social-cognitive development that children begin to discriminate their thoughts from those of other individuals and from the physical characteristics of their environment. This ability to distinguish personal thoughts from stimuli is a cornerstone of ToM (Flavell, 2004).

Premack and Woodruff (1978) first described the term ToM within the context of their intelligence research in chimpanzees. The researchers showed a chimpanzee a video tape of a human actor who was faced with a problematic situation familiar to the animal. They then presented the animal with photographs representing several different possible actions the actor could take to resolve the problem. The chimpanzee was directed to choose the photograph representing which action the actor would take. The authors proposed that if a chimpanzee was able to accurately predict what the actor would do to resolve the situation the animal would be able to surmise the actor’s goal or intention thus exhibiting a ToM (Premack & Woodruff, 1978). A similar concept was proposed by Brothers (1990). Using the term social cognition, Brothers
(1990) state that the combined recognition of an individual’s affective state, as well as awareness of their physical movement leads to the recognition of their goals.

This research led philosophers such as Jonathan Bennett, Daniel Dennett, and Gilbert Harman to propose that the demonstration of this ability would suggest that an individual was endowed with the ability to think, suppose, or believe (Flavell, 2004). This ability allows an individual to comprise an independent conception of self and the world around her (Flavell, 2004). An individual’s ideas and thoughts are separate and distinct from the physical world around her; however, these thoughts are what lead the individual to act within the environment. Belief allows the person to construct an independent conception of self and the world around her (Flavell, 2004; Wellman & Gellman, 1992).

An individual’s understanding of the way in which both she and others act is described as naïve or folk psychology (Astonington, 1993). Such representations are comprised of two distinct parts, ontological and causal. The ontological component allows the person to distinguish between the internal world of thoughts, emotions, beliefs, and other mental processes and the external world of physical objects. The causal aspect allows the individual to explain the relationship between her own mental world and the physical world in which she lives (Wellman & Gelman, 1992). The informal understanding humans have regarding their mental states and the world around them plays a large role in describing the ways in which individuals interact with others. ToM skills describe this relationship (Garfield, Peterson, & Perry, 2001; Wellman & Gelman, 1992).

ToM refers to individuals’ ability to recognize first that they themselves experience mental states and second that other individuals also demonstrate this ability. Individuals with intact ToM skills understand that the mental states of others are not identical to the mental states
that they themselves experience. One can then use these skills to make predictions regarding the behavior of others (Astington & Gopnik, 1991). Interactions between people require the application of these skills (Astington, 1998).

Each person experiences a varied repertoire of mental states that represent their interpretations of environmental stimuli (Astington, 1993). People experience emotions, such as happiness when seeing a loved one after a long absence or sadness when a pet dies. People hold beliefs about themselves, others, and the world around them. They may believe that they are smart or that their coworker is lazy or that hard work leads to success. Additionally, individuals have desires, a person may want to see a baseball game or crave a platter of cheese fries. Last, one also has intentions, thoughts regarding actions that will lead to the attainment of desires. An example of this might be planning to go to the local diner in order to buy the cheese fries that the individual wishes (Astington, 1993).

Mental states are representations that assist in organizing the environment. Such representations provide the filter through which the individual interacts with the world around him (Astington, 1993; Astington & Gopnik, 1991). In order to recognize that one experiences mental states, an individual must first recognize his direct relationship with the physical world. The individual must then recognize that he may interpret events within the physical world differently on distinct occasions. These interpretations represent the individual’s psychological relationship to the environment (Astington, 1993, 1998).

Adults without psychopathology are able to make each of the distinctions described above. They recognize the difference between the actual physical world and their interpretations of the world around them. They recognize that other people experience independent relationships with the physical world and form their own representations of this experience.
Most importantly, adults recognize that their individual interpretations are not necessarily the same as other people. This recognition relate to their understanding of the social world and their interactions with others (Astington, 1998).

**Theories of ToM Development**

According to current research two theories dominate the field of ToM (Astington, 1993; Wellman, 1990). Modular theories hypnotize that ToM skills are fast acting groups of cognitive structures that quickly and efficiently process social cognitive information (Garfield, Peterson, & Perry, 2001). Theory theories describe ToM skills as a naïve or folk psychology (Astington, 1993; Gopnik & Meltzoff, 1997). ToM skills according to this school represent the individual’s framework for interpreting social events (Gopnik & Meltzoff, 1997).

**Modular theories.** Modular theories posit that ToM competency is a cognitive module of the type proposed by Foder (1983). Cognitive modules are hypnotized to be effective information processing systems that allow individuals to make sense of specific types of environmental stimuli rapidly and effectively (Garfield, Peterson, & Perry, 2001). These information-processing structures are thought to be comprised of domain specific neural networks that attend to, encode, store, and retrieve specific types of environmental information. These networks are believed to be associated with specific neuroanatomical regions. Such networks have a mandatory function, meaning that much of the information processing that occurs within the module is automatic and not under the voluntary control of the individual (Scholl & Leslie, 1999).

Theorists have used several different types of modules to describe ToM capacities. Diachronic modules that have been depicted as information processing systems containing an innate set of parameters, which outline the general abilities supported by the module, are one
proposed model (Scholl & Leslie, 1999). This type of module is analogous to the Language Acquisition Device (LAD) described by Chomsky (1965) that contains the parameters of Universal grammar, the basis of many contemporary theories of language development (Gopnik & Meltzoff, 1997; Scholl & Leslie, 1999). Allan Leslie has proposed that ToM capacities are part of a non-paramatized module, providing the essential character of ToM that he calls the Theory of Mind Mechanism/Selection Processing Model (ToMM/SP) (Baron-Cohen, 1998; Scholl & Leslie, 1999). Baron-Cohen (1998) proposes that ToM skills are comprised of several more basic level cognitive modules that work together. A common feature of each of the modular theories is that, while modules contain an innate basis of ToM ability, these functions are not static but rather change and develop based on environmental interactions (Scholl & Leslie, 1999).

Simulation theory, a subtype of Modular theory, has received significant attention (Gallese & Goldman, 1998; Gordon, 1986; Newen & Schlict, 2009; Saxe, 2009). Simulation theory purports that ToM skills represent a person’s efforts to duplicate the mental states of another (Gallese & Goldman, 1998; Gordon, 1986). According to this theory, individuals hold independent mental states including: emotions, beliefs, and desires (Newen & Schlict, 2009). When they interact with another person they use this information to recreate a model of the emotional state of the individual with whom they are relating (Gallese & Goldman, 1998; Newen & Schlict, 2009). This replicated mental state is then used by the individual to make an imaginary decision about what she, herself, would do in a specific situation (Newen & Schlict, 2009). This replication of the other person’s mental state is then used to make a judgment regarding what the other person will do or how he will act (Newen & Schlict, 2009). The
individual then acts in accordance with the judgment that was made (Gallese & Goldman, 1998; Newen & Schlicht, 2009; Saxe, 2009).

Research in the area of ToM provides some support for modular theories. Baron-Cohen and his colleagues (1998) describe four aspects of ToM skills that provide such evidence. Individuals with disorders such as autism demonstrate specific deficits in ToM skills, while their skills in other areas are not impacted. People with disorders such as Williams or Down’s syndrome demonstrate deficits in other areas of cognition but intact ToM skills. Additionally, people perform ToM tasks quickly and often without effort or specific attention. Finally, children reared in societies with different cultural norms develop ToM along similar trajectories. Liu et al. (2008) conducted a meta-analysis examining the development of False Belief understanding, a hallmark of ToM Development, in children from Hong Kong, Mainland China, the United States, and Canada. The researchers found that despite differences cultural belief systems, children demonstrate similar patterns of performance on similar measures of False Belief understanding during development. It, however, is interesting to note that cultural differences were seen to significantly impact when such skills developed (Liu et al., 2008).

Several studies have demonstrated a biological basis for the modular theory of ToM skills. These include neuroimaging studies that indicate activation in specific regions of the brain during performance of ToM tasks. Such regions include the frontal, parietal, and temporal cortices. Additionally, research with individuals who have sustained damage to specific brain regions, such as the orbito-frontal cortex, demonstrates significantly weaker performance on ToM tasks (Ang & Pridmore, 2009; Stone, Baron-Cohen, & Knight, 1998). Lastly, the discovery of Mirror neurons provides some evidence for Simulation Theory (Gallese & Goldman, 1998). Mirror neurons are a group of visualmotor neurons that have been found in the premotor cortex.
of several species of animals including humans (Saxe, 2009). This class of neurons has been found to fire both when the animal performs a specific motor action and when it watches another individual perform these actions (Gallese & Goldman, 1998; Saxe, 2009). As Gallese and Goldman discuss (1998), mirror neurons generate identical neural activity in an individual watching a behavior as that experienced by the individual participating in the behavior. Other researchers state that while mirror neurons may in fact be a part of a modular system of ToM, Simulation Theory cannot account for the way in which such a system functions (Saxe, 2009). Researchers who support this view cite the limited evidence that motor neurons replicate the mental states of target individuals in addition to their motor activities (Saxe, 2009).

**Theory theory.** This school posits that an individual’s understanding regarding the mind is an informal self-developed framework analogous to a scientific theory (Flavell, 1999, 2004; Gopnik & Meltzoff, 1997). While individuals are born with innate capacities that relate to the development of ToM, this paradigm for understanding and interpreting mental events develops primarily from environmental interactions (Flavell, 2004).

Researchers within this school argue that, like a scientific theory, this personal foundational theory must demonstrate specific structural, functional, and dynamic properties (Flavell, 1999; Gopnik & Meltzoff, 1997). Structurally, a theory makes an ontological claim, a statement of being. Theories must be general and abstract, they must provide general conceptual ways of understanding specific occurrences rather than just describing specific exemplars. In this way a theory establishes coherence, a structure that supports the components from which it is comprised. A theory describes domain specific if-then relationships between stimuli. In addition, theories must address counterfactual claims, statements regarding what may occur if the
underlying events explained within the theory differed in some way (Flavell, 1999; Gopnik & Meltzoff, 1997).

Functionally, theories are explanatory; they make sense of events, data, and other types of stimuli. Theories allow individuals to make predictions regarding domain specific information. They allow those who use them to make interpretations of stimuli and provide a method for determining which stimuli are relevant within the framework and which are not and can thus be discarded. Lastly, theories are dynamic, they can change, such as when one is exposed to evidence that is contradictory (Gopnik & Meltzoff, 1997).

Theory theorists propose that the understandings of the individual’s mind and mental process meet these requirements. Naïve psychologies such as ToM provide an underlying ontological statement of the being of mental state that is abstract and general. It provides explanatory and predictive power regarding mentalizing and the way the mind works. Lastly, it changes and develops based upon exposure to counter evidence (Gopnik & Meltzoff, 1997).

Bartsch and colleague’s (1996) proposed developmental model of ToM provides an example. Children in this developmental model develop ToM skills in a manner that John Flavell (1999) equates to Jean Piaget’s Equilibration. Children have a theory regarding their mental world. They encounter new evidence through interaction with the environment that is not interpretable under their assumptions and throws them into a state of disequilibrium whereby they must modify their understanding based upon their experiences. It is from this modification that children form a new and richer theory. In the Bartsch et al. model, children at about age 2 demonstrate a desire psychology whereby they understand that people have mental connections to items and events, such as wanting a piece of cake, but do not yet understand that individuals form mental representations or thoughts about such stimuli. As they gain more environmental
experience and encounter examples that do not fit within the theory they have developed, they make modifications and begin to demonstrate what these researchers term a desire-belief psychology. At this time, children begin to understand that individuals form and hold beliefs about stimuli, such as believing that mom keeps chocolate bars in the kitchen cabinet. However, they still act based upon their desires; so that when asked where another child will look for the chocolate bar that they (but not the other child) watched their sister hide in the living room under the couch they will respond by stating “under the couch”. The final state in this model is the formation of a belief-desire by which children understand that both beliefs (the knowledge that their peer did not witness the deception) and desires (wanting the chocolate bar) lead to actions (Wellman, Phillip, & Rodriquez, 2000).

Wellman and his colleagues (2011) describe the development of ToM skills as comprised of five general skill areas a person possesses. Individuals demonstrate understanding of diverse desires (DD); knowledge that different people may have different degrees of want for the same item. People also demonstrate knowledge of diverse beliefs (DB); awareness that people may believe differently about the same item. They also demonstrate knowledge-access (KA); the realization that while something may be true another person may not have such information. People with fully developed ToM skills also demonstrate knowledge about false belief (FB). Lastly, such individuals understand that others may demonstrate hidden emotions (HE); for example feeling one way but demonstrating a facial expression or behavior inconsistent with that emotion.

Factors Associated with the Development of Theory of Mind

Three factors relate to the development of ToM skills. People are born with some innate skills in this area. Additionally, humans, in comparison to other species, demonstrate both strong
learning skills and an innate drive to learn new skills. Lastly, children receive informal
instruction in such skills from other people in their environment (Garfield, Peterson & Perry,
2001; Gopnik, Meltzoff, & Kuhl, 1999). A closer look at each of these factors as well as an
outline of the sequence in which ToM skills develop follows here.

A large body of research describes the skills that humans appear to demonstrate from
birth. Infants demonstrate emotional expressions such as joy, fear, sadness, and anger (Izard,
et al., 1980). Work done by Robert Fantz and others (as cited in Singelman & Rider, 2003)
demonstrates both that newborns differentiate between human faces and other objects; and that
they prefer to look at these faces compared to other presented stimuli. Infants also identify and
prefer the sound of human voices from other presented sounds (Gopnik, Meltzoff, & Kuhl,
1999). Research conducted by Meltzoff and Moore (as cited in Gopnik et al., 1999)
demonstrates that newborns can imitate facial expressions. Corielli (2005), refers to this as
emotional contagion and describes it as a necessary component of affect recognition and the
voluntary expression of emotion. These findings, taken together, provide evidence that humans
are born with an inherent understanding that people are a special category of environmental
stimuli, one that is both different and preferable to any other stimulus that may be encountered.
Additionally, newborns appear to demonstrate a basic understanding that they are similar to other
humans from birth (Gopnik et al., 1999).

Very young children also demonstrate learning skills. Meltzoff (1988) studied the
imitation abilities of 60 typically developing nine-month-old infants. He demonstrated target
actions such as pushing over a hinged rectangle, pushing a button to produce a beeping sound,
and shaking a plastic egg filled with metal pieces. Meltzoff modeled the behavior associated
with each item in three modeling periods consisting of three action demonstrations for the
infants. He found that, when compared to a control group, infants who were shown how to manipulate the various items were significantly more likely to exhibit the target behaviors. Additionally, he found that these children exhibited similar rates of targeted actions when presented with the same objects 24 hours later. Meltzoff (as cited in Gopnik et al., 1999) also found that somewhat older babies (14 months) could learn and retain how to use an object is a novel manner. He modeled pushing a panel attached to a box with his forehead to cause a light bulb within the box to light up. The children were only allowed to watch as the model exhibited the target behavior during the first experimental session. One week later, the infants were invited back to the laboratory and presented with the box used during the first session. These children exhibited the target behavior of touching their forehead to the panel of the box to cause the bulb to light.

Another example of the learning skills of young children is illustrated by the phenomena of social referencing. Social referencing refers to an individual’s ability to turn towards another person and utilize the emotions expressed by that person to adjust their own behavior (Klinnert, Emde, Buttrtterfield, & Campos, 1986; Vaish & Striano, 2004). Mary Klinnert and colleagues (1986) found that one year olds used information obtained from both their mothers and other friendly adults to respond to a novel and ambiguous situation. The authors studied the behavior of 46 typically developing 12 and 13 month old children, who were randomly assigned to one of two experimental groups. In the first group, the small child interacted with a friendly adult while the child’s mother sat nearby. The adult was instructed to respond by smiling when a remote control robot unexpected entered the playroom and was made to approach the adult and the baby. The mother was directed to display a neutral expression. In the second group, all conditions were identical except the friendly adult responded by frowning when the toy robot entered the room.
The authors found that these young children looked both towards their mothers and the less familiar friendly experimenter to provide clues as to how to respond to the entrance of the robot. Children in the smile condition were significantly more likely to approach the robot and smile as it was directed about the room than were children in the frown condition who were more likely to respond to the robot by crying and backing away from the toy. Additionally, children who saw the experimenter exhibit a fearful expression were significantly more likely to move towards their mother and reach for her than were children in the smile condition.

Vaish and Striano (2004) found that 12 month olds adjusted their behaviors when presented with ambiguous stimuli not only in response to their mothers’ facial expressions but also in response to vocal cues. The infants were randomly assigned to one of three experimental conditions. Infants were placed on one side of a visual cliff apparatus while their mothers stood at the other end in each condition. In the facial only situation, mothers smiled at their babies from across the apparatus. In the vocal only condition, mothers maintained a neutral expression and used an encouraging tone and words to urge their child across the visual cliff. In the combined condition, mothers used both their facial expressions and their vocalizations to urge their young children to cross the apparatus toward them. The authors found that all the infants tested responded to their mothers’ cues by crossing the visual cliff. However, infants who received both facial and vocal cues from their mothers crossed the cliff in the least amount of time. Those who were in the vocal condition crossed slightly more slowly than those children in the combined condition. Infants in the facial only condition took the longest time to cross the apparatus. This study suggests that perhaps infants may gain more social information from the verbal output of the other people in their world than through other means.
These findings provide some information about the learning abilities of young children. Such research supports the hypothesis that infants not only learn specific behavioral action modeled by others within their environment but also learn the meaning of different emotional expressions displayed by others and then use that information to modify their own behavior.

Vygotsky’s socio-cultural perspective of development describes that learning is transmitted to individuals through both formal mechanisms, such as schooling, and informal mechanisms, such as interactions with various family members (Schunk, 2004). Young children develop ToM skills, in part, because they receive informal instruction in such skills from others in their environments (Gopnik, Meltzoff, & Kuhl, 1999). One finding that lends support to this hypothesis is that children who are born into families with older siblings develop ToM skills at younger ages than either only children or those with younger siblings (Gopnik et al., 1999). Additionally, a study conducted by Ruffman, Perner, and Parkin (1999) found that the way in which mothers disciplined their children influenced when these youngsters developed ToM skills. Mothers who reported that they typically asked their child to consider the feelings of an injured party as part of their disciplinary practice had children who developed ToM skills slightly but significantly earlier than children whose mothers did not incorporate such practice within their reprimands.

Another example of the influence of the environment on the development of ToM skills is illustrated in the findings of several cross cultural studies on the development of ToM skills. Viden (1996) conducted research with 34 youngsters between the ages of 4 and 8 from the Juín Quechua culture in Peru. Viden examined 2 skills associated with ToM performance: appearance-reality and false belief. She modified tasks used in previous research to be culturally congruent for participants. For example in the appearance-reality she used a pen shaped like a
carrot (a vegetable familiar to the participants). This study found that, in contrast to previous studies of ToM development conducted with youngsters from literate Western cultures, the children from the Juín Quechua developed appearance-reality skills prior to false belief skills (1996). Wellman, Fang, Liu, Zhu, and Liu (2006), worked with 92 preschool students between that ages of 3 and 5 from Beijing, China. The authors found that while these students developed the same 5 skills associated with ToM performance as same aged children in the United States, they developed these skills in a different sequence. Wellman and colleagues reported that in contrast to English speaking preschoolers from the United States and Australia, Chinese preschoolers developed KA skills prior to demonstrating DB skills. Shahaeian et al., (2014) found similar differences when they compared the development of ToM skills in Iranian and American youngsters. The authors of these studies theorize that cultural differences such as the societal norm of collective responsibility in China and Iran and the individualistic norm seen in the United States lead to the transmission of different information regarding mental states to children which may lead to the variations in the sequence seen in the development of ToM skills between cultures (Wellman et al., 2006; Shahaeian et al., 2014).

Lastly, there is a clear relationship between competence in language skills and the development of ToM skills in children (Fernàndez, 2011). Astington and Jenkins (1999) examined both the language and the ToM skills of 59 three-year-old children over a seven month period. The authors found that changes in children’s language skills predicted changes in their development of ToM skills. Other researchers (e.g. Ruffman, Slade, Rowlandson, Rumsey & Garnham, 2003) have identified similar general trends. It is not yet clear which aspects of language support the development of skills associated with ToM (Fernàndez, 2011). Astington and Jenkins (1999) work found that children’s ability to bring together words in a grammatically
correct manner (syntax) predicted their abilities to preform specific ToM skills. Other researchers have found that variations in the understanding of language semantics or the meanings of words, predicts performance on ToM tasks (Ruffman et al., 2003). Ruffman and colleagues (2003) conducted a longitudinal study of 73 children with an average age of 3.07 years over a two and a half year period. Specifically, their research found that children’s semantic language skill performance predicted variability’s in their understanding of belief based tasks.

It is expected that children with communication disorders will perform poorly on tests of ToM skills. However, the research results on the development and presentation of ToM skills in children with communication impairments are varied. While early studies found that children with these difficulties performed as well as typically developing children on first order false belief tasks (Colle et al., 2007), other research demonstrated that children with communication impairments demonstrated delays in answering false belief questions compared to matched typically developing peers (Colle et al., 2007). Miller (2004) found that children with communication impairments performed as well as their typically developing peers on false belief tasks that were low in verbal complexity; however, they performed significantly less well than their same age peers on false belief tasks that were verbally complex.

Multiple language backgrounds have been found to relate to ToM development in young children (Farhadian et al, 2010). Goetz (2003) examined the impact bilingualism had on the development of ToM abilities in preschool children. She examined the performance of 104 middle class children on four tasks associated with ToM performance including: two false belief activities, one appearance reality measure and one perspective taking assessment. Children came from Mandarin only, English only or dual Mandarin/English language backgrounds. Her results
indicated that bilingual participants performed significantly better than monolingual students on the skills assessed. Additionally, in a study of 163 Iranian kindergarten students Farhadian et al. (2010) found that children from bilingual (Kurdish/Persian) backgrounds performed better on ToM assessments than those from monolingual (Persian only) backgrounds.

**Skills associated with Theory of Mind at different ages.** Research has suggested that the key skills associated with the development of theory of mind are seen in typically developing children (Astington, 1993; Callaghan et al., 2005; Wellman, Cross, & Watson, 2001). Infants under a year old demonstrate distinct emotion states that others can identify (Izard et al., 1980). They also demonstrate desires exhibited by reaching for preferred items (Astington & Gopnik, 1991). Additionally, these young children appear to recognize to some degree that others also exhibit mental states. Their use of social referencing supports this finding (Gopnik et al., 1999; Klinnert et al., 1986; Vaish & Striano, 2004).

By the age of one year, babies begin to demonstrate joint attention skills. They recognize that they can see, interact with, and feel about environmental stimuli in the same way as others (Gopnik et al., 1999). However, it appears that they do not yet understand that others may think differently about items than they themselves do. Repacholi and Gopnik (1997) demonstrated that this skill develops at approximately 18 months. They worked with two groups of young children. The mean age of the first group was 14.4 months; the mean age of the second group 18.3 months. The children were offered samples of two foods, either raw broccoli or goldfish crackers. Children’s preferences were recorded. The majority of children (93%) preferred the crackers. The members of each group were then randomly assigned to one of two conditions. In the control group, the first author modeled a preference for the goldfish crackers and dislike for the raw broccoli. In the experimental group, the experimenter modeled enjoying...
the raw vegetable and disgust at the taste of the goldfish. She then asked the young child to give her some food by placing her hand palm up midway between the two bowls of food items. The authors found that no matter which food the experimenter modeled enjoying, the 14-month-old children responded to her request by handing her the food that they themselves preferred. However, 18 month olds responded to the experimenter’s request by handing her the food that she had modeled enjoying no matter which food they themselves enjoyed.

By age two children can predict the actions of others based on their expressed desires (Astington, 1993). Additionally, at this age children begin to exhibit empathy. They independently offer comfort to others who exhibit distress (Gopnik et al., 1999).

At three years of age children can reliably differentiate between real and pretend representations (Astington & Gopnik, 1991). They also make distinctions between mental and physical entities. For example, when a three year old is told that one girl has an apple and another girl is thinking about an apple, the three year old can describe which apple can be eaten, cut into pieces, or given to another person (Astington, 1998; Astington & Gopnik, 1991).

Around their fourth birthday, children experience a qualitative leap in their abilities within ToM skills (Wellman et al., 2001). It is at this time that children demonstrate the ability to answer first order false belief questions correctly (Astington, 1993). False belief understanding is commonly considered the classical test of an individual’s ToM skills. Understanding that a person holds a conviction that is untrue or not supported by the physical evidence presented demonstrates that the child is able to distinguish between the mental states of people and the physical world (Wellman et al., 2001). Two examples of false belief tasks are often used in the research. In the first example, the child is presented with a familiar item such as a crayon box. The researcher asks the child what he thinks is inside the box and the child
generally replies, “Crayons”. The researcher opens the box and shows the child that, surprise, this box contains buttons instead of crayons. The researcher then closes the box and asks the child what another person shown the box would say was inside. By age 4, children can correctly answer that a person who has never been shown what is inside the box will believe that the box contains crayons. Younger children consistently state that the box contains the item actually inside (Astington, 1993; Astington & Gopnik, 1991; Wellman et al., 2001). The other classic first order false belief task involves presenting a scenario to the child where one individual processes an item and places that item in a specific location. That person leaves and another person enters, finds the item, and moves it to a different location. The child is asked where the first person would look for the item when she returns (Wellman et al., 2001). Children between ages 4 and 5 consistently answer questions based on similar scenarios correctly (Wellman et al., 2001).

As children approach age six, they develop second order belief attributions (Coull et al., 2006). Such abilities involve the child being able to understand what two people are thinking during a sequential exchange. More simply put, he can answer questions requiring him to state what person A thinks that person B thinks. Coull and his colleagues (2006) point out that often the language used by researchers in assessing this skill is complex and that younger children may be able to answer questions successfully when they are presented with stories in simpler terms.

**The Relationship between Theory of Mind and Social Skills**

By approximately age 6, a child has developed an impressive body of knowledge about both how she and others think (Coull et al., 2006). This understanding provides the basis that allows a person to explain her thoughts and to identify common assumptions as well as misunderstandings with peers. These skills allow the individual determine the motivations of
those around her and to act in accordance with those determinations. In short, these skills provide the foundation for successful social interactions (Astington, 1998).

Watson and her colleagues (1999) at the University of West Virginia set out to determine if for young school children between the ages of 3 and 6 years, differences in their ability to solve classic first order ToM tasks related to differences in how their teachers rated their social skills in the classroom. They rated the performance of 26 preschool and kindergarten children on their level of talkativeness based upon observation, their performance on a standardized language measure (Test for the Auditory Comprehension of Language-revised/TACL-R), and their performance on a first order false belief ToM task (showing children a familiar band aid box containing rubber bands). They found that there was a significant correlation between the children’s performance on the false belief task and their social skills ability as measured by teacher rating ($r = .35$). Additionally, they found that the combination of talkativeness, performance on the TACL-R, and performance on the false belief task accounted for 48% of the variance in teachers’ ratings of children’s social skills. Performance on the false belief task improved the predictive value of the regression equation 10% over and above talkativeness and performance on the TACL-R alone.

Cassidy and others (2003) examined the relationship between children’s understanding of belief, their understanding of emotion, and their social competency. Understanding of belief was measured by a standard first order false belief task and deception task. Understanding of emotion was measured by an affective perspective taking task where children were introduced to a puppet who experienced an emotionally provocative scenario (going to get shots, or getting a new toy). Children were asked questions regarding the feelings expressed by the puppet during each story. Their scores were then combined into one ToM skills score. Social Competency was
measured by peer ratings of likability, teacher ratings of social skills on the Social Skills Rating System (Gresham & Elliot, 1990), and by observers’ recordings of pro-social behaviors in the classroom setting. The authors controlled for the effect of language skills, which were assessed using the Test of Early Language Development-2 (Hresko, Reid, & Hammill, 1991). The relationship between children’s age, language ability, and ToM skills was complex. Together these factors accounted for the majority of variance seen in children’s social competency. While language skills accounted for most of the variance in global assessment of social skills, peer ratings of child likability were found to be related to performance in ToM skills with children who were rated as more likeable performing better on these tasks.

Walker (2005) looked at the relationship between ToM skills, children’s age, gender, and social skills. She collected data on the behaviors of two groups of children. Group one consisted of 53 three to four year olds (25 male, 28 female). Group two consisted of 58 four to five year olds (23 male, 35 female). Children were assessed using two standard first order false belief tasks, and social skills were assessed using the Profile of Peer Relations (Walker, Berthelsen & Irving, 2000). As predicted, older children scored significantly better on ToM skills than younger children. However, her data produced a particularly interesting finding. Walker (2005) found that boys who performed well on ToM skills were ranked as more disruptive and aggressive, while boys who performed less well on these tasks were described as shy and withdrawn by their teacher. For girls, the relationship between ToM and social skills was more straightforward. There was a positive correlation between ToM and pro-social behavior.

In a study conducted with middle class Israeli kindergarten students, Diesendruck and Ben-Eliyahu (2006) examined the role ToM skills plays in both peer and teacher ratings of social skills and peer relationships. The authors found that peers rated participants who demonstrated
stronger performance on an assessment which asked them to determine the motives of another as more popular. Participants who demonstrated better performance in false-belief understanding tasks received more positive behavioral evaluations from peers. Additionally, teachers rated participants who earned higher scores in both positive emotional understanding and false-belief understanding as engaging in greater levels of pro-social behavior. In contrast students with poorer performance on these tasks were rated by teachers as exhibiting higher levels of hostile-aggressive behaviors.

A (2009) study by Haruto Takagishi and his colleagues examined ToM performance and fairness behavior in Japanese preschool aged children using a forced choice method called the Ultimatum Game. The Ultimatum Game is a two-person activity in which one child is provided with a donation of desired items, such as the candies used in this study. The child is then directed by the experimenter to offer their partner a share of the commodity. The partner, who is referred to as the responder, then decides to accept or reject his peer’s offer. If the partner accepts the offer both children split the candies according to the agreed upon bargain, if the partner does not agree to the presented offer then neither child was allowed any of the candies. The authors explained the rules of the Ultimatum Game and presented an example of the game to participants before inviting them to participate in the activity. Takagishi et al. found that children who passed ToM assessments presented fairer (more equal) offers to their partners than the students who performed poorly on such assessments. Interestingly ToM performance did not have a significant relationship regarding whether game partners accepted or rejected the offers of candy.

Cowell, Samek, List and Decety (2014) examined the relationship between ToM and sharing in an ethnically and socio-economically diverse group of urban preschoolers.
Participants’ ToM skills were measured using a false-belief task in which the children were shown a puppet who hides and object and then leaves the scene, another puppet is then presented. The second puppet was shown to change the location of the object’s hiding place. The second puppet was then removed and the first puppet was returned to the scene. Children were asked by the evaluator where the first puppet would look for the hidden object. Children who stated that the puppet would look for the object in the original location passed the ToM assessment. The participants then took part in a game in which they were allowed to share some, all or none of a limited amount of preferred items (six stickers). The authors found that participants who passed the ToM assessment shared significantly fewer preferred items with same sex peers than those who failed the ToM assessment.

**Failures of Theory of Mind**

**Autism.** Individuals with autism comprise a notable and well researched population who exhibit deficits in ToM skills (Colle, Baron-Cohen, & Hill, 2007). Autistic Disorder (AD) is a developmental disability typified by impairments in communication and social interaction, the severity of which is based upon social communication deficits and degrees of rigid and stereotypic patterns of behavior (APA, 2013). While the intensity of symptoms and the initial course of the development may vary between individuals with AD, all experience the grouping of impairments described above (APA, 2000).

AD affects individuals from every ethnic and social class group (Maurice, Green, & Luce, 1996). The prevalence rate for AD has changed dramatically since Leo Kanner first identified the disorder in 1943 (Fombonne, 2003). The Centers for Disease Control estimate that 1 out of every 68 children in the United States has some form of autism spectrum disorder (Baio, 2014).
Children with AD have demonstrated comparatively poorer ability to solve several different versions of first order false belief tasks correctly, which require a person to describe a character’s incorrect or false beliefs about a single circumstance (Colle et al., 2006; Heavey, Phillips, Baron-Cohen, & Rutter, 2000). Individuals with AD also demonstrate significantly more difficulty in solving second order attribution tasks correctly, which require the individual to surmise one character in a story’s thoughts about another character within the story (Heavey et al., 2000) While many higher functioning individuals with AD are able to pass staged false belief tasks such as those described above, they too exhibit deficits in ToM skills. Heavey et al. (2000) demonstrated that these high functioning individuals were unable to correctly answer questions about the emotions and intentions of a character presented in a film depicting an awkward situation.

The Relationship between Theory of Mind and Disruptive Behavior Disorders

The relationship between ToM skills and antisocial behaviors appears to be complex and not well understood. There is evidence linking antisocial behavior to failures in executive functioning (Dodge, 1980; Hughes, Cutting, & Dunn, 2006). Dodge (1980) found that children who exhibit antisocial behaviors are more likely to make hostile attributions regarding ambiguous initiations from others. Such findings might indicate that these children demonstrate poor abilities in understanding the perspectives of others. Others (Sutton, Smith, & Swettenham, 1999) suggest that perhaps individuals who exhibit specific antisocial behaviors such as bullying have not only adequate, but advanced, ToM skills. They theorize that in order to intimidate others successfully, one must be very good indeed at understanding the desires, emotions, and beliefs expressed by another person and predicting what responses that person will make in reaction to an intentionally provocative statement on your part. Only a small number of
empirical studies have sought to address the role of ToM skills in antisocial behavior. A review follows.

Happé and Frith (1996) explored the relationship between ToM skills in children between the ages of 6 and 12 who were diagnosed with conduct disorder. They assessed the language skills, ToM performance, and teacher ratings of social skills for each child using the Vineland Adaptive Behavior Scales (Sparrow, Balla, & Cicchetti, 1998). No significant difference was found in performance on first order ToM skills between children diagnosed with conduct disorder and their typical peers. However, a significant difference was found between the two groups in day-to-day task that required ToM skills measured by the VABS. Children with conduct disorder were found to perform less well on tasks such as responding to hints or cues during conversations, initiating conversations of interest to others, and apologizing for hurting others feelings. The authors theorize that while children with conduct disorder may perform as well as their typically developing peers on first order false belief tasks, more advanced ToM skill assessments, such as those examining second order false beliefs, might demonstrate a difference in skill performance. Additionally, they hypothesize that perhaps individuals with conduct disorder have intact but distorted ToM’s, as they put it a “Theory of Nasty Minds”. This hypothesis was supported by the high ratings children with conduct disorder received compared to their typically developing peers in such antisocial behaviors such as lying and teasing. The authors theorize that in order to deceive or intentionally hurt another person successfully, an individual must have a strong awareness of the mental states experienced by that person; in addition they must be able to predict successfully how that person will act in response to some initiation made by another. Such skills are fundamental to the development of a theory of mind.
Jon Sutton and his colleagues (2000) extended the study by Happé and Firth (1996) to examine the relationship between performance of adolescents on more difficult second order ToM skills, those requiring the individual to infer what a one character thinks another character believes, their levels of disruptive behavior, and their ability to assume responsibility for such behavior. Their findings support those of Happé and Firth (1996), because they found that even on the more difficult second order false belief tasks, there was no significant relationship between ToM and level of disruptive behavior exhibited by these adolescents. Performance on ToM tasks was related to the level of responsibility that individuals took for their behavior. Individuals who performed better on the ToM assessment used in this study were also more likely to have external attributions for anti-social behavior. This finding adds some evidence to the idea that individuals with disruptive behavior disorders have intact but skewed ToM.

Claire Hughes, Judy Dunn, and Alice White (1998) examined the executive functioning of 40 preschool children who exhibited impulsive, aggressive, and disruptive behaviors. The authors found that level of executive functioning as measured both by ToM skills and tasks that assess attention, goal selection, and response to feedback did not significantly explain the variance in social skill functioning between these preschoolers and typically developing peers group.

Hughes and her colleagues (2000) examined and attempted to explain the variations in antisocial behavior exhibited by disruptive preschool children who were rated by their parents as both inattentive and extremely active. They found a weak but non-significant relationship between ToM skills and antisocial behaviors such as taking toys and hurting peers. They also observed a significant relationship between impulse control, as measured by performance on executive function tasks, and antisocial behaviors.
A 2001 longitudinal study conducted by Hughes and her group examined the changes in the relationship between ToM skills in children who exhibited disruptive behavior problems over a three year period. Children were initially assessed at age 4 using a comprehensive battery of standard first order false belief tasks. Researchers then observed the children at both ages 5 and 7 while playing a game, which was rigged so that they would lose, with a peer. While performance on ToM tasks was moderately and significantly related to antisocial behavior during a rigged competitive game at age 5, ToM skills were not significantly related to such behavior at age 7. The authors explain these results as supporting findings by Marakovitz and Campbell (as cited in Hughes et al., 2001) describing that the overt behaviors exhibited by individuals with externalizing behavior problems are more stable than factors associated with cognitive performance such as inattention and impulsivity.

A 2006 study conducted by Hughes and Ensor found that poor ToM skills, assessed using naturalistic methods, such as hiding a coin and exhibiting pretend play skills with representational objects, predicts behavioral problems in toddlers. Additionally, they found that strong ToM skill development served as a protective factor against harsh parenting styles (i.e., use of physical reprimands) for these young children.

Capage and Watson (2001) examined the relationship between social-cognitive skills assessed by ToM tasks, social competence, and aggression in preschool aged children. The authors found that performance on standard ToM assessments correlated negatively with teacher ratings of children’s levels of aggression. This relationship held when receptive language skills were controlled for. Additionally, when presented with stories describing a conflict between two individuals, preschoolers with strong performances on ToM tasks produced fewer aggressive resolutions to these stories and more solutions that involved collaboration.
Lastly, Shakoor, Jaffee and Bowes (2011) conducted a longitudinal study with participants of the Environmental Risk (E-Risk) Longitudinal Twin Study, a representative sample of 2,232 British children and their families. They administered a sequence of standardized ToM tasks developed by Hughes and colleagues to the children at age 5. At age 12 the researchers used parent, teacher and self-reports to identify adolescents who were involved in bullying either as victims, bullies or combination bully-victims. The researchers found that adolescents, who demonstrated poor ToM skills, demonstrated a significantly greater likelihood of being involved in some form of bullying at age 12. The researchers reported that victims (d=0.26), bullies (d=0.25) and bully-victims (d=0.44) demonstrated significantly poorer ToM skills than adolescents not involved in bullying at age 12. The authors report that the impact of ToM skills contributes to involvement in bullying over and above other factors including: low intelligence, child maltreatment and gender.

Current Study

As described, the relationship between ToM skills and social-emotional skills in preschoolers appears to be complex and not well understood. Research (Capage & Watson, 2001; Cassidy et al., 2003; Hughes & Ensor, 2006) suggests that there is a relationship between ToM skills and pro-social behavior in children. Strong ToM skills seem to generally relate to the social interactions of young children. However, as Hughes et al. (2001) describe, this relationship may not hold as children grow older. Others suggest that perhaps individuals who exhibit specific anti-social behaviors such as bullying have not only adequate but advanced ToM skills (Happé, & Firth, 1996; Sutton et al., 1999). They hypothesize that to successfully accomplish some types of antisocial behavior (i.e., bullying, intimidation) individual must understand and manipulate the mental states of their potential victims.
The purpose of this study is to add to this literature by further examining the relationship between preschool aged children’s understanding of ToM skills and their levels of pro-social and anti-social behavior in a free play setting. One strength of this study is the use of a norm-referenced measure to examine social perception skills. Previous research in this field has relied solely on “homemade” measures of ToM and affect recognition, which allowed researchers to compare performance within their sample but failed to provide comparisons to the general population. Another area in which this study seeks to add to the current literature is that this project seeks to examine the relationship between ToM skills and social skills in a group of participants who come from culturally diverse backgrounds. The preschoolers recruited for this project come from diverse ethnic backgrounds and neighborhoods with an average median income below New York City as a whole (U. S. Bureau of the Census, 2010).

This dissertation examines the relationship between social perception skills; emotional and mental state understanding, as measured by the NEPSY-II; and the types of social skills exhibited by preschool children. Specifically I examine the following hypotheses:

HO1: Preschool children who receive higher scores in ToM and affect recognition skills, as measured by the NEPSY-II, will exhibit higher rates of pro-social behavior and lower rates of antisocial behavior as measured by behavioral observation than preschool students who receive lower scores on ToM and affect recognition skills.

HO2: Preschool children who receive higher scores in ToM and affect recognition skills, as measured by the NEPSY-II, will exhibit higher rates of pro-social behavior and lower rates of antisocial behavior as measured by teacher ratings on the BASC-2 than preschool students who receive lower scores on ToM and affect recognition skills.
HO3: Expressive language performance as measured by the Fulharty-2 will serve as a mediator for the relationship between ToM and affect recognition skills, as measured by NEPSY-II. Students with stronger expressive language skills as measured by the Fulharty-2 will exhibit stronger ToM skills as measured by the NEPSY-II as well as stronger social skills as measured by both the behavioral observation and the BASC-2.

HO4: Receptive language performance as measured by the Fulharty-2 will serve as a mediator for the relationship between ToM and affect recognition skills, as measured by NEPSY-II. Students with stronger expressive language skills as measured by the Fulharty-2 will exhibit stronger ToM skills as measured by the NEPSY-II as well as stronger social skills as measured by both the behavioral observation and the BASC-2.
CHAPTER III

Methodology

This chapter describes the methodology of my study, the purpose of which was to examine the relationship between preschoolers’ ToM skills and their behavior in the classroom setting. The study examined levels of pro-social and antisocial behaviors exhibited by preschoolers, teachers’ ratings of students’ social skills, and how the occurrence of these behaviors and teacher ratings of students’ social skills varied based upon their ToM skill performance as measured by the NEPSY-II. A description of the selection and recruitment of participants is presented first. I then discuss the settings in which data were collected. The procedures that were followed during data collection are then described. This is followed by a brief description of the measures that are used in this study. I then describe the training of the research assistants (RAs) who assisted with data collection. An explanation of the study design and the statistical procedures, which were used to analyze the data collected, concludes this chapter.

Research Participants

Sixty-seven (67) typically developing preschool students from multicultural backgrounds between the ages of 39 and 60 months ($M = 51.19, SD = 5.72$) participated in this study. Students from culturally and linguistically diverse backgrounds were recruited for this study since the majority of previous research in this area has used samples comprised of primarily Caucasian
individuals (Slaughter et al., 2015). An a priori power analysis indicated that 43 participants were needed to have 80% power for detecting a large sized effect employing the $p < .05$ criterion of statistical significance. The inclusion of 67 participants within the study ensured that my study had adequate power to address concerns that may arise regarding Type-2 error (Cohen, 1992). Participants were recruited between March 2013 and September 2014 from an agency providing daycare and preschool programs in two boroughs of the New York metropolitan area.

The agency from which participants were recruited is a not-for-profit human services organization that runs both special and general education early childhood programs in the New York Metropolitan area. Program sites in the two neighborhoods contained between two and four universal prekindergarten classes each for typically developing youngsters. According to the New York City Department of Planning, these programs serve 169 and 190 students respectively. The student population of these classrooms comes from the surrounding community. Data retrieved from the New York City Department of Planning provide the following demographic information for each neighborhood. Residents of one community district fall within the following ethnic categories: 13.6% Caucasian, 8.4% African American, 2.1% Asian, and 74.1% Hispanic. The Department of Planning reports that as of 2014, 43.6% of the population of this district received some form of income support such as cash assistance, supplemental security income, or Medicaid. The demographic profile for residents of the second neighborhood served by the agency is: 4.2% Caucasian, 33.2% African American, 3.7% Asian and 55.2% Hispanic. In 2014, 22.6% of community residents received some form of income support (New York City Department of Planning, 2014). Students who participated in this study were recruited from preschool programs which serve children and families from low income communities. Students were enrolled in school based Head Start and Early Learn programs.
within each school. School principals confirmed that children who participated in this project came from homes where the reported income was not greater than 120% of the poverty level (J. Goldman, personal communication, June 2015; D. Thomas, personal communication, June 2015).

Eligibility requirements for study participation included: a) chronological age between 36 and 60 months b) absence of severe visual, hearing, or motor impairments, and c) no record of related service support either within a school, clinic, or home setting.

Participant recruitment occurred in two forms. Initially, parents of potential participants received a letter from the principal investigator describing the study and detailing the participants’ rights (see Appendix A). Additionally, I arranged with the school administrators from each of the preschool programs to set up times such as student drop-off and pick-up times when I met informally with parents and caregivers to introduce myself, explain the purpose of the study, and provide them with written study information and consent materials.

Initially 73 potential participants whose parents provided permission were screened for eligibility criteria. Based upon this screening, 4 children did not meet criteria (2 did not meet age requirements, and 2 received related service support within the school setting) and 69 were invited to participate in the study. One parent withdrew consent and one child left the school prior to the start of data collection, leaving a final sample of 67 students.

Of the 67 students who participated in the study, 58 provided data from all study measures. Six participants were missing 8% of study data (such as a standard score from one of the measures given), two participants were missing 25% of study data (such the standard score from three measures or scores from the behavioral observation), and one student was missing 33% (such as the standard scores from 4 measures) of the data analyzed in this study. Table 1 presents participant demographic information.
Table 1

Demographic Information for Participants

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>31</td>
<td>46.3</td>
</tr>
<tr>
<td>Female</td>
<td>36</td>
<td>53.7</td>
</tr>
<tr>
<td>School</td>
<td></td>
<td></td>
</tr>
<tr>
<td>One</td>
<td>42</td>
<td>57.5</td>
</tr>
<tr>
<td>Two</td>
<td>25</td>
<td>34.2</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>26</td>
<td>38.8</td>
</tr>
<tr>
<td>African American</td>
<td>17</td>
<td>25.4</td>
</tr>
<tr>
<td>Caribbean Black</td>
<td>16</td>
<td>23.9</td>
</tr>
<tr>
<td>Mixed</td>
<td>4</td>
<td>6.0</td>
</tr>
<tr>
<td>African</td>
<td>3</td>
<td>4.5</td>
</tr>
<tr>
<td>East Asian</td>
<td>1</td>
<td>1.5</td>
</tr>
<tr>
<td>Second Language in Home</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>13</td>
<td>68.7</td>
</tr>
<tr>
<td>Spanish</td>
<td>1</td>
<td>19.4</td>
</tr>
<tr>
<td>French</td>
<td>4</td>
<td>1.5</td>
</tr>
<tr>
<td>Creole</td>
<td>2</td>
<td>6.0</td>
</tr>
<tr>
<td>African Language</td>
<td>1</td>
<td>3.0</td>
</tr>
<tr>
<td>Hindi</td>
<td></td>
<td>1.5</td>
</tr>
</tbody>
</table>

Note N=67

Measures (Independent Variables)

NEPSY-II (Korkman, Kirk, & Kemp, 2007). The NEPSY-II assessed participants’ performance on tasks associated with both their ability to understand the intentions, beliefs, and perspectives of others (i.e., ToM subscale) and their skills in recognizing emotions displayed on another’s face (i.e., Affect Recognition subscale). Taken together these skills comprise the Social Perception scale of the NEPSY II, which can be used with individuals as young as 3 years of age (Korkman et al., 2007). The NEPSY II was normed on a national sample of 1,200 children between the ages of 3 and 16 based on the 2003 U.S. census data. The authors report that 100 individuals within each of the age groups participated in the norming process. They state that the normative sample was also matched by the percentage of ethnic group representation according
to the 2003 U. S. census. The authors report adequate validity and reliability evidence for this measure. Construct validity data for the AR and ToM scales of the NEPSY-II imply moderate relationships between the tasks within these measures ($r = .53$ clinical sample, $r = .21$ normative sample). Such findings are to be expected since the tasks within these domains measure a very wide variety of skills (Korkman et al., 2007).

The ToM subscale of the Social Perception scale of the NEPSY-II examines a participant’s abilities in three key components associated with ToM. These include understanding the ways in which others perceive specific events, their intentions towards others, as well as their feelings (Korkman et al., 2007). The NEPSY II produces scaled scores ($M = 10, SD = 3$) for this ToM subtest. The ToM subtest is comprised two separate tasks. The first is a verbal task in which the child is read descriptions of ambiguous situations or shown pictures about such situations and then asked to answer questions pertaining to the points of view presented. This subtest uses a 15 verbal questions based upon a variety of research based activities used to assess ToM skills (Korkman et al., 2007). Example questions include those which are based upon the unexpected-contents false belief task developed by Wimmer and Perner (1983) in which participants are shown two clearly marked and familiar boxes that hold unexpected substances (Joseph & Tager-Flusberg, 2004). Other questions are derived from tasks that assess perception knowledge such as those designed by Pillow (1989) and Pratt and Bryant (1990) in which participants are shown a picture of a character looking at a box and told that the character knows the contents, they are then told that another character approaches the box and are asked if that character is aware of what is inside (Joseph & Tager-Flushberg, 2004). The contextual task requires that, after looking at a depiction of a specific situation, the child is directed to select a photograph from an array describing the affect attributed to a person in the
situation (Korkman et al., 2007). This section of the subtest contains six questions. According to research conducted by Hughes and her colleagues (2000), the aggregate approach adopted by Korkman et al. (2007) has been shown to demonstrate greater reliability in measures associated with the assessment of ToM skills. The authors report that the ToM subtest has adequate reliability, which indicates that the subtest produces a stable measure of the concept of ToM (Korkman et al., 2007). When examined across the norming sample, the ToM scale obtained an alpha reliability coefficient of $\alpha = .76$ for three year olds, $\alpha = .76$ for four year olds and $\alpha = .84$ for five year olds. Responses of the current study participants yielded an alpha reliability coefficient of $\alpha = .46$. Such a coefficient indicates that when used with this sample the ToM subtest does not produce consistent scores and all results associated with this subtest need to be viewed with extreme caution.

The Affect Recognition subscale of the Social Perception scale of the NEPSY-II is comprised of four specific tasks. In the first task, the child is shown two photos of children’s faces and is asked to state if the emotion depicted in each photo is the same. On the next task, the student is asked to identify two photos depicting the same emotion from an array. In the third task, the student is presented with a photo of a student demonstrating a specific emotion and then is asked to identify a picture depicting the same emotion from an array presented. In the last task, the child is asked to remember the emotion depicted on a face shown previously and then to select two photos from an array that match the presented stimulus (Korkman et al., 2007). The authors report that this subtest has adequate reliability, which indicates that the subtest produces a stable measure of the affect recognition construct (Korkman, et. al., 2007). Specifically, when examined across the norming sample an alpha reliability coefficient of $\alpha = .80$ was obtained for three year olds, $\alpha = .68$ was obtained for four year olds, and $\alpha = .70$ was
obtained for five year olds. Based on the responses of participants from the current study an alpha reliability coefficient of $\alpha = 0.60$ was found. Such a coefficient indicates that when used with this sample the Affect Recognition subscale produces less consistent scores and that such scores need to be viewed with caution.

**Coefficient alphas.** Cronbach (1951) describes alpha as a measure of the homogeneity of a scale. While high alpha values indicate homogenous relationships between scale items low values of alpha indicate inconsistency between the items within a scale (Tavakol & Dennick, 2011). Alpha is a function of the administration of a scale within a specific population (Bland & Altman, 1997; Hattie, 1985; Tavakol & Dennick, 201; van de Vijver & Leung, 1997). While Korkman et al., 2007 report strong alpha values for ToM scale of the NEPSY-II, I found much lower levels of consistency when this scale was used with my sample.

Although the NEPSY-II was normed using a sample representative of the general United States population, the participants in the current study were not adequately represented in this process. The NEPSY’s norming sample was comprised of children and adolescents who identified as White (60%), African American (15%), Hispanic (19%) and other (8%). In contrast the participants in this sample came exclusively from Hispanic, African American, Caribbean Black, Mixed (such as African American and Hispanic), African and East Asian backgrounds. Korkman et al. (2007) reported that youngsters were excluded from the norming sample if their primary language was not English, however they do not report any information regarding the inclusion of children from dual or multiple lingual backgrounds. In comparison, while the majority of participants in my study came from English only backgrounds, many (31.3%) came from multilingual home settings. The NEPSY-II provides parental education level as an indication of participant socioeconomic status. Parental education obtainment was described as
below 11 years (13.28%), 12 years (25.61%), between 13-15 years (30.38%) and more than 16 years (30.74%). No such data were obtained for the participants in this study, however, all participants came from homes which earned no more than 120% of the federal poverty level. The composition of the sample used in my study was very different from the demographics reported for the NEPSY-II. Cultural factors including: home language use, family income, and ethnic background may account for differences seen in the alpha levels reported by Korkman et al. (2007) and those found in this study. These findings will be addressed in the Results of this dissertation.

Behavior Assessment System for Children, Second Edition, Teacher Rating Scales, Preschool (BASC-2 TRS-P; Reynolds & Kamphaus, 2004). The BASC-2 is a norm referenced standardized measure designed to assess the behavior and self-perceptions of individuals between the ages of 2 and 25. This measure consists of multiple tasks, which examine a variety of behavioral domains. Specifically, the BASC-2 system consists of a student observations schedule, a structured developmental history interview, as well as parent, teacher, and student rating scales. The components of the BASC-2 system may be used either separately or in unison depending upon the needs of the evaluator (Reynolds & Kamphaus, 2004). The BASC-2 was standardized using a sample of over 13,000 participants between the ages of 2 and 18 years. The standardization sample was stratified according to ethnicity, gender, socio-economic status, and geographic region (Reynolds & Kamphaus, 2004). The Teacher Rating Scale: Preschool (TRS-P) The TRS-P is a paper survey, in which respondents rate statements about a child’s behavior in a school based setting. Respondents indicate if a behavior never, sometimes, often, or almost always occurs. The TRS-P consists of 12 scales which examine a wide variety of school based functioning in young children. I used the Adaptability, Aggression, and Social Skills subscales of
the TRS-P to obtain a teacher rating of each participant’s social-emotional behaviors within the classroom. Specifically the Adaptability scale examines skills such as a child’s ability to shift focus from one task to another as well as to share toys with other children. The Aggression scale describes a child’s inclination to engage in behaviors which are verbally or emotionally harmful to others or property. Specific items examine hitting or verbally attacking others, as well as breaking or destroying materials. Lastly the Social Skills scale looks at the interpersonal relationship skills children possesses. The scale asks questions related to complimenting peers and adults, using manner words such as please and thank you and offering encouragement to others. The TRS-P produces a T score ($M = 50, SD = 10$) for each scale (Reynolds & Kamphaus, 2004).

Reynolds and Kamphaus (2004) report adequate validity and reliability evidence for this measure. Specifically, alpha levels for each scale at the two to three and four to five year old levels are reported. The Adaptability scale has an alpha reliability coefficient of $\alpha = .79$ for two to three year olds and $\alpha = .83$ for four to five year olds. The Aggression scale has alpha reliability coefficients $\alpha = .92$ for two to three year olds and $.93$ for four and five year olds. The authors report that the Social Skills scale has an alpha reliability coefficient of $\alpha = .89$ for two and three year olds and $\alpha = .91$ for four to five year olds (Reynolds & Kamphaus, 2004). The responses for participants in the current study produce adequate alpha reliability coefficient for each scale specifically: $\alpha = .84$ for Adaptability, $\alpha = .78$ for Aggression and $\alpha = .87$ for Social Skills (Tavakol & Dennick, 2011). Validity evidence for the BASC-II was obtained from a variety of sources including the comparison of BASC-II and other teacher report measures. Correlations between the BASC-II: TRS and the Achenbach System of Empirically Based Assessment
ASEBA) ranged between .64 to .85 for all clinical scores indicating strong relationships between the two measures (Reynolds & Kamphaus, 2004).

Fluharty Preschool Speech and Language Screening Test -2 (Fluharty-2). The Fluharty-2 is a screening tool that is used to rapidly identify preschool aged children (2-0 to 6-11) who are at risk for a communication delay (Fluharty, 2001). The Fluharty-2 describes a child’s abilities in the areas of articulation, receptive and expressive language skills. It consists of 5 subtests: Articulation, Repeating Sentences, Responding to Directives, Answering Questions, Describing Actions, and Sequencing Events. The measure produces standard scores ($M = 10, SD = 3$) for each of the five subtests. In addition, a Quotient score ($M = 100, SD = 15$) is calculated for three broader composite scores. For this study I used participants’ composite scores in Expressive Language, Receptive Language, and General Language. The Expressive Language composite score is comprised of a student’s performance on the Describing Actions and Sequencing Events subtests. The score describes a participant’s skills in choosing meaningful verbs and stating them in the correct grammatical tense as well as presenting concepts on a specified topic in a manner that is understandable to another (Fluharty, 2001). The Receptive Language quotient describes a participant’s skills in following one step directives, answering general information questions, and in recalling and repeating simple and complex sentences (Fluharty, 2001). Lastly the General Composite Language score describes a child’s overall language performance, which encompasses his or her awareness of the semantics and syntax of spoken language (Fluharty, 2001). The Fluharty-2 results demonstrate strong correlations between its subtests and the other language assessments. The Fluharty-2 was normed on a stratified sample of children ($N=705$) of three ($n =136$), four ($n =175$), five ($n =225$), and six ($n =169$) years of age representative of the United States population (Fluharty, 2001).
The authors (Fluharty, 2001) report adequate validity and reliability evidence for this measure with coefficient alphas that are generally in the high .80s and low 90s. The responses for participants in the current study produced alpha reliability coefficients for each of the composites used as follows: $\alpha = .70$ for the Expressive Language composite, $\alpha = .80$ for the General Language Composite and $\alpha = .73$ for the Receptive Language composite. These coefficient alpha scores, while lower than those of the norming sample, indicate adequate reliability for this measure for research purposes (Tavakol & Dennick, 2011). Validity evidence for the Fluharty-2 included Content validity analysis, Criterion-Predictive validity examination and evaluation of Construct-Identification validity. The Fulharty-2 is based upon the model of language proposed by Foster (1990) and test format and questions are based upon established measures of communication performance (Fluharty, 2001). Criterion-prediction validity was examined by comparison with the Test of Language Development-Primary: Third Edition (1997). Moderate to high correlations (.55 to .95) were found between the subtest and composite scores on the two measures (Fluharty, 2001). This measure was chosen for its solid psychometric properties and the speed with which it can be administered.

**Behavioral observation.** I, or a trained research assistant (see below for training details), conducted a 20-minute, partial interval observation of each participant’s behavior during a free play period (i.e. centers or recess). The observer was blind to the results of the student’s performance on the NEPSY-II subtests. The targets of the observation were four specific behaviors. I chose these behaviors because an extensive search of the social skills literature indicated that they were representative of pro-social and antisocial behaviors exhibited by children. The two pro-social behaviors include cooperation and reassuring. Cooperation was labeled as both being in close proximity (2 feet or less) to another child while engaged in a task
that is oriented toward either toward that person or participating in a mutual activity and either verbally or physically engaging to further the activity (Guglielmo & Tryon, 2000; Hughes & Ensor, 2010). Reassuring is characterized as responding to another’s verbal or physical form of distress with either a verbal or physical expression of comfort (Cassidy et al., 2003; Hughes & Ensor, 2010).

The two antisocial behaviors included physical aggression and verbal aggression. Physically aggressive actions was operationalized as hurting another child or damaging materials (Brotman et al., 2008; Rescorla, 2005; Werner, Cassidy, & Juliano, 2006; Willoughby et al., 2001). Verbal aggression was categorized as making accusations or angry remarks about another (Werner et. al., 2006; Willoughby et al., 2001).

The 20-minute observation period was divided into 80 equal intervals. Targeted behaviors were coded as present if they were noted for all or part of the interval (Alberto & Troutman, 2009). After each observation, the instances of each targeted behavior (i.e., cooperation, reassuring, physical aggression, verbal aggression) were summed. Additionally, the totals of all pro-social behaviors and all antisocial behaviors were calculated. A copy of this observational data sheet may be found in Appendix A.

**Research Assistants and Their Training**

Three graduate students in psychology or special education participated as research assistants (RAs) on this project. After completing the Collaborative Institutional Training Initiative (CITI) training. After which, RAs were taught how to complete a partial interval behavioral observation of the four target behaviors identified. RA’s were trained both individually and in small groups ($n < 3$).
The following procedure was used to instruct all RAs in conducting a partial interval observation. A script for this training procedure is attached (Appendix B). I provided verbal instruction in how to complete a partial interval observation to the RAs. This instruction reviewed operational behavior definitions of the four target behaviors, time intervals, the summation of observational data, and inter-observer agreement (IOA). The RAs then watched a 5-minute video on partial interval data collection prepared by Butterfly Effects Behavior Instructor Basics (BIB) Certification Course (2012). Butterfly Effects is an organization that provides approved continuing education for the Behavior Analysis Certification Board in a wide variety of skills including: data collection, instructional, and ethics (2012). RA’s next practiced collecting partial interval data using the 70-second observation included in the video (Appendix C). RAs practiced collecting partial interval data using a short video from Thomas McIntyre’s website, BehaviorAdvisor.com (Appendix D). Dr. McIntyre is a professor of Special Education at CUNY: Hunter College (McIntyre, 2012).

RAs then practiced collecting partial interval data on the four behaviors operationalized for this study. They were then instructed in the use of the specific data collection sheet for this study (Appendix B). Instructions for completion of all data collection sheets were also printed on the pages themselves. I then used Public Domain videos of preschool aged children, which were approximately 5 minutes in length. For copies of these videos please visit http://www.behavioradvisor.com/TrainingVideo.html. The purpose of the videos was to provide RAs with practice in collecting behavioral data such as that needed for this study. In order to insure that observations were conducted reliably, inter-observer agreement (IOA) between each RA and me was conducted for each of the practice observations. I compared the observation data I collected using the same videos and same interval length to the RA’s data. I used the Kappa
statistic (Watkins & Pacheco, 2000) to calculate IOA between myself and the RA. Kappa is defined as the proportion of agreement between observers that occurs beyond chance. Watkins and Pacheco describe the Kappa statistics between .60 and .75 as good and those above .75 as excellent. IOA with a Kappa of at least .60 was established between the observations by the RA and me before the RA was allowed to collect observation data to be used in this study.

IOA data was also collected for 13.4% of observations conducted during the study. This served a reliability check to determine the specificity of the behavior operational definitions used in this study and the judgment and objectivity of the observers (Alberto & Troutman, 2009). Mean Kappa for this study was .89 with a standard deviation of .13.

Procedure

I provided the director of the general education preschool with sealed information packets for each of the students enrolled. The information packet contained a description of the dissertation study (Appendix A), contact information for my advisor, Georgiana Tryon, and myself, and a parental consent form with a reply envelope (Appendix E).

The director of the preschool program gave the packets to classroom teachers who sent them home in students’ backpacks. Teachers collected the returned consent forms in the envelopes from students’ backpacks and place them in a collection area in the director’s office. I picked up returned consent forms from the director weekly. I then assigned each participant a unique identification number.

Additionally, I held recruitment sessions at both preschool sites. I set up a table near the central office at each preschool and met with parents both in the morning after they had dropped their children off at school and after they had picked their children up during dismissal. I provided each parent with a description of the proposed study (Appendix E), contact information
for my advisor, Georgiana Tryon, and myself, and a parental consent form (Appendix F). Parents completed the consent either during the recruitment session or they took the information home and returned the signed consent for in a sealed envelope to the director’s office, where I then collected it.

After receiving consent from a participant’s parent or guardian, I met with the child in his/her classroom and, accompanied by the classroom teacher or assistant teacher, escorted the child to a quiet room within the school building to administer the Fluharty-2 and subtests comprising the Social Perception scale of the NEPSY II in a counterbalanced order. After administering the test protocol, I escorted the student back to the classroom. Each test protocol was labeled with the identification number of the participant and placed with the participant’s consent materials.

I followed standard administration procedures for the Fluharty-2 and the NEPSY II. The manual for each measure (Fluharty, 2001; Korkman et al., 2007) provides a detailed description of the administration procedures. I then placed the completed assessment protocol in the numbered data packet assigned to that child.

Either a researcher assistant or I observed each participant during a daily independent play period, such as recess or classroom centers, for 20 minutes. The observer was blind to the child’s score on either of the two subtests of the NEPSY-2. The observer tracked two specific pro-social (cooperation and reassurance,) and two specific antisocial (physical aggression and verbal aggression) behaviors displayed by the participant using the partial interval recording datasheet (see Appendix A). Operational definitions of target behaviors were provided to RA’s (Appendix G). Data from each observation were labeled using the participant’s assigned
identification number. I completed 13.4% of observations with the RAs. IOA was calculated for these observations. This served as an integrity check.

Additionally, each student’s teacher completed the Adaptability, Aggression, and Social Skills subscales of the TRS-P (Reynolds & Kamphaus, 2004) to obtain a teacher rating of each participant’s social-emotional behaviors within the classroom. These rating scales were returned to the principal investigator and labeled with the participants’ assigned identification numbers.

**Research Design**

A correlational research design was used to examine the relationship between preschoolers’ social skills and their ToM skills. Correlational research designs are quantitative studies that allow the investigator to examine the relationship between two or more variables within the same group of subjects. Correlation studies allow researchers to analyze the relationship among a large number of variables and provide information regarding the degree and direction of the relationships between variables.

The Statistical Package for the Social Sciences (SPSS) was used to conduct analyses. Descriptive statistics examined each of the variables included in this study. Additionally, specific analyses were conducted to determine if these data met the assumptions for multiple regression analysis including: linear relationships, normality of data, and homoscedasticity. The observational data collected consisted of count variables, which do not meet the requirements for analysis by the Ordinary Least Squares method of multiple regression (Glass & Hopkins, 1996); so the Poisson method was proposed to analyze the relationship between these outcome variables and the predictor variables. Poisson multiple regressions are a type of general linear model that examines data such as those described (i.e., a rate of behavior exhibited within an observational period). One specific assumption for variables analyzed by such models is that the likelihood of observing an event in each interval is not dependent upon the observation of an event within any other interval (Coxe, West, &
Aiken, 2009). The data collected did not adequately fit Poisson models. I attempted to correct for this poor fit using a Negative Binomial Regression model (Coxe, et al., 2009). While Negative Binomial Regression models did fit the study data, they did not add to the interpretation of the data over and above that seen with an Ordinary Least Squares Regression model. Therefore I used an Ordinary Least Regression model to examine these dependent variables. Supplementary analyses using partial correlations were also conducted.
CHAPTER IV

Results

This study was designed to investigate the extent to which Social Cognitive skills measured by the NEPSY-II are associated with the social skills of a group of typically developing, urban preschoolers from culturally and linguistically diverse backgrounds. The chapter reports on the findings of this project. First, the chapter presents descriptive statistics for participants’ performance on communication, ToM, affect recognition, and social skills behavior. The chapter then describes the relationships between the study’s variables. Lastly, the chapter describes the results of the analyses that answer the study’s research questions and hypotheses.

Descriptive Statistics

Communication skills. Table 2 presents the means, standard deviations, and ranges for each of the communication skills assessed in this study. Specifically, the table presents data regarding the Fluharty-2 Receptive Language Quotient (RLQ), Expressive Language Quotient (ELQ), and overall language skills (General Language Quotient; GLQ). RLQ, ELQ, and GLQ scores were normally disturbed. The skewness and kurtosis measurement for each variable are as follows: RLQ has a skewness of .438 (SE = .295) and kurtosis of -.153 (SE = 582); ELQ has a skewness of -.462 (SE = .295) and kurtosis of .319 (SE = .582); and GLQ skewness is .009 (SE = .295) with a kurtosis of -.470 (SE = .582).

Overall, participants scored within the average range on each of the communication variables. Participants demonstrated lower mean scores in receptive language ($M = 95.74, SD = 7.93$) than overall language skills ($M = 102.30, SD = 6.98$). They demonstrated the highest scores in expressive language skills ($M = 108.27, SD = 7.10$).
**Communication Skills as Measured by the Fluharty-2**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLQ</td>
<td>66</td>
<td>95.74</td>
<td>7.93</td>
<td>80</td>
<td>118</td>
</tr>
<tr>
<td>ELQ</td>
<td>66</td>
<td>108.27</td>
<td>7.10</td>
<td>88</td>
<td>124</td>
</tr>
<tr>
<td>GLQ</td>
<td>66</td>
<td>102.30</td>
<td>6.98</td>
<td>87</td>
<td>118</td>
</tr>
</tbody>
</table>

*Note. RLQ = Receptive Language Quotient; ELQ = Expressive Language Quotient; GLQ = General Language Quotient. Quotient scores range from 40 to 163 (M = 100, SD = 10) with higher scores representative of stronger skills within the domain.*

**Social Perception Skills:** Table 3 presents the means, standard deviations, and ranges of participants’ scores on the two Social Perception subscales of the NEPSY-2: Affect Recognition (AR) and Theory of Mind (ToM). Scores for both variables are normally distributed. AR demonstrates a skewness statistic of -0.471 (SE = .295) and a kurtosis of 0.582 (SE = .599). ToM has a skewness value of -0.098 (SE = .307) and a kurtosis value of 0.611 (SE = .599). Based upon these scores, participants obtained similar levels of performance in each of the two social perception skills measured.

AR scores indicate that, on average, the participants in this study display adequate skills for children of their age in recognizing and differentiating the emotions of others. Additionally, participants in this study demonstrate skills in understanding and interpreting the emotions and perspectives of others (ToM). Taken together, participants’ AR and ToM scores suggest that, on average, the children who took part in this study exhibit Social Perception skills within the range that is expected for children their age.

Table 3

**Social Perception Skills as Measured by the NEPSY-2**

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR</td>
<td>66</td>
<td>11.29</td>
<td>2.19</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>ToM</td>
<td>62</td>
<td>11.18</td>
<td>1.79</td>
<td>6</td>
<td>16</td>
</tr>
</tbody>
</table>

*Note. AR = Affect Recognition Scaled Score; ToM = Theory of Mind Scaled Score*
Teacher ratings of social skills. Table 4 provides information on the means, standard deviations, and ranges for each of the following Behavioral Assessment System for Children (BASC II) scales that measure of Social Skills behavior: Aggression (AGG), Adaptability (ADA), and Social Skills (SS). Participants’ AGG is slightly positively skewed 1.156 (SE = .295); kurtosis is .375 (SE = .582). The ADA and SS variables meet requirements for a normal distribution. ADA demonstrates skewness of -.066 (SE = .295) and kurtosis of -.430 (SE = .582). SS demonstrates skewness of -.619 (SE = .299) and kurtosis of .506 (SE = .590).

Teachers rated participants as slightly lower in aggression (M = 48.29, SD = 7.32) than in adaptive skills (M =52.92, SD=10.28) or social skills (M = 55.80, SD =11.57). Overall, teachers rated participants as performing within normal limits in each of the three areas measured. Taken together, these scores indicate that teachers believe that the students within this study demonstrate pro-social skills such as adapting to new situations and engaging in a variety of social skills with peers and adults. Additionally, the results suggest that the students in this study demonstrate aggressive behaviors at the same rate as their same-age peers.

Table 4

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGG</td>
<td>66</td>
<td>48.29</td>
<td>7.32</td>
<td>41</td>
<td>71</td>
</tr>
<tr>
<td>ADA</td>
<td>66</td>
<td>52.92</td>
<td>10.28</td>
<td>26</td>
<td>75</td>
</tr>
<tr>
<td>SS</td>
<td>64</td>
<td>55.80</td>
<td>11.57</td>
<td>20</td>
<td>75</td>
</tr>
</tbody>
</table>

Note. AGG = Aggression T Score; ADA=Adaptive Skills T Score, SS = Social Skills T Score. AGG, clinical scale; T scores range from 20 to 120 (M = 50, SD =10) with higher scores indicating poorer skills within this domain SS and ADAP T adaptive Scale; T scores range from 10 to 100 (M = 50, SD = 10) with higher scores representative of stronger skills within the domain.

Observations of social skills behavior. Table 5 provides information on the means, standard deviations, and ranges for each of the following observation measures of participant
social skill behavior: Cooperation (COS), Reassuring (RES), Physical Aggression (PAS), and Verbal Aggression (VAS). Skewness and kurtosis were examined to determine if data met the assumptions of normality. Only COS demonstrates skewness and kurtosis values within the limits acceptable for a normal distribution (Glass & Hopkins, 1996). Skewness for COS is -.115 (SE = .295), kurtosis is -.532 (SE = .582). PAS is slightly positively skewed, 1.306 (SE = .295); kurtosis is within the acceptable range for a normal distribution, .745 (SE = .582). VAS is positively skewed, 2.366 (SE = .295) and leptokurtic with a kurtosis value of -5.557 (SE = .582). Finally, RES is positively skewed 7.822 (SE = .295) and highly leptokurtic with a kurtosis value of 62.330 (SE = .582). Due the very few participants who exhibited reassuring behavior (n = 2) and the few total instances of the behavior (n = 7), RES was dropped from the analyses of the research questions.

As Table 5 indicates, participants demonstrated higher levels of cooperative behavior (M = 34.21, SD = 14.59) than any other measured behavior. Students showed similar levels of verbal (M = .58, SD = 1.10) and physical (M = .52, SD = .77) aggression. Participants exhibited lower levels of reassuring behaviors (M = .11, SD = .75) than they did of any of the other behaviors. As a whole, participants exhibited far higher levels of cooperative behavior than any other social behavior measured in this study. They also exhibited far greater variability in their performance of reassuring behavior.
Table 5

Participant Observations of Social Skill Behaviors

<table>
<thead>
<tr>
<th>Variable</th>
<th>( N )</th>
<th>( M )</th>
<th>( SD )</th>
<th>( \text{Min} )</th>
<th>( \text{Max} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>COS</td>
<td>66</td>
<td>34.21</td>
<td>14.59</td>
<td>3</td>
<td>63</td>
</tr>
<tr>
<td>RES</td>
<td>66</td>
<td>.11</td>
<td>.75</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>PAS</td>
<td>66</td>
<td>.52</td>
<td>.77</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>VAS</td>
<td>66</td>
<td>.58</td>
<td>1.10</td>
<td>0</td>
<td>5</td>
</tr>
</tbody>
</table>

*Note. COS = Frequency of Cooperative behavior during the 20-minute observation; RES = Frequency of Reassuring behavior during the 20-minute observation; PAS = Frequency of Physically Aggressive behavior during the 2-minute observation, VAS = Frequency of Verbally Aggressive behavior during the 20-minute observation. Behavior counts ranged from 0 to 63. Higher counts indicate a greater number of observations of the behavior.*

**Bivariate correlations among social perception, communication, and social skills**

**variables.** Table 6 presents the correlation matrix for the variables examined in this study. The relationships between the variables used in this study are more easily examined by discussing the relationships between each category of variables.
### Table 6

**Bivariate Correlations of Study Variables**

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETH</td>
<td>.19</td>
<td>-02</td>
<td>-11</td>
<td>.01</td>
<td>-04</td>
<td>-02</td>
<td>-03</td>
<td>.17</td>
<td>-07</td>
<td>-11</td>
<td>-05</td>
<td>.06</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCH</td>
<td>.48**</td>
<td>---</td>
<td>.25*</td>
<td>-21</td>
<td>.10</td>
<td>-05</td>
<td>.00</td>
<td>.13</td>
<td>.09</td>
<td>.16</td>
<td>.15</td>
<td>.22</td>
<td>.17</td>
<td>-.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE</td>
<td>.19</td>
<td>.25*</td>
<td>---</td>
<td>.00</td>
<td>.11</td>
<td>-.19</td>
<td>.00</td>
<td>.08</td>
<td>.11</td>
<td>-.11</td>
<td>-.06</td>
<td>-.12</td>
<td>-.14</td>
<td>-.04</td>
<td>-.10</td>
<td></td>
</tr>
<tr>
<td>SL</td>
<td>-.02</td>
<td>.21</td>
<td>-.00</td>
<td>---</td>
<td>.08</td>
<td>.13</td>
<td>.11</td>
<td>-.14</td>
<td>-.11</td>
<td>-.14</td>
<td>-.28*</td>
<td>-.19</td>
<td>-.15</td>
<td>-.13</td>
<td>-.78</td>
<td>.00</td>
</tr>
<tr>
<td>GEN</td>
<td>.13</td>
<td>.10</td>
<td>.11</td>
<td>.08</td>
<td>---</td>
<td>-.04</td>
<td>.13</td>
<td>.22</td>
<td>.22</td>
<td>.20</td>
<td>.25*</td>
<td>-.11</td>
<td>-.09</td>
<td>-.03</td>
<td>.00</td>
<td>.25*</td>
</tr>
<tr>
<td>AR</td>
<td>-.11</td>
<td>-.05</td>
<td>-.19</td>
<td>.13</td>
<td>-.04</td>
<td>---</td>
<td>.24</td>
<td>.05</td>
<td>.15</td>
<td>.12</td>
<td>.07</td>
<td>.10</td>
<td>.19</td>
<td>.05</td>
<td>.01</td>
<td>.11</td>
</tr>
<tr>
<td>ToM</td>
<td>.01</td>
<td>-.00</td>
<td>.00</td>
<td>.11</td>
<td>.13</td>
<td>.24</td>
<td>---</td>
<td>.19</td>
<td>.17</td>
<td>.23</td>
<td>.20</td>
<td>.13</td>
<td>.14</td>
<td>-.30*</td>
<td>.21</td>
<td>.10</td>
</tr>
<tr>
<td>RLQ</td>
<td>-.04</td>
<td>.02</td>
<td>.06</td>
<td>-.14</td>
<td>.22</td>
<td>.05</td>
<td>.19</td>
<td>---</td>
<td>.40**</td>
<td>.84*</td>
<td>.17</td>
<td>-.15</td>
<td>-.29*</td>
<td>-.07</td>
<td>.09</td>
<td>-.07</td>
</tr>
<tr>
<td>ELQ</td>
<td>-.02</td>
<td>.13</td>
<td>.08</td>
<td>-.11</td>
<td>.20</td>
<td>.15</td>
<td>.17</td>
<td>.40**</td>
<td>---</td>
<td>.82**</td>
<td>-.05</td>
<td>.06</td>
<td>-.03</td>
<td>.13</td>
<td>.10</td>
<td>.06</td>
</tr>
<tr>
<td>GLQ</td>
<td>-.03</td>
<td>.09</td>
<td>.11</td>
<td>-.14</td>
<td>.25*</td>
<td>.12</td>
<td>.23</td>
<td>.82**</td>
<td>.82**</td>
<td>---</td>
<td>.08</td>
<td>-.04</td>
<td>.18</td>
<td>.02</td>
<td>.09</td>
<td>-.02</td>
</tr>
<tr>
<td>AGG</td>
<td>.17</td>
<td>.10</td>
<td>.11</td>
<td>-.28*</td>
<td>-.11</td>
<td>.07</td>
<td>.20</td>
<td>.17</td>
<td>-.05</td>
<td>.08</td>
<td>---</td>
<td>-.30*</td>
<td>.39**</td>
<td>-.01</td>
<td>.17</td>
<td>.02</td>
</tr>
<tr>
<td>ADA</td>
<td>-.07</td>
<td>.16</td>
<td>-.06</td>
<td>.19</td>
<td>-.09</td>
<td>.10</td>
<td>.13</td>
<td>-.15</td>
<td>.06</td>
<td>-.04</td>
<td>-.30</td>
<td>---</td>
<td>.69**</td>
<td>.06</td>
<td>.15</td>
<td>.00</td>
</tr>
<tr>
<td>SS</td>
<td>-.11</td>
<td>.15</td>
<td>-.12</td>
<td>.15</td>
<td>.06</td>
<td>.19</td>
<td>.14</td>
<td>-.29*</td>
<td>-.03</td>
<td>-.18</td>
<td>-.39**</td>
<td>.69**</td>
<td>---</td>
<td>-.05</td>
<td>.02</td>
<td>.01</td>
</tr>
<tr>
<td>COS</td>
<td>-.05</td>
<td>.22</td>
<td>-.14</td>
<td>-.13</td>
<td>-.03</td>
<td>.05</td>
<td>-.30*</td>
<td>-.07</td>
<td>.13</td>
<td>.02</td>
<td>-.01</td>
<td>.06</td>
<td>-.05</td>
<td>---</td>
<td>.24</td>
<td>.05</td>
</tr>
<tr>
<td>PAS</td>
<td>.06</td>
<td>.17</td>
<td>-.04</td>
<td>-.08</td>
<td>.00</td>
<td>.01</td>
<td>.21</td>
<td>.09</td>
<td>.10</td>
<td>.09</td>
<td>.17</td>
<td>.15</td>
<td>.02</td>
<td>.24</td>
<td>---</td>
<td>.26*</td>
</tr>
<tr>
<td>VAS</td>
<td>.12</td>
<td>-.04</td>
<td>-.10</td>
<td>.00</td>
<td>.25*</td>
<td>.11</td>
<td>.10</td>
<td>-.07</td>
<td>.06</td>
<td>-.02</td>
<td>.02</td>
<td>.00</td>
<td>.00</td>
<td>.05</td>
<td>.26*</td>
<td>---</td>
</tr>
</tbody>
</table>

* = p ≤ .05  ** = p ≤ .01

**Note:** AGE = Participants’ centered age in months; SL= Participant also speaks a language other than English; GEN = Participants’ gender; AR = Affect Recognition Scaled Score; ToM = Theory of Mind Scaled Score; RLQ = Receptive Language Quotient; ELQ = Expressive Language Quotient; GLQ = General Language Quotient; AGG = Aggression $T$ Score; ADA = Adaptation $T$ Score; SS = Social Skills $T$ Score; COS = Frequency of Cooperative Behavior during the 20-minute observation; PAS = Frequency of Physical Aggression during the 20-minute observation; VAS = Frequency of Verbal Aggression during the 20-minute observation.
The first category of correlations that I will discuss present the relationships among the participant demographics and the independent and dependent variables explored in this study. The correlations between the demographic data and students’ performance on the communication variables suggest that GEN is positively correlated with GLQ \( (r = .25, p = .045) \), indicating that girls perform somewhat better on the General Language Scale of the Fluharty-2 than boys. Thus, female participants demonstrate slightly better language skills than male participants. Such a finding is consistent with the literature on early language development (Ozcaliskan & Goldin-Meadow, 2010).

Next, the relationships between demographic data and Social Perception Skills show no significant correlations between the demographic variables examined in this study and the ToM and AR skills earned by participants on the NEPSY-2. Lastly, the relationships between demographic data and participants’ social skills found a negative correlation between SL and AGG \( (r = -.28, p = .025) \), indicating that participants who spoke a second language were rated as slightly, but significantly, less aggressive according to their classroom teachers. Interestingly, there were no significant relationships between participants’ ethnic background and any of the independent and dependent variables examined in this study. Gender is positively correlated with VAS \( (r = .25, p = .05) \). This finding is consistent with research indicating that young girls are more likely to engage in forms of relational aggression, such as verbalizations, than boys (Ostrov & Keating, 2005). Ethnicity was significantly correlated with the school participants attended. This finding is in line with the demographics of the two communities in which the schools used in this study were conducted.

The next group of correlations that I will discuss examines the relationship between student performance on Social Perception variables and the students’ Communication skills as
measured by the Fluharty-2 and their social skills as measured by observations and teacher ratings on the BASC II scales. The communication variables were not significantly correlated with either of the Social Perception skills measured. However, study participants who exhibit stronger performance on ToM tasks demonstrate lower levels of cooperative behavior during the behavioral observations ($r = -0.30, p = 0.019$).

The last category of correlations explores the relationships among the different communication variables as well as the relationships between these variables and social skills. Overall student communication skills, as represented by the GLA score, correlates strongly with students’ receptive language skills, represented by the RLQ score ($r = 0.84, p = 0.001$). Additionally, students’ expressive language skills have a strong, positive relationship with their General language performance on the Fulharty-2 ($r = 0.82, p = 0.001$). The relationship between the expressive and receptive language variables measured by the Fuharty-2 demonstrates a smaller, but significant, effect ($r = 0.403, p = 0.001$), indicating that although the two variables are related, they comprise distinct categories within overall communication performance and their relationships with other study variables should be examined separately. Receptive language skills as measured by the RLQ and students’ social skills as measured by the BASC-2 demonstrate a significant negative relationship ($r = -0.29, p = 0.022$). This finding indicates that students who have lower scores on the RLQ are more likely to engage in specific social skills behaviors as rated by their classroom teachers. These students are somewhat more likely to say “please” and “thank you”, ask for assistance politely, and compliment others..

**Results of Research Question Analyses and Hypotheses Testing**

This section presents the results for the statistical analyses that examine the study hypotheses. Three of the dependent variables (COS, PAS, and VAS) analyzed in this study are
counts, and a Poisson model was selected to examine the relationships between them and the control and predictor variables. However, upon inspection of the analyses conducted using the Poisson model, I found that this model did not fit the data (Coxe, et al., 2009). Specifically, the data collected in this study were overdispersed so that the deviance divided by the degrees of freedom contained within the model were outside the accepted limits of a Poisson Regression Model (Coxe, et. al., 2009). I next attempted to correct for the overdispersion by using a Negative Binomial Regression model (Coxe, et al., 2009). Although this model fit the study data, it did not add substantively to the interpretation of the data over and above that seen with an Ordinary Least Squares Regression model. For the sake of interpretation simplicity, I therefore used an Ordinary Least Regression model to examine these dependent variables. The other dependent variables (AGG, ADA, and SS) are standard scores resulting from teacher ratings of participant behavior. I examined these data using descriptive statistics such as Q-Q Plots before conducting analyses to determine if the data met the assumptions for multiple regression. Overall the data met these assumptions (Glass & Hopkins, 1996).

Preliminary Analyses

Prior to testing the specific study hypotheses, I examined the effects of demographic variables (i.e., speaking a second language, gender, and age) collected on the study’s dependent variables using six regression analyses. Table 7 summarizes the descriptive statistics and analyses results for these models. Overall, the demographic variables taken together do not predict either participants’ social skills measured by observation (COS, VAS, and PAS) or participants’ social skills measured by teacher ratings (AGG, ADA, and SS).

Model 1 presents the multiple regression for COS. This multiple regression model, $R^2 = .04, F(3,62) = .81, p > .05$, suggests that the three demographic variables used do not
significantly predict the frequency with which participants engaged in cooperation during observations. Model 2 presents the regression equation for VAS, $R^2 = .08, F(3, 62) = 1.73, p > .05$. Similar to the findings of Model 1, it is clear that the demographic variables alone do not predict the frequency with which participants engage in verbal aggression during observations.

Model 3 presents regressions for PAS. Similar to the findings seen in the previous models this equation, $R^2 = .01, F(3, 62) = .16, p > .05$ suggests that the demographic variables do not predict the frequency of physical aggression engaged in by participants. However, it is interesting to note that gender provides a unique contribution to the level of verbal aggression exhibited by participants, $b = .57, t(62) = 2.12, p < .05$. Model 4 depicts the relationship between the demographic variables and AGG. Taken together the demographic variables do not predict teacher ratings regarding the level and types of aggression displayed by study participants, $R^2 = .1, F(3, 62) = .2.30, p > .05$. However, it is noted that speaking a language other than English does uniquely predict teacher ratings of participant aggression on the BASC, $b = -4.14, t(62) = -2.23, p < .05$. As seen in Model 4, speaking a language other than English predicts that teachers will rate such students lower in Aggression.

Model 5 presents the analysis for ADA. The results for this model, $R^2 = .05, F(3, 62) = 1.10, p > .05$, indicate that, taken together, the demographic variables do not predict teacher ratings of Adaptability on the BASC. Similarly, Model 6 $R^2 = .04, F(3, 60) = .85, p > .05$ suggests that taken together speaking a second language, gender, and age do not predict the level of social skills, based on teacher ratings, demonstrated by participants.
Table 7

Multiple Regression: Summary of Demographic Variables Participants' Social Skills

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>B</td>
<td>-1.45</td>
<td>B</td>
<td>.01</td>
<td>B</td>
<td>.57*</td>
</tr>
<tr>
<td></td>
<td>SEB</td>
<td>3.64</td>
<td>SEB</td>
<td>20</td>
<td>SEB</td>
<td>1.77</td>
</tr>
<tr>
<td>Age</td>
<td>B</td>
<td>- .36</td>
<td>B</td>
<td>-.01</td>
<td>B</td>
<td>-.02</td>
</tr>
<tr>
<td></td>
<td>SEB</td>
<td>.31</td>
<td>SEB</td>
<td>.02</td>
<td>SEB</td>
<td>.02</td>
</tr>
<tr>
<td>Second Language</td>
<td>B</td>
<td>-4.11</td>
<td>B</td>
<td>-.13</td>
<td>B</td>
<td>-.05</td>
</tr>
<tr>
<td></td>
<td>SEB</td>
<td>3.88</td>
<td>SEB</td>
<td>.21</td>
<td>SEB</td>
<td>.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>.19</td>
<td>.01</td>
<td>.08</td>
<td>.10</td>
<td>.05</td>
<td>.04</td>
</tr>
<tr>
<td>Adj R²</td>
<td>-.01</td>
<td>-.04</td>
<td>.03</td>
<td>.06</td>
<td>.00</td>
<td>-.01</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>64</td>
</tr>
<tr>
<td>SE</td>
<td>14.66</td>
<td>.79</td>
<td>1.08</td>
<td>7.11</td>
<td>10.26</td>
<td>11.61</td>
</tr>
<tr>
<td>F Change</td>
<td>.81</td>
<td>.16</td>
<td>1.73</td>
<td>2.30</td>
<td>1.10</td>
<td>.85</td>
</tr>
<tr>
<td>Sig. F Change</td>
<td>.50</td>
<td>.92</td>
<td>.17</td>
<td>.09</td>
<td>.36</td>
<td>.47</td>
</tr>
</tbody>
</table>

Note: Model 1, Response Variable COS; Model 2, Response Variable PAS; Model 3, Response Variable VAS; Model 4, Response Variable AGG; Model 5, Response Variable ADA; Model 6, Response Variable SS.

* = p ≤ .05
Hypotheses Testing

Hypothesis 1. Hypothesis one stated that the preschool children who receive higher scores in ToM and affect recognition skills, as measured by the NEPSY-2, will exhibit higher rates of pro-social behavior and lower rates of antisocial behavior as measured by behavioral observation than preschool students who receive lower scores on ToM and affect recognition skills. To this this hypothesis three multiple regression models were fitted. Table 8 summarizes the results of these findings.

Table 8

<table>
<thead>
<tr>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>B .40</td>
<td>B -.06</td>
</tr>
<tr>
<td>SEB</td>
<td>3.61</td>
<td>SEB .20</td>
</tr>
<tr>
<td>Age</td>
<td>B -.24</td>
<td>B .00</td>
</tr>
<tr>
<td>SEB</td>
<td>.32</td>
<td>SEB .02</td>
</tr>
<tr>
<td>Second Language</td>
<td>B -.68</td>
<td>B -.08</td>
</tr>
<tr>
<td>SEB:</td>
<td>3.89</td>
<td>SEB .22</td>
</tr>
<tr>
<td>Affect Recognition</td>
<td>B .83</td>
<td>B -.01</td>
</tr>
<tr>
<td>SEB</td>
<td>.85</td>
<td>SEB .05</td>
</tr>
<tr>
<td>Theory of Mind</td>
<td>B -2.6*</td>
<td>B .10</td>
</tr>
<tr>
<td>SEB</td>
<td>1.04</td>
<td>SEB: .06</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.12</td>
<td>.05</td>
</tr>
<tr>
<td>Adj $R^2$</td>
<td>.04</td>
<td>-.04</td>
</tr>
<tr>
<td>$N$</td>
<td>61</td>
<td>61</td>
</tr>
<tr>
<td>$SE$</td>
<td>13.86</td>
<td>.78</td>
</tr>
<tr>
<td>$F$ Change</td>
<td>1.51</td>
<td>.58</td>
</tr>
<tr>
<td>Sig. $F$ Change</td>
<td>.20</td>
<td>.72</td>
</tr>
</tbody>
</table>

Note: Model 1, Response Variable COS; Model 2, Response Variable PAS; Model 3, Response Variable VAS.
* $p \leq .05$

Model 1 presents the multiple regression for COS. This model, $R^2 = .12$, $F(5,55) = 1.51$, $p > .05$, indicates that taken together participant demographic variables, AR scores on the NEPSY-2, and ToM scores on the NEPSY-2 do not predict the level of
cooperative behavior exhibited during observations. However, it is interesting to note that participants’ ToM scores uniquely predict the level of cooperation exhibited by participants, $b = -2.6$, $t(5, 55) = -2.5$, $p < .05$. This finding illustrates that as participants’ scores on the ToM assessment from the NEPSY-2 increased, they were significantly less likely to engage in cooperative behaviors during the study observation.

Model 2 presents the equation for PAS. This model, $R^2 = .08$, $F(5,55) = .9$, $p > .05$, suggests that, taken together, the variables measuring Social Perception (AR and ToM), as well as the demographic variables used in this study do not predict levels of verbal aggression exhibited by participants.

Lastly, Model 3 presents the equation for VAS, $R^2 = .08$, $F(5, 55) = .9$, $p > .05$. Similar to Models 1 and 2, overall this model does not demonstrate a significant relationship between the predictor variables and social skill behaviors exhibited by the preschoolers in this sample. It is interesting to note that gender still maintains a small but unique contribution to the prediction of verbal aggression levels exhibited in this sample, although this relationship is no longer significant at the $p \leq .05$ level, $(b = .54, t(5,55), p = .068)$. Based upon the findings of the models fitted, readers will note, that Hypothesis 1 was not supported.

**Hypothesis 2.** Hypothesis two stated that the preschool children who receive higher scores in ToM and affect recognition skills, as measured by the NEPSY-2, will exhibit higher rates of pro-social behavior and lower rates of antisocial behavior as measured by teacher ratings on the BASC-2 than preschool students who receive lower scores on ToM and affect recognition skills. To this hypothesis, three multiple regression models were fitted. Table 9 summarizes the results of these analyses.
Table 9

*Multiple Regression: Summary of Hypothesis Two Variables on Participants’ Social Skills*

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>-1.62</td>
<td>-2.57</td>
<td>1.72</td>
</tr>
<tr>
<td></td>
<td>SEB 1.87</td>
<td>SEB 2.63</td>
<td>SEB 3.18</td>
</tr>
<tr>
<td>Age</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>.22</td>
<td>-.05</td>
<td>-.19</td>
</tr>
<tr>
<td></td>
<td>SEB .16</td>
<td>SEB .23</td>
<td>SEB .28</td>
</tr>
<tr>
<td>Second Language</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>-4.67*</td>
<td>5.97*</td>
<td>2.9</td>
</tr>
<tr>
<td></td>
<td>SEB 1.99</td>
<td>SEB 2.79</td>
<td>SEB 3.35</td>
</tr>
<tr>
<td>Affect Recognition</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>.28</td>
<td>.25</td>
<td>.76</td>
</tr>
<tr>
<td></td>
<td>SEB .47</td>
<td>SEB .66</td>
<td>SEB .79</td>
</tr>
<tr>
<td>Theory of Mind</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td></td>
<td>.91</td>
<td>.57</td>
<td>.49</td>
</tr>
<tr>
<td></td>
<td>SEB .57</td>
<td>SEB .79</td>
<td>SEB .95</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>.17</td>
<td>.11</td>
<td>.07</td>
</tr>
<tr>
<td>Adj $R^2$</td>
<td>.09</td>
<td>.03</td>
<td>-.02</td>
</tr>
<tr>
<td>$N$</td>
<td>61</td>
<td>61</td>
<td>59</td>
</tr>
<tr>
<td>SE</td>
<td>7.19</td>
<td>10.09</td>
<td>12.01</td>
</tr>
<tr>
<td>$F$ Change</td>
<td>2.19</td>
<td>1.38</td>
<td>.80</td>
</tr>
<tr>
<td>Sig. $F$ Change</td>
<td>.07</td>
<td>.25</td>
<td>.55</td>
</tr>
</tbody>
</table>

*Note: Model 1, Response Variable AGG; Model 2, Response Variable ADA; Model 3, Response Variable SS.*

* $p \leq .05$

Model 1 presents the multiple regression for AGG. This model, $R^2 = .17$, $F(5, 55) = 2.19$, $p = .07$, indicates that taken together participant demographic variables (i.e., speaking a second language, age, and gender), AR scores on the NEPSY-2 and ToM scores on the NEPSY-2 do not predict how their classroom teachers rate the level and types of Aggressive behavior exhibited by study participants. Speaking a language other than English does uniquely predict teachers ratings of aggressive behavior exhibited by participants, $b = -4.6$, $t(5, 55) = -2.35$, $p = .02$. This finding illustrates that the presence of a second language leads to teachers rating students as less aggressive on the BASC-2. This relationship was seen in Table 7 and remains constant even when the predictor variables, AR and ToM, are added to the regression model.
Model 2 presents the multiple regression for ADA. This model $R^2 = .11$, $F (5.55) = 1.38, p = .25$, demonstrates that predictor variables included in the regression equation do not contribute to the prediction of teacher ratings of student adaptability on the BASC-2. However, the presence of a second language spoken by the child does continue to uniquely add to the prediction of teachers’ ratings of participants’ adaptive behavior, ($b = 5.97$, $t (5, 55) = 2.14, p = .04$).

Lastly, Model 3 presents the regression equation for SS. This model, $R^2 = .07$, $F (5, 53) = .80, p = .55$, suggests that, taken together, the demographic and predictor variables do not significantly predict teacher ratings of participate social skills on the BASC-2. The models tested do not support Hypothesis 2.

**Hypothesis 3.** Hypothesis three stated that expressive language performance, as measured by the Fulharty-2, will serve as a mediator for the relationship between ToM and affect recognition skills, as measured by NEPSY-II and participants social emotional skills. Students with stronger expressive language skills, as measured by the Fulharty-2, will exhibit stronger ToM skills, as measured by the NEPSY-II, as well as stronger social skills, as measured by both the behavioral observation and the BASC-2. To test this hypothesis, six multiple regression models were fitted. Table 10 summarizes the results of these findings.
Table 10

**Multiple Regression: Summary of Hypothesis 3 Variables on Participants’ Social Skills**

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>B -92</td>
<td>B -11</td>
<td>B .57</td>
<td>B -.98</td>
<td>B -2.99</td>
<td>B 1.82</td>
</tr>
<tr>
<td>SEB:</td>
<td>3.69</td>
<td>SEB .21</td>
<td>SEB .30</td>
<td>SEB 1.91</td>
<td>SEB 2.73</td>
<td>SEB 3.30</td>
</tr>
<tr>
<td>Age</td>
<td>B -.33</td>
<td>B .001</td>
<td>B -.12</td>
<td>B .27</td>
<td>B -.08</td>
<td>B -.16</td>
</tr>
<tr>
<td>SEB:</td>
<td>.32</td>
<td>SEB .02</td>
<td>SEB .03</td>
<td>SEB .17</td>
<td>SEB .24</td>
<td>SEB .29</td>
</tr>
<tr>
<td>Second Language</td>
<td>B 30</td>
<td>B -.05</td>
<td>B -.05</td>
<td>B -5.20**</td>
<td>B 6.32*</td>
<td>B 2.74</td>
</tr>
<tr>
<td>SEB:</td>
<td>3.93</td>
<td>SEB .22</td>
<td>SEB .32</td>
<td>SEB 2.01</td>
<td>SEB 2.87</td>
<td>SEB 3.48</td>
</tr>
<tr>
<td>Affect Recognition</td>
<td>B .69</td>
<td>B .002</td>
<td>B .04</td>
<td>B -.37</td>
<td>B .21</td>
<td>B .90</td>
</tr>
<tr>
<td>SEB:</td>
<td>.87</td>
<td>SEB .05</td>
<td>SEB .07</td>
<td>SEB .49</td>
<td>SEB .70</td>
<td>SEB .83</td>
</tr>
<tr>
<td>Theory of Mind</td>
<td>B -2.74**</td>
<td>B .10</td>
<td>B .03</td>
<td>B .96</td>
<td>B .54</td>
<td>B .52</td>
</tr>
<tr>
<td>SEB:</td>
<td>1.06</td>
<td>SEB .06</td>
<td>SEB .09</td>
<td>SEB .56</td>
<td>SEB .81</td>
<td>SEB .96</td>
</tr>
<tr>
<td>Expressive Language</td>
<td>B .34</td>
<td>B -.0009</td>
<td>B -.003</td>
<td>B -.17</td>
<td>B -.09</td>
<td>B -.13</td>
</tr>
<tr>
<td>SEB:</td>
<td>.28</td>
<td>SEB .02</td>
<td>SEB .02</td>
<td>SEB .15</td>
<td>SEB .21</td>
<td>SEB .26</td>
</tr>
</tbody>
</table>

| $R^2$               | .15     | .06     | .0    | .19    | .12    | .08    |
| Adj $R^2$           | .06     | -.05    | 8    | .10    | .02    | -.03   |
| $N$                 | 60      | 60      | -03  | 60      | 60      | 58     |
| SE                  | 13.79   | .77     | 60   | 7.16    | 10.22   | 12.18  |
| $F$ Change          | 1.58    | .52     | 1.14 | 1.21    | .32     | .64    |
| Sig. $F$ Change     | .17     | .80     | .75  | 2.12    | .32     | .64    |

Note: Model 1, Response Variable COS; Model 2, Response Variable PAS; Model 3, Response Variable VAS; Model 4, Response Variable AGG; Model 5, Response Variable ADA; Model 6, Response Variable SS.

*p ≤ .05 level **p ≤ Significant at .01 level
NORM REFERENCED THEORY OF MIND

Model 1 presents the multiple regression for COS. This model, $R^2 = .13$, $F(6, 53) = 1.29$, $p = .28$ indicates that, taken together, participant demographic variables, AR scores on the NEPSY-2, and ToM scores on the NEPSY-2 do not predict the level of cooperative behavior exhibited during observations. However, it is interesting to note that participants’ ToM scores do uniquely predict the level of cooperation exhibited by participants, $b = -2.6$, $t(6, 53) = -2.46$, $p = .02$. This finding illustrates that as participants’ scores on the ToM assessment from the NEPSY-2 increased, they were significantly less likely to engage in cooperative behaviors during the study observation. This relationship remains strong even when participants’ expressive language skills were added to the regression model.

Model 2 presents the equation for PAS, which demonstrates that taken together, participant demographics, social perception skills, and expressive language skills do not predict levels of physical aggression exhibited by participants. Model 3 describes the regression for VAS. The overall model for this equation is non-significant.

Model 4 presents the findings for AGG. This model, $R^2 = .17$, $F(5, 55) = 2.19$, $p = .07$, is also not significant. However, upon examination, it is seen that second language presence, $b = -4.67$, $t(5, 55) = -2.35$, $p = .02$, predicts lower teacher ratings of aggression in this population. This relationship holds steady even when expressive language skills are added into the regression equation. Similar to Model 4, Model 5 is not significant. However, examination of the model reveals that speaking a second language contributes to teachers’ rating students with higher levels of adaptive behavior. Lastly, Model 6 presents the equation for SS. This model is not significant and demonstrates that, taken together, demographics, social perception skills, and receptive language do not predict
NORM REFERNCED THEORY OF MIND

teacher ratings of social-emotional skills. As seen the Table 10 the models fitted do not support Hypothesis 3.

**Hypothesis 4.** Hypothesis 4 stated that receptive language performance, as measured by the Fulharty-2, will serve as a mediator for the relationship between ToM and affect recognition skills, as measured by NEPSY-II. Students with stronger expressive language skills, as measured by the Fulharty-2, will exhibit stronger ToM skills, as measured by the NEPSY-II, as well as stronger social skills, as measured by both the behavioral observation and the BASC-2 Table 11 summarizes these findings.
Table 11

Multiple Regression: Summary of Hypothesis 4 Variables on Participants’ Social Skills

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3</th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>$B$</td>
<td>-.09</td>
<td>$B$</td>
<td>-.12</td>
<td>$B$</td>
<td>.63*</td>
</tr>
<tr>
<td>SEB</td>
<td>3.72</td>
<td>$SEB$</td>
<td>.20</td>
<td>$SEB$</td>
<td>.30</td>
<td>$SEB$</td>
</tr>
<tr>
<td>Age</td>
<td>$B$</td>
<td>-.27</td>
<td>$B$</td>
<td>.001</td>
<td>$B$</td>
<td>-.01</td>
</tr>
<tr>
<td>SEB</td>
<td>.32</td>
<td>$SEB$</td>
<td>.02</td>
<td>$SEB$</td>
<td>.03</td>
<td>$SEB$</td>
</tr>
<tr>
<td>Second Language</td>
<td>$B$</td>
<td>-.30</td>
<td>$B$</td>
<td>-.03</td>
<td>$B$</td>
<td>-.15</td>
</tr>
<tr>
<td>SEB</td>
<td>4.05</td>
<td>$SEB$</td>
<td>.22</td>
<td>$SEB$</td>
<td>.32</td>
<td>$SEB$</td>
</tr>
<tr>
<td>Affect Recognition</td>
<td>$B$</td>
<td>.92</td>
<td>$B$</td>
<td>.001</td>
<td>$B$</td>
<td>.04</td>
</tr>
<tr>
<td>SEB</td>
<td>.87</td>
<td>$SEB$</td>
<td>.05</td>
<td>$SEB$</td>
<td>.07</td>
<td>$SEB$</td>
</tr>
<tr>
<td>Theory of Mind</td>
<td>$B$</td>
<td>-2.62*</td>
<td>$B$</td>
<td>.09</td>
<td>$B$</td>
<td>.05</td>
</tr>
<tr>
<td>SEB</td>
<td>1.07</td>
<td>$SEB$</td>
<td>.06</td>
<td>$SEB$</td>
<td>.09</td>
<td>$SEB$</td>
</tr>
<tr>
<td>Receptive Language</td>
<td>$B$</td>
<td>-.02</td>
<td>$B$</td>
<td>.003</td>
<td>$B$</td>
<td>-.03</td>
</tr>
<tr>
<td>Language</td>
<td>$SEB$</td>
<td>.25</td>
<td>$SEB$</td>
<td>.01</td>
<td>$SEB$</td>
<td>.19</td>
</tr>
</tbody>
</table>

$R^2$ | .13 | .06 | .11 | .18 | .12 | .17 |
Adj $R^2$ | .03 | -.05 | .01 | .09 | .02 | .08 |
$N$ | 60 | 60 | 60 | 60 | 60 | 58 |
$SE$ | 13.98 | .77 | 1.12 | 7.21 | 7.21 | 11.55 |
$F$ Change | 1.29 | .52 | 1.06 | 1.97 | 10.21 | 1.77 |
Sig. $F$ Change | .28 | .79 | .40 | .09 | .31 | .13 |

Note: Model 1, Response Variable COS; Model 2, Response Variable PAS; Model 3, Response Variable VAS; Model 4, Response Variable AGG; Model 5, Response Variable ADA; Model 6, Response Variable SS.

* = Significance at .05 level
Model 1 is the model for COS. Overall, this model finds that the combined demographic and predictor variables do not provide information regarding participants’ levels of cooperative behavior. However, it is interesting to note that ToM, $b = -2.62, t (6, 53) = -2.46, p = .02$, does provide unique information regarding the prediction of the cooperative behavior of study participants. This relationship was first noted in Table 8 and remains consistent when participant expressive language skills (seen in Table 10) and receptive language skills, as seen above are added to the equation.

Model 2 presents the regression for PAS, $R^2 = .06, F (6, 53) = .52, p = .79$. The results of this model show that, taken together, the demographic variables, social perception variables, and expressive language do not predict the level of physical aggression exhibited by participants. Model 3, $R^2 = .11, F(6,53) = .52, p = .40$, presents the regression equation for VAS, which illustrates that, taken together, the included predictor variables do not significantly predict the level of physical aggression.

Model 4 presents the regression model for AGG. This model, $R^2 = .18, F (6, 53) = 1.97, p = .09$, is not significant. However it is interesting to note that the presence of a second language, $b = -4.48, t (6, 53) = -2.17, p = .03$, continues to predict lower levels of teacher rated aggression in study participants.

Model 5 presents the equation for ADA. Similar to the other models displayed in Table 8, this model does not significantly predict teacher ratings of adaptive behavior for the students in this sample. The presence of a second language, however, does uniquely add to the prediction of teacher ratings of adaptive behavior on the BASC-2, $b = -5.76, t (6, 53) = 1.97, p = .05$. This relationship has remained consistent throughout the analyses.
Model 6 presents the equation for SS. While the overall equation, $R^2 = .17$, $F (6, 51) = 1.77$, $p = .13$, is not significant. Participants’ receptive language performance does predict teacher ratings of student social skill performance. Interestingly as receptive language performance increases, teacher ratings of student social skill performance decreases, $b = -.51$, $t (6, 51) = -2.46$, $p = .02$.

**Summary of Findings from Hypotheses Testing**

Hypotheses 1-4 were tested using multiple regression analyses. The models fitted did not provide support for any of the four hypotheses presented. Table 12 presents a summary of these findings
### Summary of Findings from Hypotheses

<table>
<thead>
<tr>
<th>HO</th>
<th>Hypothesis</th>
<th>Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>HO1</td>
<td>Preschool children who receive higher scores in ToM and affect recognition skills, as measured by the NEPSY-II, will exhibit higher rates of pro-social behavior and lower rates of antisocial behavior, as measured by behavioral observation, than preschool students who receive lower scores on ToM and affect recognition skills.</td>
<td>Not Supported. ToM provides unique contributions to participants’ cooperation.</td>
</tr>
<tr>
<td>HO2</td>
<td>Preschool children who receive higher scores in ToM and affect recognition skills, as measured by the NEPSY-II, will exhibit higher rates of pro-social behavior and lower rates of antisocial behavior, as measured by teacher ratings on the BASC-2, than preschool students who receive lower scores on ToM and affect recognition skills.</td>
<td>Not Supported. SL provides unique contributions to teacher ratings of participants’ aggression and adaptability.</td>
</tr>
<tr>
<td>HO3</td>
<td>Expressive language performance, as measured by the Fulharty-2, will serve as a mediator for the relationship between ToM and affect recognition skills, as measured by NEPSY-II. Students with stronger expressive language skills, will exhibit stronger ToM skills, as measured by the as well as stronger social skills, as measured by both the behavioral observation and the BASC-2.</td>
<td>Not Supported. ToM provides unique contributions to participants’ cooperative behavior. SL provides unique contributions to teacher ratings of participants’ aggression and adaptability.</td>
</tr>
<tr>
<td>HO4</td>
<td>Receptive language performance, as measured by the Fulharty-2, will serve as a mediator for the relationship between ToM and affect recognition skills, as measured by NEPSY-II. Students with stronger expressive language skills, will exhibit stronger ToM skills, as well as stronger social skills, as measured by both the behavioral observation and the BASC-2.</td>
<td>Not Supported. ToM provides unique contributions to participants’ cooperation. GEN provides unique contributions to participants’ verbal aggression. SL provides unique contributions to teacher ratings of participants’ aggression and adaptability. Receptive Language provided unique contributions to ratings of Social Skills</td>
</tr>
</tbody>
</table>
Supplemental Analyses

After completing the analyses designed to address specific research hypotheses, I conducted analyses to examine the low coefficient alpha score obtained on the ToM scale of the NEPSY-2. As many have noted, alpha is not a property of a measure but rather it is a function of the administration of the scale within a specific population (Bland & Altman, 1997; Hattie, 1985; Tavakol & Dennick, 201; van de Vijver & Leung, 1997). Alpha is considered to be a measure of the homogeneity of scale (Cronbach, 1951). Low values of alpha indicate a lack of homogeneity between the items within a scale (Tavakol & Dennick, 2011). As noted in the methods section of this paper, the alpha for the ToM scale of the NEPSY-2 used in this study were very low, $\alpha = .46$. Such an alpha score is indicative of low reliability and as such, the ToM scores obtained in this study must be viewed with extreme caution. Further examination was conducted to determine the possible cause of the low alpha score.

Two general factors are reported to influence coefficient alpha scores, the length of a scale and the difficulty of items within the scale (Bland & Altman, 1997; Cortina, 1993; Hattie, 1985; Tavakol & Dennick, 2011; van de Vijver & Leung, 1997). Researchers report that items, which are neither too difficult nor too easy, lead to an increase in a scale’s Coefficient alpha. Items, which the average test taker has a probability of $p \approx .6$ of answering correctly, are generally considered best for scale consistency (Wells & Wollack, 2003).

I examined participant responses to the 21 items, which comprise the ToM scale on the NEPSY-2. I determined the number of participants who completed each item, and then calculated the percentage of participants who answered the item incorrectly. Several items were answered correctly by fewer than half of the participant sample. These questions were then examined to determine potential reasons for the high rate of incorrect answers. Two possibilities
were considered. The first is that participants’ ToM skills were not developmentally advanced enough to correctly answer the presented questions (Astington, 1993; Callaghan, 2005; Wellman, Cross, & Watson, 2001). The second was that the way in which specific questions were presented (vocabulary or idiom use) was unfamiliar to participants in this project. Upon examination, over 50% of participants incorrectly answered 9 of the 21 questions that comprise this scale.

Informal content analysis shows that for 3 (33.33%) of the 9 questions, participants’ developmental performance most likely accounts for the majority of participants answering the question incorrectly. These questions address specific concepts noted to be in transition among children within this age group. The other 6 (66.67%) of the 9 questions contained phrasing or situations that may have caused confusion for study participants and lead to incorrect answers that reflect their lack of familiarity with the vocabulary used or situation depicted rather than the concept the test developers sought to examine.

Tables 13 and 14 present a summary of these data. Table 13 presents the data regarding student performance on each of the questions discussed. Table 14 presents a summary of the information presented in each question. A more complete discussion of these findings will be presented in the discussion section.
Table 13

**Percentages of Student Correct and Incorrect Answers to ToM Items**

<table>
<thead>
<tr>
<th>Question Number</th>
<th>N</th>
<th>Item Incorrectly Answered</th>
<th>Item Correctly Answered</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToM 5</td>
<td>67</td>
<td>70.1% (n = 47)</td>
<td>29.9% (n = 20)</td>
</tr>
<tr>
<td>ToM 6</td>
<td>67</td>
<td>74.6% (n = 50)</td>
<td>25.4% (n = 17)</td>
</tr>
<tr>
<td>ToM 7</td>
<td>67</td>
<td>80.6% (n = 54)</td>
<td>13.4% (n = 9)</td>
</tr>
<tr>
<td>ToM 12</td>
<td>67</td>
<td>91.0% (n = 61)</td>
<td>9.0% (n = 6)</td>
</tr>
<tr>
<td>ToM 13</td>
<td>67</td>
<td>73.1% (n = 49)</td>
<td>26.9% (n = 6)</td>
</tr>
<tr>
<td>ToM 14</td>
<td>67</td>
<td>98.5% (n = 66)</td>
<td>1.5% (n = 1)</td>
</tr>
<tr>
<td>ToM 15</td>
<td>67</td>
<td>100.0% (n = 67)</td>
<td>0% (n = 0)</td>
</tr>
<tr>
<td>ToM 16</td>
<td>63</td>
<td>65.1% (n = 41)</td>
<td>34.9% (n = 22)</td>
</tr>
<tr>
<td>ToM 20</td>
<td>63</td>
<td>77.6% (n = 52)</td>
<td>16.4% (n = 11)</td>
</tr>
</tbody>
</table>
Table 14

*Summary of ToM Scale Questions the Majority of Participants Answered Incorrectly*

<table>
<thead>
<tr>
<th>Question Number</th>
<th>Description of Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToM 5</td>
<td>Participants are shown 3 drawings. The first shows a girl riding a dolphin, the second shows a girl lying in bed with a thought bubble above her. Depicted inside the thought bubble is a picture of the girl hugging a dolphin. The third picture is of a girl reading a book with a dolphin on the cover. The question asks “Who can hug a dolphin in real life?”</td>
</tr>
<tr>
<td>ToM 6</td>
<td>The question presents a version of the standard False Belief Test (Hale &amp; Tager-Flushberg, 2003). The participant is shown two identical boxes, both with drawings of red blocks on the covers. One box contains pencils, the other contains red blocks. The child’s attention is directed to box containing the blocks. He is asked to guess what is inside. He is then shown the blocks inside the box. The box is then closed. The child’s attention is next drawn to the other box. He is shown that it contains pencils. With the child’s attention remaining on the box containing the pencils, it is then closed. The child is asked if “your friend came in now, what would your friend think was in this box?”</td>
</tr>
<tr>
<td>ToM 7</td>
<td>The question presents the story of a boy who “has a hard time with spelling” and “didn’t do well on his spelling test at school”. The child’s mother says “You’ll feel better if you go play with X”. The child goes to visit X. We are told X “wanted to play Word Spell” and that the child decided to go home. The participant is asked “Why?”</td>
</tr>
<tr>
<td>ToM 12</td>
<td>The participant is show a picture of a boy cutting paper; tape, glue, a pen, and pencil are on a table in front of the boy. The examiner states that Mrs. X’s “class is making presents for people at the nursing home”. “It was almost time for recess”. The participant is asked the meaning of wrap it up now.</td>
</tr>
<tr>
<td>ToM 13</td>
<td>The participant is shown a photograph of two girls who look very similar. The examiner is directed to point the photo and state that the two girls are sisters. The participant is asked the meaning of two peas in a pod.</td>
</tr>
<tr>
<td>ToM 14</td>
<td>The participant is told a long story. The story describing “Laurie Lamb” who goes into the woods while playing hide and seek. She sees “a big sheep with a white wooly coat, skinny gray legs, a long nose and big teeth, smiling as it came toward her.” Then “Suddenly, a funny-looking wooly brown bear came roaring into the woods and chased off the big sheep.” Participants are told Laurie ran home to her grandmother who stated that her mother would be home soon and “smiled to herself”. The mother sheep returns “out of breath” with wool that was wet, and “brown in places.” Laurie says she won’t go in the woods because of the bear. The examiner is directed to ask the participant to identify the bear and explain why the grandmother smiled.</td>
</tr>
</tbody>
</table>
Table 14 continued

<table>
<thead>
<tr>
<th>ToM 15</th>
<th>The participant is shown a photograph of a boy holding a phone to his ear looking at a woman. The examiner is directed to read a short description describing what the boy is saying to the mother. The participant is asked the meaning of “wrapped around your little finger?”</th>
</tr>
</thead>
<tbody>
<tr>
<td>ToM 16</td>
<td>The participant is shown four photographs of the same girl with different emotional expressions in each photograph and a drawing of the same girl on a roller coaster. The participant is asked to point to which picture shows how the girl feels on the roller coaster.</td>
</tr>
<tr>
<td>ToM 20</td>
<td>The participant is shown four photographs of the same girl with different emotional expression in each photograph and a drawing of the same girl holding an empty cookie jar looking at a boy eating cookies. The participant is asked to point to the picture that shows how the girl feels.</td>
</tr>
</tbody>
</table>
CHAPTER V

Discussion

This chapter presents a summary and a discussion of the results of this project. The clinical significance of the findings is presented, study limitations are then addressed, and lastly, directions for future research are discussed.

Key Findings

The majority of previous studies in this area have examined the relationship between ToM and social skills using participants from predominately Caucasian middle class families (see Slaughter et al., 2015 for a review of the literature). Most earlier research has also relied on homemade measures of ToM (Theory of Mind) skills (Cassidy et al., 2003; Disendruck & Ben-Eliyahu, 2006; McAlister & Peterson, 2013; Watson et al., 1999; Walker, 2005). This study sought to expand upon this research by examining the relationship between a norm-referenced measure of ToM and AR (Affective Recognition) (NEPSY-II) and the social skills demonstrated by multicultural children in their preschool classrooms.

Four specific hypotheses related to the relationship between preschoolers’ social cognitive skills and their social emotional skill performance in the classroom were proposed. Two hypotheses stated that ToM and AR skills as measured by the NEPSY-II would lead preschoolers to exhibit higher levels of pro-social behavior and lower levels of antisocial behavior as measured by both observation during independent play periods and by teacher ratings of social skill performance on the BASC-II. The other hypotheses surmised that participants’ Expressive and Receptive language skills would serve to influence the relationships between these variables.
**Theory of Mind, AR, and observed social skills.** I hypothesized that participants with stronger ToM and AR skills would demonstrate higher levels of pro-social behaviors and lower levels of antisocial behaviors. However, I found no significant relationships between these social cognitive skills and levels of observed verbal and physical aggression. These findings are in line with results established by Hughes et al. (2000) who reported non-significant relationships between ToM performance and observed anti-social skills including snatching and hitting in a sample of preschoolers. The results however do not support the work done by Hughes and colleagues (2001) in a later study, which found that for five year olds, poorer ToM performance did predict higher levels of antisocial behavior during game playing situations.

Several factors may account for these discrepancies. First, the participants in this study exhibited relatively low levels of aggressive behaviors overall. Because few instances of such behaviors were captured, by the observations conducted in this study, there may not have been a large enough sample pool to accurately examine the relationships between these variables (Glass & Hopkins, 1996). Additionally, although Hughes et al. (2001) engineered the situation in their study to increase levels of frustration for participants, the children in this project were engaged in play situations that were part of their regular class routine. In comparison to the competitive game designed by Hughes et al. (2001), the current situations may place significantly fewer demands on the perspective taking and emotion reading skills of participants due to their familiarity with the observation settings and their experiences with the expected antecedents and consequences (Bandura, Adams, & Beyer, 1977). Lastly, the data obtained in this study may not correctly reflect participants’ levels of ToM or AR skills. Coefficient alpha levels for NEPSY-II subsets used in this study were very low (ToM $\alpha = .46$ and AR $\alpha = .60$). Therefore, it is possible
that for the participants in this study, the NEPSY-II did not produce accurate measurements of participants’ performance in either area.

Like Cassidy et al. (2003), who found that for the participants in their sample, stronger skills in emotional recognition and ToM did not significantly impact observed pro-social behaviors, I did not find a significant relationship between these variables and the levels of cooperative behavior exhibited by the current participants. Interestingly, unlike the Cassidy et al. (2003) study, this study found that ToM alone was significantly correlated with lower levels of cooperation, $r = -0.30$, $p \leq 0.05$.

Although all results of the current study must be considered with high levels of caution, the correlation above supports the work of Cowell et al. (2015). This group found that the preschoolers in their sample who successfully completed a presented false-belief measure were less likely to share desired items with peers. Although, consistent with Cowell et al. (2015), my results differ from outcomes obtained by the research groups led by Moore et al. (1998) and Takagishi et al. (2009), who found that young children with stronger ToM skills were significantly more likely to engage in pro-social behaviors such as sharing. One reason for the discrepancies in these results may be that the research teams led by Moore and Takagishi examined the relationship between ToM and sharing in forced choice situations where it was made clear to the participant that engaging in sharing would impact another individual either positively or negatively and that this person would be aware of the action taken by the participant. In contrast, the confederates in the Cowell group’s study (2015) study were unaware of how the actions of participants impacted them.

Cowell et al. (2015) theorize that children with stronger ToM skills are better able to understand the expectations of not only their peers but also of general social settings more
clearly. They are therefore able to determine what behaviors may successfully be performed in different settings. The participants in my study were playing and working in familiar settings with familiar peers and known adults. Like the participants in the Cowell et al. (2015) study, they were aware of the expected consequences of the actions they engaged in. Participants with stronger theory of mind skills might then have been able to determine that there were few, if any, negative consequences for failing to cooperate with peers, which may have lead them to engage in cooperative actions less often than participants with poorer ToM skills.

**Theory of Mind and teacher ratings of preschoolers’ social emotional skills.**

Hypothesis two stated that participants with stronger ToM and AR skills, as measured by the NEPSY-II, would exhibit stronger pro-social and fewer antisocial skills, as measured by teacher ratings on the BASC-II. The results of this study indicate no significant relationships between ToM and AR skills and the three social skill scales administered. These findings are consistent with the results found by Cassidy et al. (2003), who report that ToM alone did not account for any variance in teachers’ ratings of preschoolers’ social skills. My findings, however, contrast with those found by several other researchers. Hughes and Ensor (2006) found that poor ToM performance was related to increased behavior problems based upon parental report in two year olds. Walker (2005) stated that ToM performance accounted for significant differences in teacher ratings of both pro-social and antisocial actions. Specifically, Walker found that, according to teacher report, preschool aged boys with stronger ToM skills were more likely to engage in aggressive and disruptive behaviors, while girls of this age group with strong ToM performance were more likely, according to their teachers, to exhibit pro-social behaviors. Watson et al. (1999) established that ToM skills significantly accounted for 14% of the variance seen in teacher ratings of preschoolers’ social skills for the participants in their sample.
One reason for the differences between the findings of this study and other studies may lie with the way in which ToM and AR were measured. The current study used a standardized, norm referenced tool. However the reliability of this measure with the current sample was questionable. Reliability as measured by coefficient alpha was particularly low for the ToM scale. The unreliability of the tool used in this study may mean that the scale used with this sample was not an accurate measure of social-cognitive skills and may even examine another, unknown, construct entirely.

Interestingly, I found that speaking a language other than English did uniquely account for the variances in teacher ratings of both aggressive and adaptive behaviors on the BASC-II. Speaking English and another language accounted for teachers rating students as less aggressive and more adaptable in the current study. As Cassidy et al. (2003) discuss, past research reports that numerous factors play a role in the development of pro-social behaviors. Cultural factors are one category that may influence such skills (Bronfenbrenner, 1994). Exposure to multiple languages may be considered an example of a cultural influence.

It is possible that participants who come from multilingual backgrounds have stronger pro-social skills because they have wider experience in navigating a variety of social situations than participants from single language backgrounds. Another possibility is that multilingual children may make fewer attribution errors than their monolingual peers during play situations. The participants in this study come from culturally and linguistically diverse classrooms where while English is the language of instruction and the primary language of communication between children and teachers. However, it is not unusual for children to communicate with each other in both English and in their home languages (Spanish, Creole, or French). It is therefore possible these children demonstrate less aggressive behavior with peers in the classroom because they
may understand more of what classmates are saying during play and make fewer errors in interpreting the statements of their peers. The work of Greenberg, Bellana, and Bialystok (2013) provide additional insights. Their study of 82 school age children found that, compared to monolingual participants, bilingual participants demonstrated better perspective taking skills during structured non-verbal tasks, which may account, in part, for stronger social skill ratings from teachers.

**Theory of Mind, language skills, and social behaviors.** Hypotheses three and four state that participants’ language skills will influence the relationship between ToM and AR performance and social skill presentation. Neither hypothesis was supported by the results of this study. However, the relationships found between ToM, gender and dual language exposure and specific social skills within this study were maintained when Expressive and Receptive language skills were added to each regression equation. These findings differ from the results of several past studies. Watson et al. (1999) found that language skills, as measured by the Test for Auditory Comprehension of Language-revised (TACL-R), increased the amount of variance explained by ToM skills alone on teachers’ ratings of social skills. Capage and Watson (2001) found that overall communication skills contributed to the explained variance over and above ToM alone in predicting teacher ratings of preschoolers’ aggression. Cassidy et al. (2003) found that, for their participants, language performance contributed to the variance in both observed pro-social behaviors and teacher ratings of preschoolers’ social skills. One reason the results from the current study differ from previous studies could be that the participants in my sample exhibit relatively little variability in both Expressive and Receptive language skills as measured by the Fluharty-II. This may lessen the relationship that these predictor variables have on the overall analysis (Nathans, Oswald, & Nimon, 2012).
Receptive language skills did explain a proportion of the variance in teachers’ ratings of participants’ social skills. Stronger receptive language skills accounted in part for lower ratings by teachers on the social skills scale of the BASC-II. Such a finding is different than what might be expected based upon past research. Longoria, Page, Hubbs-Tait and Kennison (2007) found that for kindergarten students’ language skills were positively related to teacher ratings of social competence. Monopoli and Kingston (2012) found that stronger language skills, as measured by the Peabody Picture Vocabulary Test-Revised, related to stronger ratings in adaptive behaviors by teachers, as measured by the BASC, for a sample of second graders. One reason my results may differ from those of Manopoli and Kingston may be due to the way in which skill performances were measured by the two studies. Manopoli and Kingston used the Adaptive Behavior Composite of the BASC-II: Teacher Rating Scale-C for children between 6-11 as their measure of social skills. I used three subscales of the BASC-II: Teacher Rating Scale-P for children between 2 and 5 for children in this study. The composite used by Manopoli and Kingston contains both a Study Skills scale and a Functional Communication scale, which require the use of strongly integrated language skills. In contrast, the social skills scale used in this study consists of items that require rote language skills. Scale items ask teachers to rate students’ likelihood to say ‘please’ and ‘thank you’, politely ask for help, compliment peers, and congratulate others. Such rote phrases are often the first category of language used by youngsters (Lieven, Pine, & Barnes, 1992) and may be more frequently used by children with poorer receptive language skills leading to higher ratings by their teachers on this scale.

**Theory of Mind measurement in the current study.** The alpha value for the ToM scale of the NEPSY-2 used in this study was very low, $\alpha = .46$. Such an alpha score is indicative of low reliability and as such, the ToM scores obtained in this study must be viewed with extreme
caution. The length of a scale and the difficulty of items within the scale are factors with affect reliability (Bland & Altman, 1997; Cortina, 1993; Hattie, 1985; Tavakol & Dennick, 2011; van de Vijver & Leung, 1997). Scale questions, which the average test taker has a probability of \( p \approx 0.6 \) of answering correctly, are generally considered best for scale consistency (Wells & Wollack, 2003). Over 50% of participants incorrectly answered 43% of the questions that comprise the ToM scale of the NEPSY-II used in this study. The majority of the items failed by students on this measure questions may have been difficult for the multicultural participants in this study due to their lack of familiarity with the vocabulary used or situations depicted rather than their inability to perform the concept the test developers sought to examine. I conducted an informal content analysis of each of the disputed questions on this measure. This exploration provides a starting point to more fully examine the relationship between the composition of the ToM on the NEPSY-II and its ability to successfully measure this concept in children from ethnically and socioeconomically diverse backgrounds.

Henrich, Heine, and Norenzayan (2010) remind us that all populations develop within different communities and that variability in experience leads to differences in understanding. It is possible that the participants in this study have significant and meaningful differences in the way that they interpret the items presented on the NEPSY-II. This differences may have lead them to interpret presented items in ways not related to the their abilities in ToM performance but in some other, unknown manner.

**Implications for Practice**

The results of this study provide information that is potentially useful for a wide variety of professionals. Although the hypotheses stated at the beginning of this study were not
supported, several findings related to child behavior and assessment may assist teachers, psychologists, and other clinicians in their work with young children.

The results of this study provide an illustration that multiple factors contribute to the development of children’s social emotional skills (Bronfenbrenner, 1994). Both intra- and inter-personal factors were found to influence participants’ social skill performance. No one factor was seen to account for the majority of variance seen in social skills. Such information can be used to design instruction and interventions for early childhood students. Findings indicate that teachers and clinicians should use student strengths to assist in the development of lagging skills (Bagnato, McLean, Macy, & Neisworth, 2011; Zepada, Castro, & Cronin, 2011). Additionally, individualized instruction provides young learners with targeted goals to and methodologies that relate to learning (Greenwood et al., 2011).

The poor reliability of the scale used to assess ToM in this study provides valuable information for clinicians, researchers, and others who work with individuals from diverse backgrounds. Although the NEPSY-II is generally considered to be an instrument with strong psychometric properties and was normed using a large, stratified sample of individuals between the ages of 3 and 16, the measure may not be appropriate for use with children and adolescents from all backgrounds (Brooks et al, 2010; Korkman et al, 2007). Korkman and colleagues provide no information regarding instrument norms for individuals from dual language households. Additionally, although they report parental educational level, they provide no direct information regarding the socioeconomic status of the norming sample. The low reliability of the ToM scale of the NEPSY-II combined with limited information regarding norms for the children in my sample raise significant issues regarding the interpretation of scores obtained by participants. These findings can inform professionals who routinely use norm-referenced and
standardized tools, such as the NEPSY-II, to assess students in schools, clinics, and home-based settings. Best practice standards in assessment require that student performance measured by assessments not normed using a representative sample must be interpreted with caution and qualitative descriptions of student performance rather than standard scores should be used (Comas-Diaz, 2012; Ortiz & Ochoa, 2005; Schon, Shaftel, & Markam, 2008; Schulman, 2002). The findings of this study illustrate the importance of assessing the fit between the individual child being evaluated and the standardization sample.

**Study Limitations**

There are several limitations, which must be kept in mind when evaluating this study as a whole. Four specific areas should be considered: the theoretical construct on which the study was based, the study sample, study instrumentation, and specific procedures.

The theoretical construct of ToM is difficult to separate from other development theories and perspective taking. Theorists need to differentiate ToM from other constructs in order for it to be studied in a coherent fashion. It is difficult to develop standardized tests, such as the NEPSY-2, when one is not sure exactly what one is looking for. Homemade measures such as the Sally-Ann or Bandaid Box tasks can be construed in terms other than ToM. Concrete operational definitions for ToM that differentiate the construct from other theoretical concepts such as Piaget’s conservation or Egocentrism and perspective taking are needed in order to move the field forward.

The sample was small and homogeneous. The small sample size made it difficult to detect anything other than large effect sizes and also may also have limited the power of the study (Glass & Hopkins, 1996). The homogeneity of the sample limited the generalizability of the results (Henrich, Heine & Norenzayan, 2010). Participants in this sample came from low
in 92 income households and a large percentage were linguistically diverse. These two factors are confounded and make it impossible to separate them and therefore interpret study results.

Procedural constraints include relying on verbal parental report to determine participant ethnicity and language use and failing to videotape behavioral observations of students. In addition repeated observations of participant behavior would have provided a more representative description of the actual actions children exhibited in the classroom. Additionally, participants were only observed in low frustration activities (i.e. playing with preferred toys and friends) Observations of participants during activities which might be less preferred, and which are more apt to solicit frustration, such as a teacher directed lesson, might have resulted in observations of more varied behavior. Parents provided information regarding their children’s ethnic background and language use during an in-person or telephone interview with me. A potential issue with the use of self-reports to obtain such information is social desirability. It is possible that the parents of study participants may have felt a need to impress me by stating that their child comes from a specific ethnic background and/or speaks only English in the home (Van de Mortel, 2008). Due to restrictions placed upon the study by the Institutional Review Board, I was not allowed to videotape the behavioral observations conducted for any of the students in this study. Videotaping would have allowed me to revisit the observations conducted by both my RAs and myself and strengthen their reliability. It would have been possible to obtain inter-observer agreement data for a much greater number of observations. Additionally, the reassessment data from the video would have addressed any potential integrity issues such as observer fatigue or distractibility (Haidet et al., 2009).

The last category of limitations involves the measures used during this study. One methodological concern involves the definition of Physical Aggression used in both the
behavioral observation and the BASC-2. Physical Aggression is defined in both these measures as harmful acts directed to another or an object. Because aggression towards objects and aggression towards people are confounded in both the BASC-2 and the observation, it is not possible to look at each aspect of aggression separately. ToM relates to social aggression only. So future measure should only include a measure of social aggression. Only one tool, the NEPSY-II, was used to assess participants’ ToM performance. Several problems regarding the use of this assessment have been previously discussed within this paper. Obtaining additional measures of ToM would have reduced the error associated with using only this single measure (Haynes, Heiby, & Hersen, 2003). Additionally, many of the questions asked and the situations presented on the ToM scale of the NEPSY-II are hypothetical and somewhat abstract. It is possible that examining perspective taking skills using tasks that consist solely of situations familiar to children of this age group may provide greater insight into how these youngsters utilize such skills during day-to-day social interactions (Cowell et al., 2015; Hertzig, Snow, and Sherman, 1989). The low reliability of the NEPSY-II may have contributed to some insignificant results.

**Directions for Future Research**

The results of this study raise many questions both about the development of pro-social behaviors in young children and about the use of norm-referenced assessments in multicultural settings. One future research agenda may be to systematically examine the effectiveness of the ToM scale of the NEPSY-II with diverse populations. Another set of research questions may seek to describe more fully the complexities associated with ToM performance and social skills in children from culturally diverse backgrounds.
The population of students from culturally and economically diverse backgrounds continues to grow in our schools. It is estimated that by 2030 students whose initial language is not English will comprise approximately 40% of school age children (Thomas & Collier, 2002). The development of effective assessment tools and techniques for use with the ever-diversifying student population must be a priority for school professionals. Standardized assessments, such as those used in this study, are components of educational assessments and evaluations for special education services (Haynes et al., 2003). Van de Vijver and Leung (1997) suggest that comparison studies of reliability between specific populations of interest are one way to obtain information regarding the generalizability of scores of norm-referenced measures. In the case of the ToM scale of the NEPSY-II, such studies might contrast the reliability coefficients for this scale with the current sample to samples drawn from middle and upper class backgrounds.

Other work may seek to compare the relationships between demographic, predictor, and response variables within the current sample to a more heterogeneous group of preschoolers. Past research has found mixed results in describing the relationship between ToM skills of preschool children and their social skill performance (see Slaughter et al., 2015 for a review of the literature). Although a recent meta-analysis by Slaughter and her colleagues found that ToM was a modest but significant factor in determining effective peer relations, the findings of the current study raise important points. Research should examine the role that children’s ToM skills play in how youngsters are perceived by their teachers and other adults. Both the current study and Cassidy et al.’s (2003) work found that ToM skills alone did contribute to variance in teacher ratings of student behavior. However, other studies, such as one conducted by Walker (2005), did find a relationship. Future research should systematically use multiple assessments to clarify this relationship.
Additional studies may seek to examine the relationship between dual language use, ToM skills, and social skills behavior. Similar to findings by Greenberg, Bellana, and Bialystok (2013), I found that sample participants with children with dual language backgrounds demonstrated stronger social skills than their peers who were monolingual. Past research, such as that conducted by Goetz (2003) and Farhadian et al. (2010), has found that bilingual preschoolers demonstrate stronger ToM performance than those who speak only one language. Future research may seek to expand upon such findings and examine the role that ToM plays in the performance of social emotional skills for these children.

Lastly, tasks used to examine ToM performance are often somewhat abstract and not related to the day-to-day experience of preschoolers (Cowell et al., 2015). Future studies might seek to develop versions of traditional ToM assessments, which are directly related to the experiences of the participants in their sample and examine the relationship between children’s performance on such assessments and their actions in the classroom.

**Conclusions**

This study sought to examine the effect of ToM and AR performance as measured by a norm-referenced assessment on the social skills of a multicultural group of preschool students. The results of this study expand the research in this area and provide several important results, which can be applied by researchers and clinicians. Study findings suggest that demographic variables play a role in how teachers rate student social emotional skills. Participants in this study with dual language backgrounds were rated as less aggressive and more adaptable than peers who were monolingual. The results also suggest that girls were more likely to engage in verbal aggression than their male counterparts. Lastly, the study found that ToM skills as measured by the NEPSY-II predicted lower levels of cooperative behavior for sample
participants. This finding must be interpreted with extreme caution. Reliability coefficients for this measure for my sample were extremely low and the validity of the measure is therefore questioned (van de Vijver & Leung, 1997).

The results of this study raise several questions. Although having a dual language background supported the social emotional skills of the participants in this study, the exact nature of this relationship is unknown. It may be that, as past researchers propose (Farhadian et al. 2010; Goetz, 2003; Greenberg, Bellana et al., 2013), children with bilingual backgrounds have better perspective taking skills than monolingual youngsters, which in turn lead to stronger social emotional skills. However, this study and others (Cowell et al., 2015; Hughes et al., 2001) found that improved ToM performance did not predict increased pro-social behavior. The exact nature of this relationship remains to be determined.

Additionally, although the ToM scale on the NEPSY- II reports strong psychometric properties (Brooks et al, 2010; Korkman et al., 2007), the low reliability of the ToM scale when used with the current sample raises questions about the validity of this tool when working with individuals from diverse backgrounds. As the numbers of students from culturally and linguistically diverse backgrounds continues to increase, the findings of this study underscore the importance of examining behavior from multiple viewpoints and using appropriate tools (Thomas & Collier, 2002).
Participant Number ____________________________ Date ____________________________

**Behavior Observation Form for Free-play Setting** *(Directions: Place an X in the box if the specific behavior was observed during the interval)*

<table>
<thead>
<tr>
<th></th>
<th>0&quot;</th>
<th>15&quot;</th>
<th>30&quot;</th>
<th>45&quot;</th>
</tr>
</thead>
<tbody>
<tr>
<td>0'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>1'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>2'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>3'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>4'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>5'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>6'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>7'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>8'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>9'</td>
<td>C</td>
<td>R</td>
<td>C</td>
<td>R</td>
</tr>
<tr>
<td></td>
<td>PA</td>
<td>VA</td>
<td>PA</td>
<td>VA</td>
</tr>
</tbody>
</table>

**Observation Scoring Legend**

- **C** = Cooperation
- **PA** = Physical aggression
- **R** = Reassuring
- **VA** = Verbal aggression

*Please turn over for minutes 10-19 →*
Directions: *Place and X in the box if the specific behavior was observed during the interval*

<table>
<thead>
<tr>
<th>Interval</th>
<th>C</th>
<th>R</th>
<th>PA</th>
<th>VA</th>
<th>C</th>
<th>R</th>
<th>PA</th>
<th>VA</th>
<th>C</th>
<th>R</th>
<th>PA</th>
<th>VA</th>
<th>C</th>
<th>R</th>
<th>PA</th>
<th>VA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0&quot;</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>11'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>12'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>13'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>14'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>15'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>16'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>17'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>18'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
<tr>
<td>19'</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
<td>C</td>
<td>R</td>
<td>PA</td>
<td>VA</td>
</tr>
</tbody>
</table>

Count the number of intervals in which each behavior was observed and mark below

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Total Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cooperation</td>
<td></td>
</tr>
<tr>
<td>Reassuring</td>
<td></td>
</tr>
<tr>
<td>Physical Aggression</td>
<td></td>
</tr>
<tr>
<td>Verbal Aggression</td>
<td></td>
</tr>
</tbody>
</table>

*Scoring Summary*
Appendix B

Script for Partial Interval Instruction

I. Today we are going to learn how to conduct a partial interval observation.

II. The training is going to last approximately 60 minutes and will consist of several different activities.
   a. First, I will discuss what partial interval observations are and describe the steps in conducting them.
   b. Then you will watch a video, which describes the steps for conducting this procedure.
   c. You will then practice collecting partial interval data using the 70-second clip presented in the video.
   d. You will then practice collecting partial interval data using a short video clip from another website.
   e. Next we will review the 6 behaviors you will need to observe for this study.
   f. Then you will practice collecting partial interval data on these 6 behaviors using two 10-minute video clips.
   g. Interobserver agreement data with the Primary Investigator will be collected during these observations.
   h. If Interobserver agreement does not meet standards described as excellent then another practice session will occur. To determine this I will calculate a statistic known as Kappa. The Kappa statistic looks at the agreements and disagreements between observers.

III. Discussion of partial interval observations
   a. Partial Interval data collection is a recording strategy that involves whether or not a behavior occurred during a specified time period.
   b. To conduct a partial interval observation. First, the behaviors to be observed are chosen. In this study we will be observing 6 behaviors.
   c. These behaviors are then operationally defined. This means that they are described so that any other researcher can independently measure them.
   d. Here is a sheet listing the operational definitions of the behaviors used in this study.
   e. An appropriate amount of time a behavior might occur is chosen. In this study we are conducting a 20-minute observation.
   f. The time is divided into equal, small intervals. In this study that interval is 15 seconds.
   g. A timer or recorded interval counter is used to track the beginning and end of an observation period. In this case we will use a recorded sound track.
   h. If you observe the behavior occurring during any part of the interval, you mark an X under that behavior category on your data collection sheet.
   i. Do not make any mark if you did not see the behavior occur.
   j. It does not matter how many times the behavior occurs. If it occurs once or many times during the interval, you mark just one X.
   k. Once the session is complete, you tally up the number of intervals where the behavior occurred and record that information in the Scoring Summary section of the Data Collection Sheet (Appendix B).
   l. On some occasions, I will collect observation data with you. I will do this to make sure that both you and I are conducting the observations in the same way. This is a measure of reliability.
called Interobserver Agreement (IOA). I will conduct IOA checks on approximately 20% of the observations done for this study.

IV. View Video from Butterfly Effects.
   a. Hand out practice data collection sheet for this video.
   b. Have RA’s practice data collection.
   c. Review RA’s data collection using the Butterfly Effects video.

V. View Video from Tom McIntyre’s website
   a. Hand out practice data collection sheet for this video.
   b. Have RA’s practice data collection with interval timer played aloud.
   c. Review RA’s practice data collection in comparison to PI’s data collection.
      i. For discrepancies replay the video segment of that interval and review if behavior occurred or did not.

VI. View first 5 minute video of children in a group.
   a. Hand out data collection sheet for this video.
   b. Have RA’s practice data collection with interval timer played aloud.
   c. Review RA’s practice data collection in comparison to PI’s data collection.
      i. Calculate the Kappa for the observation between RA and PI.
      ii. For discrepancies replay the video segment of that interval and review if behavior occurred or did not.

VII. View second video of children in a group.
   a. Hand out data collection sheet for this video.
   b. Have RA’s practice data collection with interval timer played aloud.
   c. Review RA’s practice data collection in comparison to PI’s data collection.
      i. Calculate the Kappa for the observation between RA and PI.
      ii. For discrepancies replay the video segment of that interval and review if behavior occurred or did not.

VIII. For RA’s whose IOA isn’t within defined limits repeat steps VI and VII.

IX. Thank them for taking the time to assist with this project.
Appendix C

Practice Data Collection Sheet for Butterfly Effects Video

Operational Definition of the Behavior:

<table>
<thead>
<tr>
<th>Interval 1</th>
<th>Interval 2</th>
<th>Interval 3</th>
<th>Interval 4</th>
<th>Interval 5</th>
<th>Interval 6</th>
<th>Interval 7</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Summary

<table>
<thead>
<tr>
<th>Number of Intervals behavior occurred</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>
Appendix D

Practice Data Collection Thomas McIntyre’s Video

Operational Definition of behavior: Off task behavior is defined as looking away from the teacher, talking to peers or playing with objects other than those used in the activity.

Interval length: 10 seconds

<table>
<thead>
<tr>
<th>Interval 1</th>
<th>Interval 2</th>
<th>Interval 3</th>
<th>Interval 4</th>
<th>Interval 5</th>
<th>Interval 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval 7</td>
<td>Interval 8</td>
<td>Interval 9</td>
<td>Interval 10</td>
<td>Interval 11</td>
<td>Interval 12</td>
</tr>
<tr>
<td>Interval 12</td>
<td>Interval 15</td>
<td>Interval 16</td>
<td>Interval 17</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Behavior Observation Summary

<table>
<thead>
<tr>
<th>Number of Intervals Behavior Observed</th>
<th>Number of Interval Behavior Not Observed</th>
</tr>
</thead>
</table>
Appendix E

Information about the Study

Aggressive and disruptive behavior problems are a persistent and disturbing group of behavior exhibited by children in all educational settings (Burt et al., 2004; Frick, 2001; O’Reilly et al., 2005).

A potentially promising program of research indicates that disruptive behavior problems are related to deficits in the information processing skills of individuals (Dodge, 1980). One paradigm that may be useful in both describing the problems faced by children with such behavioral difficulties and in developing interventions to target such problems is Theory of Mind (ToM).

ToM refers to a person’s ability to recognize that, first, he himself experiences mental states and, second, that other individuals also demonstrate this ability. Individuals with intact ToM skills understand that their mental states are not always the same as the mental states that others experience. They can then uses this skill to make predictions regarding the behavior of others (Astington & Gopnik, 1991).

The relationship between ToM skills and social skills is complex and not fully understood. One factor that may account for the variability within the relationship between ToM skill and social skills behavior found in research studies is the lack until now of a standardized and norm referenced measure of ToM skills. The NEPSY-II (Korkman et al., 2007) is such a measure.

The purpose of this dissertation study is to examine the relationship between ToM skills, as measured by the NEPSY-II, and social skills exhibited by preschool students. Preschool
students between the ages of four and five will be observed during a free play or independent instructional period. They will also be asked to complete a communication screening measure and a short evaluation of ToM skills. Additionally, their teachers will be asked to complete a short questionnaire about their social skills in the classroom.

This study will help to develop an understanding of the processes involved in the development of strong social skills in young children as well as better methods of identifying social skills problems that may hinder a child’s academic progress.
Appendix F

Consent Letter

Dear Parent:

My name is Sarah Birch and I am a graduate student from the Ph.D. Program in Educational Psychology at the Graduate School and University Center of the City University of New York. My advisor, Dr. Georgiana Tryon, and I would like to include your child, along with his or her classmates, in a research project on development of social skills in young children.

If your child takes part in this project, then either I or one of the research assistants helping with this project will meet with him or her for a short evaluation session. This session can either take place within the school building, during the school day or in your home during a time your child is not in school. The researcher will meet privately with your child in a quiet location, such as a small office. During the meeting, which will last about 20 minutes, the researcher will administer a brief communication screening tool to your child as well as a short social perception measure. Your child’s teacher will be asked to complete a short questionnaire about his or her social skills in the classroom. Additionally, either I or one of the research assistants will observe your child during a free play period in his or her classroom.

Participation in this project is voluntary and involves no unusual risks to you or your child. You may rescind your permission at any time with no negative consequences. Your child can also refuse to participate or withdraw from the project at any time with no negative consequences.

Only Dr. Tryon and I will have access to the information collected in this project, which will be kept in locked storage at the university for a period of three years following the completion of the research. Neither your name nor your child's name will appear in any reports of this research. You have a right to review a copy of any survey, questionnaire, checklist being administered to your child.

Your child's participation in the project will help us develop a understanding of the processes involved in the development of strong social skills in young children as well as to develop better methods of identifying and social skills problems that may hinder a child's academic progress. The results of this study will be used for my dissertation, and may also be used for a scholarly report, journal article, or conference presentation.

In the space at the bottom of this letter, please indicate whether you do or do not want your child to participate in this project. Ask your child to bring one copy of this completed form to his or her teacher by __________. The second copy is to keep for your records. If you have any questions about this research project, please feel free to contact us either by mail, e-mail, or telephone.

Sincerely,

Sarah Birch MS ed, Graduate Student Educational Psychology
Phone: 646-872-7361
Email: SBirch@gc.cuny.edu

Georgiana Tryon, Professor of Psychology
Phone: 212-817-8293
Email: GTryon@gc.cuny.edu

Kay Powell, IRB Administrator
Phone: 212-817-7525
Email: KPowell@gc.cuny.edu

I do / do not (circle one) give permission for my child (______________) to participate in the research project described above.

Date: __________________________

Parent’s signature: ______________________________

I would like you to meet with my child at home to administer the communication-screening tool and the social perception measure.

Please contact me at the following phone number: __________________________
Appendix G

Operational Definitions of Behaviors

Cooperation

- being in close proximity (2 feet or less) to another while engaged in a task that is oriented toward either toward the other or participating in a mutual activity and either verbally or physically engaging to further the activity.

Reassuring

- responding to another’s verbal or physical form of distress with either a verbal or physical expression of comfort.

Physically aggression

- hurting another or damaging materials.

Verbal aggression

- making accusations or angry remarks about another.
References


