Summer 2015

Theory of mind in middle childhood : assessment and prediction

Jeung Eun Yoon
University of Iowa

Follow this and additional works at: https://ir.uiowa.edu/etd
Part of the Psychology Commons

THEORY OF MIND IN MIDDLE CHILDHOOD: ASSESSMENT AND PREDICTION

by

Jeung Eun Yoon

A thesis submitted in partial fulfillment
of the requirements for the Doctor of
Philosophy degree in Psychology
in the Graduate College of
The University of Iowa

August 2015

Thesis Supervisor: Professor Grazyna Kochanska

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Jeung Eun Yoon

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Psychology at the August 2015 graduation.

Thesis Committee:

Professor Grazyna Kochanska, Thesis Supervisor

Todd Kopelman

Molly Nikolas

Julie Gros-Louis

Kristian Markon

ABSTRACT

Theory of Mind (ToM), a social cognitive skill defined as one's ability to attribute mental states to self and others, is considered key for a successful navigation of one's social world. Extensive research has elucidated the early developmental trajectory, predictors, correlates, and outcomes of ToM in the first five years of a child's life. By contrast, although ToM continues to develop beyond age five, and children increasingly begin to function in more complex and interconnected social ecologies, very little is known about ToM in middle childhood. The present study examines ToM development in middle childhood, using a new measure that is age appropriate, innovative, and embedded in the flow of a naturalistic social interaction. Drawing from rich behavioral and report data collected from children, parents, and teachers in a longitudinal study from toddlerhood to middle childhood, interpersonal factors (the child's relationships with the mother, father, and peers), and intrapersonal factors (temperament characteristic of effortful control) are systematically examined to predict individual differences in children's performance in the new ToM measure at age 10. Associations between children's ToM and their broadly ranging, concurrently assessed clinical symptoms are also examined. As a preliminary venture, using a small sample of children with autism spectrum disorders (ASD) and their caregivers, the present study also seeks to establish preliminary criterion validity for the new measure of ToM.

PUBLIC ABSTRACT

Theory of Mind (ToM), a social cognitive skill defined as one's ability to attribute mental states to self and others, is considered key for a successful navigation of one's social world. Extensive research has elucidated the early developmental trajectory, predictors, correlates, and outcomes of ToM in the first five years of a child's life. By contrast, although ToM continues to develop beyond age five, and children increasingly begin to function in more complex and interconnected social ecologies, very little is known about ToM in middle childhood. The present study examines ToM development in middle childhood, using a new measure that is age appropriate, innovative, and embedded in the flow of a naturalistic social interaction. Drawing from rich behavioral and report data collected from children, parents, and teachers in a longitudinal study from toddlerhood to middle childhood, interpersonal factors (the child's relationships with the mother, father, and peers), and intrapersonal factors (temperament characteristic of effortful control) are systematically examined to predict individual differences in children's performance in the new ToM measure at age 10. Associations between children's ToM and their broadly ranging, concurrently assessed clinical symptoms are also examined. As a preliminary venture, using a small sample of children with autism spectrum disorders (ASD) and their caregivers, the present study also seeks to establish preliminary criterion validity for the new measure of ToM.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Theory of Mind (ToM) refers to a set of skills considered critical for understanding and navigating the intricate social world (Baron-Cohen, 1997). To function adaptively in the social world, individuals must realize that others (as well as themselves) have internal states (thoughts, emotions, perceptions, perspectives). ToM is broadly defined as one's capacity to attribute mental states to self and others to understand and predict others' behaviors (Premack & Woodruff, 1978).

Given the key importance of ToM for social cognition and behavior, relevant research has been growing exponentially in developmental psychology and psychopathology, neuroscience, and psychiatry, and in various populations including nonhuman primates, young children, typically developing individuals and multiple atypical populations including individuals with autism, schizophrenia, personality disorders and congenital blindness and deafness (e.g., Brune, Abdel-Hamid, Lehmkamper, & Sonntag, 2006; Flavell, 1991; Peterson, Wellman & Slaughter, 2012; Ghiassi, Dimaggio & Brune, 2010; Minter, Hobson, & Bishop, 1998; Tomasello, & Josep, 1997).

Extensive developmental research on ToM has elucidated its trajectory. Although there is some evidence of rudimentary awareness of others' internal states in infancy (Meltzoff, 1995), most researchers agree that ToM skills begin to emerge in toddlerhood. Toddlers begin to demonstrate their understanding that one can have different perceptual information (i.e., see different things) than the other depending on their physical locations (Sodian, Thoermer & Metz, 2007). Between age 3 and 4, toddlers develop a more sophisticated appreciation of mental state terms (e.g., desire, intention, belief, and etc.) and use that knowledge to understand and predict others' mental states and behaviors (Wellman & Liu, 2004). Eventually, children aged between 4 and 5 reach a pivotal level of ToM skills, commonly called a first-order false-belief understanding. They now understand that others can hold a false belief that deviates from reality (and from their own belief that reflects reality), and behave correspondingly with that false belief (Wellman, Cross, & Watson, 2001). Beyond this pivotal turning point, ToM continues to evolve in late childhood and early adolescence; children develop appreciation of more nuanced mental states (e.g., second-order false belief, humor, sarcasm, white lies, bluff, reading facial expression) and begin to apply them in their social settings (e.g., conversations) (e.g., Happe, 1994; Peterson, Wellman, & Slaughter, 2012).

The early development of ToM in infants and toddlers has received extensive attention from developmental psychologists. Critical precursors (e.g., eye gaze, joint attention, imitation skills), correlates (e.g., verbal ability, working memory), and predictors (e.g., a positive parent-child relationship, effortful control) of ToM skills in early development have been well documented (e.g., Carlson, Moses, & Breton, 2002; Charman et al., 2000; Slade & Ruffman, 2005; Symons & Clark, 2000).

In contrast, factors related to ToM development in older children, who have already achieved an understanding of false belief, have received relatively little attention. Consequently, little is known about the qualities of the early parent-child relationship or characteristics of the child that predict ToM development in middle childhood. As well, it is not known whether experiences such as functioning in peer groups contribute to the development of ToM in older children beyond the impact of early factors, although social interactions have been implied as key for skills very closely related to ToM, such as role taking (Selman, 1980).

My dissertation is nested within a large ongoing study of 102 community mothers, fathers, and typically developing children (Family Study), for whom rich behavioral and report data have been collected at multiple assessments from infancy to age 10. Those data include broadly ranging measures of children's characteristics and their relationships with parents. Measures of children's ToM at preschool age have also been collected. The availability of those data makes it possible to pursue three goals.

As stated above, generally, the great majority of research on ToM in developmental psychology has focused on toddler and preschool age, whereas ToM in older ages has been greatly under-studied. *To examine ToM in middle childhood (at age 10) is the first, and general, goal of my dissertation. To achieve this first goal, I pursued two tasks. One, I developed a novel ToM measure; two, I validated that new measure.*

The overwhelming majority of measures have relied heavily on abstract, hypothetical tasks such as various false-belief tasks (FBT). The popularity of FBTs has raised the following concerns. First, FBTs assess an understanding of one type of mental states, a (false) belief whereas ToM applies to a much broader array of mental states (Baron-Cohen, O'Riordan, Stone, Jones, & Plaisted, 1999). Second, most of ToM measures (e.g., Happe's Strange Stories) largely focus on the attribution of different kinds of mental states (e.g., embarrassment, sarcasm, humor) without explicitly paying attention to another critical component of ToM skills, understanding and prediction of others' behaviors—a skill rooted in perspective-taking ability. Third, and perhaps most importantly, although ToM is a key ability needed to navigate social interactions, surprisingly, not a single measure has been developed to assess ToM as embedded in the flow of an actual social interaction. *To develop such an interactive measure that targets perspective-*

taking skills and is embedded and assessed in the child's ongoing social interaction with a social partner (parent) was the initial task involved in accomplishing my first goal.

To that effect, I have adopted a naturalistic yet carefully scripted paradigm where the parent-child dyad is working together to complete a puzzle set. The paradigm has been developed in the Minnesota Longitudinal Study (Kovan, Chung, & Sroufe, 2009; Sroufe, Egeland, Carlson, & Collins, 2005) to observe parent-child interactions (but unrelated to ToM). During the task, a member of a dyad ("receiver") is blindfolded while physically completing a puzzle set, and the other member ("guide") verbally guides his or her partner to help complete the puzzle. Consequently, the task requirements for the guide call for taking on the perceptual perspective of the receiver throughout the procedure, and thus, they embody a certain set of ToM skills applied dynamically during the interaction.

Note that individuals in various professions encounter similar challenges. Take, for example, educators of the visually impaired. American Foundation for the Blind (AFB) emphasizes how important it is for teachers to be cognizant of their visually impaired students' unique perceptual experiences and perspectives. The AFB guidelines strongly stress the importance of teachers' perspective-taking (or ToM) skills as key to successful teaching, and reiterate the need for keen awareness of students' not having the same perceptual information as sighted individuals.

Likewise, to be successful in a puzzle task described above, the guide should always keep in mind that the receiver does not have the same perceptual information as he or she does. Thus, the guide needs to deploy perceptual perspective-taking aspects of ToM. To reiterate, my first task is to develop a measure of ToM skills as they are dynamically implemented in the flow of the child's interaction with the blindfolded parent in the puzzle paradigm.

The second task involved in accomplishing the first goal is a two-prong methodological objective to validate my new measure of ToM. One, I examined heterotypic continuity between the new measure and two traditional ToM measures administered at preschool and kindergarten age to the same children (52 and 67 months, respectively). Briefly, heterotypic continuity refers to the conceptual coherence of psychological variables that are measured over time and by different measures, yet theorized to represent the same underlying or latent construct (Caspi & Shiner, 2006). To find that children's ToM skills measured at different developmental periods and by different tasks are significantly related with each other would be evidence for the conceptual coherence. Two, as a preliminary and frankly exploratory venture, I implemented my new measure in a small group of age-matched children with the autism spectrum disorders (ASD) and their parents. Vast evidence has conclusively documented that children with ASD diagnoses show inferior performance on ToM tasks, compared to typically developing children or children with disorders other than ASD (i.e., intellectual developmental disorder). To find that those children's performance is significantly worse than the children's in Family Study would constitute evidence of criterion validity.

My second goal is to examine predictors of children's individual differences in ToM in middle childhood. There is extensive literature on predictors of ToM, including several characteristics of the child's interpersonal relationships (e.g., family and peer relations) and individual qualities (e.g., temperament), but most of it focuses on ToM at early age.

Early parent-child relationship characterized as sensitive, responsive, and affectively positive has been shown to positively predict children's subsequent ToM skills (Meins, Fernyhough, Russell, & Clark-Carter, 1998; Symons & Clark, 2000; Fonagy, Redfern, & Charman, 1997). However, because most research has focused on maternal relationships, whether or not similar effects will be found in paternal relationships is largely unknown. With regard to peer relations, social competence has been inherently associated with ToM skills. Of course, better ToM skills have been repeatedly shown as promoting social competence (e.g., Cassidy, Werner, Rourke, Zubernis, & Balaraman, 2003). It is likely, however, that the reverse is also true, and that having positive experiences in peer contexts promotes role-taking and other ToM skills (Selman, 1980). With regard to the child's individual qualities, the role of child temperament, in particular effortful control, has been well established as a predictor of ToM skills (e.g., Benson, Sabbagh, Carlson, & Zelazo, 2012; Razza & Blair, 2009).

Although separate literatures exist on various factors that contribute to the development of ToM skills, they are fragmented and lack integration. To my knowledge, no study has brought together both interpersonal and intrapersonal variables to systematically examine their relative impacts on the longitudinal trajectory of the development of ToM in individuals. *The second goal of my dissertation project is to examine both interpersonal (i.e., early parent-child relationship, peer competence) and intrapersonal, or individual (i.e., effortful control) factors in relation to the future development of the advanced stage of ToM development assessed at the age of 10.*

My third goal is to examine children's ToM in the context of various clinical symptoms. Not much is known about ToM development in the context of clinical disorders except for the autism spectrum disorders (ASD). Substantial social difficulties, however, are present not only in the ASDs, but also in many other clinical disorders, such as attention deficit hyperactivity disorder (ADHD), conduct disorder (CD), or internalizing problems, such as anxiety of depression. Given ToM's key role in social cognition and social functioning, the investigation of ToM in the context of those disorders should be extremely informative. *Thus, the investigation of ToM development in the context of broadly-ranging clinical symptoms is my third goal.*

ToM in Middle Childhood

In general, researchers have acknowledged that much more remains to be learned about ToM development beyond 5-year-olds' mastery of first-order false belief understanding (e.g., Baron-Cohen et al., 1999; Filippova & Astington, 2008; Smith & LaFreniere, 2009). Goldstein and Winner (2010) stated that "there is no consensus yet about what constitutes more advanced theory of mind skill once the ability to represent false belief has been attained (p. 449)." Certainly, less is known about ToM skills in middle childhood than about those skills in preschoolers. However, there has been some effort to uncover the nature of ToM skills in middle childhood (e.g., Happe, 1994; Smith & LaFreniere, 2009), and thus I will review relevant studies, and critique the limitations of the current literature in the following section. Note that I refer to middle childhood as ages 6-11 (Centers for Disease Control and Prevention, 2015).

Advanced ToM tasks in middle childhood. To assess advanced ToM skills beyond the first-order false-belief understanding, second-order FBTs have been devised (Perner & Wimmer, 1985). Second-order FBTs assess one's understanding of someone's inference about someone else's mental state (compared to first-order FBTs, which assess one's understanding of only someone else's mental state) and one's prediction of others' behaviors; thus, second-order FBTs assess doubly-embedded representations of thoughts or beliefs (Perner & Wimmer, 1985). Children typically pass these more complicated second-order FBTs around the age of 6-7, supporting the appropriateness of those tasks in the early middle childhood (Perner & Wimmer, 1985; Sullivan, Zaitchick, & Tager-Flusberg, 1994).

Whereas the FBTs are extremely widely used, and considered the gold-standard "litmus tests" to determine whether or not a child has reached a pivotal level of ToM skills, a few limitations have been noted. First, as previously noted, FBTs typically assess only one type of mental state, (false) belief, whereas children develop an appreciation and understanding of a wide range of mental states, particularly as they increasingly function in more complex social worlds (e.g., peer interaction, Hartup, 2005). Second, the explicit nature of the FBTs (i.e., asking participants a question "where would Sally look for her marble?") has failed to reflect the implicit and dynamic nature of the social demand in real life (Klin, 2000). Third, FBTs measure ToM in an either/or test—subjects are judged to pass or fail, and either to have or not have ToM. This dichotomous method of measuring obscures the actual spectrum of ToM abilities displayed in individuals (Klin, 2000).

To address the shortcomings of the FBT, new measures have been developed to assess ToM skills in middle childhood. For example, Happe's *Strange Stories* task (1994) has been widely used to investigate school-aged children's ability to appreciate remarks reflecting others' nuanced internal states such as white lie, joke, lie, figure of speech, double bluff, or irony (e.g., Ronald, Viding, Happe, & Plomin, 2006; Meins, Fernyhough, Johnson, & Lidstone, 2006). Baron-Cohen and his colleagues (1999) developed the *Faux Pas* test to assess whether school-aged children can detect the faux pas (i.e., a social blunder; for example, a speaker said, "what a cute boy!" to a little girl). Baron-Cohen et al. (1999) posited that the *Faux Pas* test measures a more sophisticated level of ToM because children need to take both the listener's and speaker's perspectives to successfully detect a faux pas (i.e., the listener's awareness that the speaker did not intend to offend him or her). In addition, originally developed to assess adults' social cognition, the *Reading the Mind in the Eyes* test (Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001) has been used to assess children's ability to decipher internal states from photographs of a person's eye region (e.g., Moor et al., 2012; Peterson & Slaughter, 2009; Sharp, 2008). Lastly, Smith and LaFreniere (2009) examined children's ability to detect deception and use the information to improve their prediction about their partner's intentions in a game situation. Children watched video clips of an experimenter hinting the color of cards by glancing or pointing to the cards, detected whether the hints were deceptive, and used the information to win the game (Smith & LaFreniere, 2009). Notably, and pertinent to my project, the study by Smith and LaFreniere (2009) is the only one that extended a typical ToM assessment to include not only children's ability to detect deception, but also their ability to *use* that understanding to their advantage in an interactive task resembling real life.

The aforementioned studies have revealed that in middle childhood, children begin to appreciate more subtle and nuanced social remarks such as sarcasm (Happe, 1994), or unintentional insults (Baron-Cohen et al., 1999), to identify internal states from facial expressions around eyes (Baron-Cohen et al., 1997; 2001), and to recognize deceptive information and apply it to their own advantage (Smith & LaFreniere, 2009). Age trends, consistent across studies, indicated that performance improves over the course of middle childhood, suggesting that ToM skills in middle childhood indeed continue to develop after the original mastery of false-belief understanding at preschool age.

Limitations of ToM research in middle childhood. Although the advanced ToM measures reviewed above capture more aspects of ToM skills in middle childhood than elementary ToM measures do (i.e., a range of mental states assessed), research on advanced ToM skills is still very sparse. Whereas decades of research have produced empirically well-supported developmental trajectory of early ToM skills in typically developing children (see Introduction, p. 1-2), research interests in ToM skills in middle childhood is far less advanced.

One of the limitations in the existing ToM research in middle childhood is that most frequently used advanced ToM tasks (e.g., the *Strange Stories* task, the *Faux Pas* task) are confounded with verbal and reading skills, largely emphasizing the aspects of ToM skills underpinned by conceptual understanding. In addition, perhaps most importantly, to my knowledge, all tasks employ a question-and- answer format, and are not deployed in the context of a dynamically flowing social interaction, with one notable exception of Smith and LaFreniere (2009).

Because of this dominant manner of assessing ToM skills, aspects of ToM that are perhaps more closely related to social behavior (i.e., prediction of others' behaviors) may have been inadequately assessed. As a result, nothing is known about the extent to which children implement ToM skills during their social interactions with partners. In middle childhood, when children are immersed in an increasingly complex social world, the behavioral aspect of ToM that is appropriately deployed in a social setting may be as important as the aspect of ToM that is mastered cognitively. Little is also known about the relation between children's performance in traditional ToM tasks and their ability to use their ToM skills in the context of social interaction.

Individual Differences in ToM Development

Given the significant role of ToM in socio-cognitive development, factors that contribute to individual differences in ToM have been widely investigated, especially in the early years. Different explanations have been offered. At one extreme, ToM is suggested to be "turned on" via a specific module in the brain dedicated to ToM (e.g., Fodor, 1983; Leslie, 1987). At the other extreme, ToM is believed to develop via early emotional, social, and cognitive experiences (e.g., Bartsch & Wellman, 1995; Gopnik & Meltzoff, 1997; Perner, 1991). In the present work, I focus on experiential and temperament factors as contributors to ToM development.

Family variables, such as the parent-child relationship, family size (number of siblings), and family socioeconomic status (SES, e.g., household income, parental education) have been intensely examined as predictors of children's ToM development (e.g., Cole & Mitchelle, 1998; Ontai & Thompson, 2008; Perner, Ruffman & Leekam, 1994). Then, as children begin to interact more with peers (e.g., in daycare, school), those experiences, especially peer competence, have been examined as both a predictor and outcome of ToM skills (e.g., Razza & Blair, 2009).

Individual, intrapersonal characteristics such as temperamental traits, mostly those related to effortful control, have been also examined as factors that influence individuals' ToM development (e.g., Carlson, Moses, & Claxton, 2004). However, most research has been conducted in early childhood, and therefore little is known about contributors to individual differences in ToM development in middle childhood.

Interpersonal factors.

Parent-child relationship. The importance of family context for children's sociocognitive development has been repeatedly demonstrated in numerous studies (e.g., Dunn, Brown, Slomkowski, Telsa, & Youngblade, 1991; Fonagy & Target, 1997; Symon & Clark, 2000). Among the familial environmental factors contributing to individual differences in ToM, the early parent-child relationship has been most intensely investigated.

Several aspects of parent-child relationships have been examined as contributors to individual differences in children's ToM. Parental warmth (Hughes, Deater-Deckard, & Cutting, 1999), mothers' elaborative discourse style (Ontai & Thompson, 2008), and discipline style (*laxness*, Guajardo, Snyder, & Petersen, 2009; *severity of discipline*, Hughes et al., 1999) have all been linked concurrently to 3- to 5-year-old children's performance on ToM tasks. Some relations were moderated by the child's gender. Parental warmth was positively correlated with girls' ToM performance, but not with boys'; and severity of discipline was positively correlated with boys' ToM performance, but not with girls' (i.e., the more severe discipline for boys, the better boys' ToM performance) (Hughes et al., 1999).

With regard to longitudinal links, maternal sensitivity and secure attachment have been positively associated with later ToM development in young children (Meins et al., 1998; Symons & Clark, 2000). Parental use of mental state language also predicted children's future ToM development (Meins et al., 2002; Ruffman, Slade, Devitt, & Crowe, 2006; LaBounty, Wellman, Olson, Lagattuta, & Liu, 2008).

The association between early positive parent-child relationship and children's socio-cognitive development has been largely conceptualized in the context of attachment theory

(Meins, 1998; Sharp & Fonagy, 2008). Research has shown that securely attached children engage in more sophisticated pretense play (Belsky, Garduque & Hrncir, 1984), and are more receptive to others' suggestions during pretense play (Mein & Russell, 1997; Slade, 1987) than insecurely attached children. Security has been seen as fostering children's ability to assume different perspectives in their play and interaction with others (Meins et al., 1998).

Further, following Ainsworth and colleagues' seminal work on antecedents of secure attachment (Ainsworth, Bell, & Stayton, 1971), maternal sensitivity has been considered a potent early parenting characteristic that influences children's early social, emotional, and cognitive development more broadly (Meins et al., 1998; Symons & Clark, 2000). Sensitive mothers are attuned to their children's needs and wants, and respond promptly and appropriately to their children's signals, needs, and desires (Ainsworth, Blehar, Waters, & Wall, 1978). The child's resulting sense of security leaves him or her open to exploration of and learning about the world. Meins (1997) refers to "mind-mindedness" to describe sensitive mothers' tendencies to treat their children as individuals with their own minds, and to focus on the mental-state aspects of their interactions (e.g., commenting on the child's mental states and emotions). Meins et al. (2002) found that mothers' mind-mindedness at 6 months positively predicts children's ToM scores at 48 months, and explained their significant findings in terms of a scaffolding context: Mothers provide their children with a learning environment in which children incorporate mental-state comments into their behaviors. Taken together, sensitive mothers appropriately and consistently meet their children's needs, creating an environment conducive to the child's exploration and learning. That environment is rich in references to internal states and emotions, and thus children are in a good position to learn mental states of their own and others (Meins et al., 2002; Symons & Clark, 2000).

Kochanska (1997, 2002) suggested that secure attachment may be one aspect or form of a more general construct of *the mutually responsive orientation* (MRO). MRO refers to the set of relational qualities that emerge in some parent-child dyads. Such dyads are mutually responsive, willingly cooperative, coordinated, harmonious, “in sync”, attuned to each other, and affectively positive in their interactions (Kochanska, 1997, 2002; Kochanska & Kim, 2014; Kochanska, Kim, Boldt, & Yoon, 2013). MRO has been robustly shown to have important and broadly-ranging implications for children’s future social-emotional development: Kochanska and colleagues’ findings (Kochanska, 2002; Kochanska & Aksan, 2006, Kochanska et al., 2013) that children growing up in highly mutually responsive relationships are more willing to adopt, embrace, and internalize parental values and standards of behavior seem particularly relevant in that MRO promotes children’s willingness to represent others’ perspectives. Indeed, Novak (2012) posited that MRO and its positive effects in children’s subsequent socialization seemed precursors of the child’s development of perspective taking and related skills (thus, ToM). To empirically test if MRO predicts subsequent ToM development will be a novel contribution of this work.

Typically, studies that investigated the effects of parent-child relationship on children’s ToM have followed children from infancy to preschool years. This is understandable given the importance of that relationship in early childhood and the predominant focus of ToM research on the first few years. Nonetheless, as argued earlier, ToM continues to develop in middle childhood. Few studies have investigated the role of the quality of the early parent-child relationship for the development of advanced ToM skills assessed in middle childhood. Investigating whether early relationships continue to predict ToM development in middle childhood may be as important as in early childhood given that middle childhood is a period when children begin to function in more complex social worlds (e.g., peer groups, school, community organizations, Hartup, 2005) outside the family context.

Another important limitation of the extant literature is the predominant focus on mother-child relationships, with the exclusion of fathers. A few exceptions suggest that links between mother- and father-child relationships and ToM may indeed be different. LaBounty et al. (2008) reported different associations between mothers' versus fathers' use of mental-state language and children's understanding of mind at 3.5 and 5 years; mothers' language was associated with children's emotion understanding, and fathers' language was associated with children's false-belief understanding. The dearth of research on father-child relationships is an important gap in literature on children's ToM.

Peer competence. From a social development perspective, rich social interaction in the early years of the child's life, primarily with parents, forms a foundation for the basic ToM skills, and the ToM skills, in turn, support the child's functioning in his or her social interactions with others. As reviewed above, research has empirically supported the claim that children whose parents are warm and sensitive develop better ToM skills (e.g., Symons & Clark, 2000). As well, children with better ToM skills have better social outcomes, such as peer acceptance (Cassidy et al., 2003; Slaughter, Dennis, & Pitchard, 2002), stable mutual friendship (Peterson & Siegal, 2002), and social competence (Smiley, 2001; Watson, Nixson, Wilson & Capage, 1999).

However, a converse relation has also been proposed and supported in (few) existing longitudinal studies: Children's experiences in peer contexts can promote (or undermine) the development of ToM. The relevant studies, however, are very few. I was able to locate three studies that examined the effects of peer competence on subsequent ToM skills in children, with only one study using a sample of middle-childhood children. Using a sample of very young children, Suway and colleagues (2012) found that negative peer interactions at the age of 2 predicted poor ToM skills at the age of 3 (Suway, Degnan, Sussman & Fox, 2012). Using a slightly older sample of children, Razza and Blair (2009) reported that children who are rated as highly socially competent by their teachers at the age 5 performed better on the ToM tasks in the following year. Finally, Banerjee, Watling, and Caputi (2011) reported that peer rejection at the age of 7, and at the age of 9 negatively predicted their ToM performance at the age of 8, and at the age of 10, respectively.

Thus far, available studies suggest that peer competence is related to subsequent ToM skills. However, much more research is needed to elucidate the effects of peer competence on ToM skills in middle childhood. Furthermore, although children in middle childhood function in one, integrated world of social relationships -- with peers and parents (Hartup, 1979; Parke & O'Neil, 1999; Sroufe, Egeland, & Carlson, 1999), no studies have simultaneously examined those two kinds of social relations and their roles in predicting children's ToM in middle childhood.

Intrapersonal factors.

Children's temperament qualities have been linked to ToM. In particular, effortful control (EC), a class of self-regulatory mechanisms, considered an aspect of child temperamental individuality (Rothbart, 1989; Rothbart & Ahadi, 1994; Rothbart & Bates, 1998), has been studied as a contributor to ToM. Rothbart and colleagues define effortful control as the ability to suppress a dominant response to perform a subdominant response. EC is a multifaceted construct that encompasses various abilities such as slowing down motor activity (i.e., regulating one's gross motor activity, for example, walking on the line slowly), conflict-inhibition (i.e., suppressing a dominant response to correctly initiate a subdominant response to a signal during tasks such as go-no go tasks), delay-inhibition (i.e., purposefully delaying one's urge to obtain a reward), modulating voice (i.e., deliberately lowering one's voice volume in response to exciting stimuli), cognitive-flexibility (i.e., flexibly switching a response to a signal), and working memory (i.e., remembering and maintaining the rules to correctly perform on tasks such as go-no go tasks) (e.g., Putnam, Gartstein, & Rothbart, 2006; Rothbart, Chew, & Gartstein, 2001; Rueda, 2012; Wolfe & Bell, 2007; Zhou, Chen, & Main, 2012).

In the past two decades, research investigating associations between EC and ToM has blossomed. In this context, note that executive function (EF) skills, closely related construct of EC, are also widely investigated as correlates of ToM (e.g., Benson & Sabbagh, 2010, Pellicano, 2007). Briefly, EF skills are defined as higher-order cognitive skills that include "regulation of attention, inhibition of inappropriate responses, coordination of information in memory, organizing and sequencing of behavior, and flexibility in doing all of the above" (NICHD Early Child Care Research Network, 2005, p.100). To delineate the commonalities and differences between EC and EF is beyond the scope of my project; however, researchers have noted the overwhelming similarities between the two constructs more than the differences (e.g., Liew, 2012; Zhou et al., 2012). For example, Zhou and colleagues (2012) stated in their review of the two constructs that "both constructs imply self-control in the face of conflicting or competing demands and are consistent with the broader definition of self-regulation (p.113)." While acknowledging the significant overlap between the two constructs, I will consistently use the term EC in my discussion of individual differences in self-regulatory ability given the focus of my project on children's temperament qualities.

The intense interest in the associations between EC and ToM, especially children's understanding of false-belief, may be due to the fact that the two characteristics emerge almost simultaneously and develop in parallel, throughout the toddler and preschool years (e.g., Benson et al., 2012; Carlson et al., 2004; Cole & Mitchell, 2000). A sizeable body of evidence has shown robust links between children's EC and performance on FBTs, above and beyond age and verbal intelligence (e.g., Carlson et al., 2004; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). A meta-analysis by Perner and Lang (1999) found a strong effect size of 1.08 for the relation between ToM and executive skills.

To explicate the nature of this strong relation, researchers have examined, and found, significant relations between distinct EC abilities (e.g., conflict-inhibition, Carlson et al., 2002; working memory, Davis & Pratt, 1995; cognitive flexibility, Frye, Zelazo, & Palfai, 1995), and children's FBT performance. Further, among these various abilities, conflict-inhibition, which involves children's ability to suppress a dominant response and activate a subdominant response, has been most consistently and strongly associated with early emerging ToM ability (Benson et al., 2012; Carlson, Mandell, & Williams, 2004; Carlson et al., 2004; Sabbagh et al., 2006; Lewis, Huang, & Rooksby, 2006).

Perhaps the reason for this robust association between conflict-inhibition and FBTs may be due to a common underlying challenge involved in these two tasks: To perform successfully, children are required to inhibit their dominant responses (e.g., their own view of the reality) and initiate subdominant responses (e.g., a mistaken view of someone else). Another reason for this association may be the immense popularity of FBTs in studying the ToM development of young children. Whereas FBTs are critical measures to investigate young children's understanding of intent and knowledge, questions arise whether EC would also have robust relations with older children's advanced ToM skills assessed by tasks others than FBTs. Stronger associations between EC and ToM have been found in a younger group of children compared to an older group (Fahie & Symons, 2003). Perhaps EC, particularly conflict-inhibition, becomes less predictive of ToM assessed with tasks appropriate for older children in middle childhood.

To my knowledge, EC has not been examined in relation to ToM development in middle childhood. Also, perhaps even more importantly, studies conducted to date have found associations between ToM and "cool" EC, referring to a more abstract form of self-regulation such as conflict-inhibition (Brock, Rimm-Kaufman, Nathanson, Nathanson, & Grimm, 2009; Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Zelazo & Muller, 2002). No study, to my knowledge, has examined associations with "hot" EC, which refers to a self-regulatory ability in affectively charged tasks, typically involving hedonically attractive, highly salient rewards such as delay-inhibition (Allan & Lonigan, 2011; Kim, Koenig Nordling, Yoon, Boldt, & Kochanska, 2013; Shoda, Mischel, & Peake, 1990; Sulik et al., 2010). An investigation of the relations between both "cool" and "hot" aspects of EC and ToM in middle childhood will substantially complement the literature.

ToM in the Context of Clinical Symptoms

Deficient social functioning is a key symptom in many forms of children's psychopathology (e.g., ASD, ADHD, Oppositional Defiance Disorder (ODD), CD). Surprisingly, despite the abundant evidence that ToM is strongly implicated in children's social relations, research on links between ToM and broadly-ranging clinical symptoms (with an exception of ASD) is extremely limited.

The discovery that autistic individuals characteristically fail the FBTs (Baron-Cohen, Leslie, & Frith, 1985) spurred intensive research regarding ToM in ASD. It is now well established that children with ASD are far more likely to fail FBTs, and the proportion of who fail is disproportionately larger than both developmentally delayed and typically developing individuals (e.g., Baron-Cohen, 1989; Pellicano, Maybery, Durkin, & Maley, 2006; Prior, Dahlstrom & Squires, 1990). Because not all with ASD fail the FBTs, advanced ToM tasks such as Happe's (1994) *Strange Stories,* and Baron-Cohen et al.'s (1999) *Faux Pas* test were developed to more adequately capture individual variability within the ASD population. Supporting the centrality of ToM deficits in ASD, children with ASD consistently perform worse on these measures than control groups matched on age and verbal ability (e.g., Baron-Cohen et al., 1999; Happe, 1994).

But much less is known about ToM in children with other clinical symptoms. Children with ADHD have social problems due to their inattentiveness, impulsivity, and hyperactivity (Nijmeijer et al., 2008). However, according to a recent review by Uekermann and colleagues (2010), there have been only four studies (Buitelaar, van der Wees, Swaab-Bameveld, & van der Gaag, 1999; Charman, Carroll, & Struge, 2001; Dyck, Ferguson, & Shochet, 2001; Perner, Kain,

& Barchfeld., 2002) that investigated ToM in those children, with mixed results (see Uekermann et al., 2010). Before further discussing the results of these studies, I will also mention that it is widely known that children with ADHD symptoms commonly have poor EF skills (e.g., Fahie & Symons, 2003). As before, I will consistently use the term “EC”, (see Zhou et al., 2012, for a discussion of EC and EF).

Some studies have found that whereas children with ADHD perform worse on EC tests, they do not differ from typically developing children on ToM tasks (Charman et al., 2001; Perner et al., 2002). Fahie and Symons (2003) reported that EC, not ToM abilities, was correlated with social problems in children who were referred to the mental health clinic for attention difficulties. Whereas these findings may suggest that children with ADHD symptoms are not impaired on ToM understanding, they may also indicate that ToM tasks used in the studies were not sensitive to assess ToM deficits in children with ADHD, especially given the wide age span in these studies (e.g., 9 – 16 years old, Dyck et al., 2001). Indeed, Fahie and Symons (2003) reported that the relations between EC and ToM seem stronger in young children (i.e., 5 - 6-years-old) than in older children, probably due to the ceiling effects in ToM measures. Thus, investigating links between children’s ADHD symptoms and their performance in age-appropriate ToM task may elucidate this matter.

Children with ODD and CD symptoms also show deficiencies in social functioning (pervasive rule violations, lack of empathy, disregard for feelings of others, Happe & Frith, 1996; Sharp, 2008). A few studies have examined whether children with symptoms of CD are indeed socio-cognitively impaired. Although studies have reported that children with CD perform worse on ToM tasks (Hughes, Dunn & White, 1998; Sharp, 2008), and reportedly engage in fewer social behaviors that require ToM (e.g., initiating conversation of interest to others; Happe & Frith, 1996) than their typical peers, mixed findings have been reported as to whether their ToM performance is associated with the symptoms of CD (Hughes, White, Sharpen & Dunn, 2000; Sharp, 2008). Hughes and colleagues (1998, 2000) have suggested that their poor ToM performance and social problems may both be due to their low EC abilities, at least in young children. As in the ADHD literature, a question has been raised regarding the adequacy of ToM tasks to capture the variability of children's ToM performance (Sharp, 2008).

Most of the (limited) extant research has focused on links between externalizing problems and ToM. Almost nothing is known about internalizing symptoms (i.e., depressed and anxious mood) and ToM, especially in school-aged children. Given that social and interpersonal difficulties are among the most salient features in these internalizing disorders (Joiner, 2002 for review), whether and if socio-cognitive skills are involved in internalizing symptoms is a noteworthy topic of research. At this point, all evidence about links between ToM and depression or anxiety comes from the literature on adults (e.g., Harkness, Sabbagh, Jacobson, Chowdrey, & Chen, 2005; Inoue, Yamada & Kanba, 2004; Samson, Lackner, Weiss, & Papousek, 2012). Practically nothing is known about ToM skills in children with internalizing problems. Thus, research on whether ToM deficits are linked to children's internalizing symptoms will address a substantial gap.

METHOD

Participants

Family Study participants. Two-parent families of infants (N = 102) volunteered for a longitudinal study in response to ads and fliers distributed in the community. A family was accepted if the parents were living together and both wished to participate (and to speak English during sessions), the infant was their biological child, normally developing, and free of major birth complications or health problems, and the family had no plans to move in the next five years. The parents ranged in education (25 % of mothers and 30% of fathers having no more than high school education, and 21% of mothers and 20% of father having post-graduate education) and annual income (25% made less than $40,000, and 49% made over $60,000). Ninety percent of mothers and 84% of fathers were White, 3% and 8% Hispanic, 2% and 3% were African American, 1% and 3% were Asian, 1% of mothers were Pacific Islanders, and 3% and 2% were "other" non-White. In 20% of families, one or both parents were non-White.

The study was approved by University of Iowa IRB; parents completed informed consent, and children (at 100 months) completed assent. Of note, it was revealed that one Family Study child was diagnosed with ASD over the course of the study. Therefore, the data from this participant were excluded from the data analyses concerning typically developing children. The participant's data were also excluded from the analyses concerning children with ASD because he did not meet the eligibility criteria to participate in that portion of the study.

ASD participants. Children with the autism spectrum disorders, closely age-matched with the Family Study children, participated with their primary caregiver(s) (N = 7 with mothers, N = 4 with fathers; thus, 4 with both mothers and fathers). Families responded to the University-wide research E-mails, fliers, and mailing lists through local autism support groups. A family was accepted if the child : (1) had been diagnosed with autism spectrum disorders according to the Diagnostic and Statistical Manual of Mental Disorder-V (DSM-V) (American Psychiatric Association, 2013) criteria, or with autistic disorder, Asperger's syndrome, or pervasive developmental disorder, not otherwise specified according to the DSM-IV-TR (American Psyciatric Association, 2000) criteria, (2) demonstrated intact cognitive ability via past medical records including intellectual functioning or educational assessments, or classroom placement (i.e., inclusion in the regular classroom activities for the majority of his or her time at school), (3) demonstrated relatively intact expressive speech skills (no current evidence of speech impairment, no speech services, or past medical records suggesting average- range of speech/language ability, and (4) was aged between 8 and 11 (in Family Study, the new ToM task was administered when children were, on average, 10 years old, M =123 months, SD = 2.7; there were no significant correlations between children's ToM performance and their age, from 8 to 11 years, and thus we decided on 8-11 age range for ASD children).

Mothers answered basic demographic questions. Four mothers had an advanced or professional degree, two had a bachelor's degree, and one had a high school diploma or GED. Three mothers worked full time, three worked part time, and one did not work outside the home. Four mothers reported household income as more than $70,000, two -- between $40,000 and $60,000, and one did not report income. The study was approved by University of Iowa IRB. Parents completed informed consent, and children completed assent.

Procedure

Family Study. Data were drawn from assessments at 15 months (*N* = 100, 51 girls), 25 months (*N* = 99, 50 girls), 38 months (*N* = 99, 50 girls), 52 months (*N* = 98, 49 girls), 67 months (*N* = 91, 45 girls), 80 months (*N* = 89, 43 girls), 100 months (*N* = 86, 41 girls), and 123 months (*N* = 81, 37 girls).

One- and- a- half-hour to three-hour sessions were conducted in the home and laboratory by female experimenters. There were two laboratory sessions, one with each parent, at 15, 25, 52, 67, 80, and 123 months, one laboratory session at 100 months, and one home and one laboratory session, with each parent taking part in half of each, at 38 months. All sessions were videotaped for future coding. Observations in the laboratory took place in a naturalistic living room and a sparsely furnished play room.

Data on children's ToM skills were collected at three different assessments, using traditional FBTs at 52 and 67 months, and the new interactive ToM task at 123 months (age 10). Behavioral data on parent-child relationship (MRO, Kochanska, 1997) were obtained at 15, 25, 38, 52, 67, and 80 months; and behavioral data on effortful control were collected at 38, 52, 67, and 80 months. Mother- and father-reported data on children's peer competence were obtained at 67, 80, and 100 months. Mother-, father-, teacher-, and self-reported data on children's clinical symptoms were obtained at 123 months. The overview of data collection is presented in Table 1.

Multiple coding teams used at least 20% of cases for reliability and frequently "realigned" to prevent drift. Inter-rater reliabilities and scale reliabilities regarding continuous variables were established using alphas, or intraclass correlations (ICCs) according to the best practices over the last 15 years. Note that approaches regarding continuous variables using alphas or ICCs are essentially equivalent (Bravo & Potvin, 1991; Shrout & Fleiss, 1979). Inter-rater reliabilities regarding categorical variables were established using (weighted) kappas.

Procedure for ASD participants. Participants had an option either to visit the laboratory or to have an in-home session (20-25 minutes), with the latter chosen by two families. All sessions were conducted by the same female experimenter and videotaped for future coding. One coder, blind to the children's diagnoses, was trained by me and coded all cases.

Measures

Theory of mind, 52, 67 months, and age 10.

False-belief tasks, 52, 67 months. Children's understanding of a false-belief was assessed at two assessments, at 52 and 67 months. Four to five FBTs were administered at 52 and 67 months, respectively. Three different types of FBTs were used: (1) first-order FBT, one unexpected contents task (Hogrefe, Wimmer & Perner, 1986), and one unexpected location-change task (Wimmer & Perner, 1983); (2) belief-desire reasoning task (Harris, Johnson, Hutton, Andrews, & Cooke, 1989); and (3) second-order FBT (Baron-Cohen, 1989; Perner & Wimmer, 1985). All tasks employed pictures, puppets, and props to help present the task contexts. Because all tasks were administered in identical manner, but only differed in content across two assessments, detailed descriptions will be presented only for the assessment at 52 months.

First-order FBTs. In unexpected contents tasks at 52 months, the child saw a clearly identifiable object (e.g., a Band-Aid box), and shortly, the child found out that the box contains unexpected contents (e.g., crayons). Then the box was closed, and the child was asked a false-belief question, "Before you looked inside, what did you think was in the box?" Next, a child was introduced to a naïve puppet who had not yet seen the contents, and was asked what the puppet thought was in the box, another false-belief question. To pass the false-belief question, the child had to pass all the control questions (e.g., “What is really in the box?”). At 67 months, the unexpected contents task was administered using different props.

In unexpected location-change task at 52 months, colored drawing pictures were used to introduce a story-like of task. In the story, the first character placed an object in one place, but unbeknown to the first character, the second character transferred the object to a different place. Then, the first character returned to look for the object. Here, the child was asked a false-belief question: Where will the first character look for his item?. To pass the false-belief question, the child had to pass all the control questions (e.g., “Where is the object really?”). At 67 months, the unexpected location-change task was administered using a different story.

Belief-desire reasoning task. At 52 months, the child was introduced to two puppets, Chris-the-Crocodile who liked to play tricks, and his friend Leo-the-Lion. The child was told that Leo’s favorite drink was Coke, and his least favorite drink was milk. Then, the child watched Chris played a trick on Leo while Leo was gone for a walk: Chris poured out the Coke, and instead he poured in some milk. Then, the child saw that Leo came back from his walk, and he saw the Coke can on the table. The child was asked critical questions to predict Leo’s false-belief: “When Leo first comes back from his walk and sees the can, what does Leo think is in the can?”; to identify Leo’s emotional state from his false-belief, “Does he feel happy or unhappy?”; and to justify Leo’s emotional state, “Why does he feel happy/unhappy?” To pass the critical questions, the child had to pass all the control questions (e.g., “How does Leo feel after he's had a drink—happy or not happy?”). At 67 months, the belief-desire reasoning task was administered using a different story.

Second-order FBT. Colored pictures were used to help explain the contexts of task. At 52 months, the child was told a story in which the first character hid an object from the second character, but unbeknown to the first character, the second character saw where the first character put the object. Then, the two characters returned to the area in which the object was. The child was asked two critical questions, a second-order belief question of where the first character thought the second character would look for the object, and a justification question of why the first character thought that. To pass the critical questions, the child had to pass all the control questions (e.g., “Where is the chocolate really?”). At 67 months, two second-order FBTs were administered using different stories

Coding. Coders assigned scores of either 0 (incorrect answer) or 1 (correct answer) to each question asked per story. Children were awarded with points per task if they responded correctly to both critical test question (e.g., false-belief question) and control questions. Reliability, kappas, alphas and ICC were 1.00 across the assessments.

Data aggregation. Total ToM scores at respective assessments were the sums of scores across tasks, with a possible range of 0 to 8 at 52 months, and 0 to 11 at 67 months, with higher scores representing better ToM skills. In addition, an overall early ToM skills score was also created by standardizing the ToM scores at 52 and 67 months and aggregating them (r = .56, p < .001).

Object assembly (the new interactive ToM task), age 10.

Observed context. The new interactive ToM task was administered in a paradigm adapted from Kovan et al. (2009) to the Family Study participants and participants with ASD. The child verbally guided the blindfolded parent in assembling a puzzle. Successful guidance required that the child consider the parent's perspective and the fact that the parent lacked the perceptual information the child had.

I had designed a new behavioral coding system to capture child ToM skills during this interactive task (5 min with each parent, thus 10 min with both parents total). In Family Study, children engaged in the task with both parents, and children in the ASD sample engaged in the task with both parents when both participated, or with one, if only one was available.

Coding. I and one independent coder reliably coded two variables. Children's *perspective-taking (PT) errors* were all occurrences of the child's directions that ignored the parent's perspective, e.g., "put the red one there," "turn this way", coded using microscopic codes for every 5-sec segment (thus, up to 60 segments with each parent). Total scores of children's PT errors were the total occurrences of children's errors divided by the total number of segments.

Children's *perspective-taking proficiency* in guiding the parent to complete the task was also rated, as one rating for the entire task. The child's PT proficiency ranged from 1 to 5 (e.g., 1, the child's quality of ToM skills is very poor, the child's instruction interferes with the task completion, and the child is oblivious to the fact that child's strategy is inefficient and is the cause of the dyad's struggle; 3, the child's quality of ToM skills is acceptable, the child is able to sense that his or her strategy is not working and fixes his or her approach accordingly, and the child is able to self-correct mistakes; 5, the child's quality of ToM skills is excellent, child is exceptionally careful in delivering his or her directions and monitors his or her own behavior carefully to avoid mistakes due to differences in perspectives).

Coders started coding when experimenter left the room and stopped coding after 5 minutes. Detailed descriptions of a coding scheme can be found in Appendix. In the Family Study, inter-rater reliability, ICCs ranged from .98 to 1.0 for microscopic PT errors, and weighted kappa was .76 for overall rating of child PT proficiency.

Data aggregation. Several ways of data aggregation were used to create several types of scores, appropriate for different analytic questions. In the analyses in which ToM skills are compared between the Family Study participants and the participants with ASD, four scores were used: (1) PT errors score with mother, (2) PT proficiency rating score with mother, (3) PT errors score with father, and (4) PT proficiency rating score with father. The scores were not combined further in those analyses, because one of the goals for the first aim was to examine whether children from the two samples performed differently across the different measures (i.e., PT errors and proficiency rating scores). PT errors score was reverse-computed to represent the amount of time that participants did *not* make perspective-taking errors.

In other analyses, the scores of PT errors and PT proficiency were aggregated to represent more robustly the *child's perspective taking skills*. Three scores were created: (1) the score for the child's PT skills with the mother, (2) the score for the child's PT skills with the father, and (3) the child's overall PT skills deployed with both parents.

First, PT errors score was reverse-computed to represent the amount of time that participants did *not* make perspective-taking errors. Children's scores of (reversed) PT errors and proficiency were correlated, for the task with the mother ($r = .51$, $p < .001$) and for the task with the father ($r = .52$, $p < .001$), and thus, they were standardized and averaged to create the PT skills scores with each parent. Those two parent-specific scores, the child's PT skills score with the mother and that with the father, were used as the dependent variables in the analyses of the second aim, where children's ToM was considered in the context of the relationships with mothers and fathers.

The final score created was an overall child PT skills score; this score was created by aggregating across the tasks with both parents (the composite of the PT errors score and PT proficiency score with mothers and those with fathers). Cronbach's alpha for that score was .64. This overall PT skills score with both parents combined was used in the correlation analyses for the first aim in which heterotypic continuity was examined between early FBTs and the new ToM measure, and in the analyses that involved the third aim (links with clinical symptoms). Note that I also conducted additional analyses for the second aim (above), using this overall measure. Those are available upon request.

Parent-child relationship, 15 - 80 months.

Observed contexts. Positive, mutually responsive parenting (MRO) for each parent-child dyad was observed in naturalistic, carefully scripted contexts that encompassed play, chores, preparation of snacks, snack time, parent busy with questionnaires, free time, a craft project, etc. The numbers of coded contexts and total time with each parent are in parentheses: at 15 months (6, 47 min), at 25 months (6, 47 min), at 38 months (9, 77 min), at 52 months (6, 65 min), at 67 months (6, 60 min), and at 80 months (6, 60 min). Overall, each mother- and father-child dyad was observed in 39 contexts (356 min), and each child, in 78 contexts (712 min) from 15 to 80 months.

Coding. Coders assigned one overall MRO rating for each observed context, ranging from 1 (very untrue of the dyad) to 5 (very true of the dyad). That rating integrated four dyadic dimensions described below.

Coordinated routines. Low: The dyad has no routines, or routines are choppy, rough, and conflict producing. High: The dyad easily settles into comfortable, coordinated routines.

Harmonious communication. Low: The dyad communicates very little or not at all. High: Communication is smooth, connected, and harmonious.

Mutual cooperation. Low: The dyad is unable to cooperate, struggles escalate. High: The parent and child have a willing, receptive stance toward each other, with subtle cues sufficient for cooperation.

Emotional ambience. Low: Negative ambience, bouts of negative affect. High: The parent and child enjoy each other, ambience is positive and warm, with bouts of joy, good humor, and affection.

The conventions specified how to integrate the dimensions to arrive at the overall score for each context. Reliability, weighted kappas, ranged from .72 to .83.

Data aggregation. At each assessment, the scores across all observed contexts cohered substantially. Cronbach's alphas were as follows (mother-child dyad first, father-child dyad second): at 15 months, .86 and .80, at 25 months, .80 and .81, at 38 months, .71 and .77, at 52 months, .77 and .73, at 67 months, .80 and .75, and at 80 months, .72 and .76. They were averaged across all contexts into one score for each parent at each time. MRO scores for each parent at each time were again aggregated to represent mother-child dyad MRO across time, father-child dyad MRO across time, and overall MRO across time. Cronbach's alphas for mother-child dyad and father-child dyad were .82 and .78, respectively.

Peer competence, 67 – 100 months.

MacArthur's Health Behavior Questionnaire (HBQ; Boyce et al., 2002; Essex et al., 2002) was administered to each parent at 67, 80, and 100 months to assess children's peer functioning. Parents rated items from 1 (not at all like) to 4 (very much like), or from 1 (rarely applies) to 3 (certainly applies) depending on the items. The following scales targeting peer competence were selected: Peer Acceptance/Rejection, Bullied by Peers, Prosocial Behavior, Overt Aggression, Relational Aggression, Social Inhibition, and Withdrawn with Peers.

Data aggregation. Cronbach's alphas for the specific scales ranged from .53 to .91 at 67 months, .64 to .89 at 80 month, and .49 to .90 at 100 months. The scales that indicate peer incompetence, Bullied by Peers, Overt Aggression, Relational Aggression, Social Inhibition, and Withdrawn with Peers, were reversed so that higher scores reflected higher peer competence. The scales were aggregated to represent mother-reported and father-reported peer competence scores at each time. Cronbach's alphas for the aggregated scale for mother-reported peer competence at 67, 80, and 100 months were .62, .68, and .64, respectively, and for father-reported peer competence at 67, 80, and 100 months were .70, .70, and .68, respectively. Mother-reported and father-reported peer competence scores at 67, 80, and 100 months were correlated at all three assessments (rs = .26 to .71, ps < .05) except for one correlation between mother-reported scores at 80 months and father-reported scores at 100 months ($r = .17$, $p = .139$); and Cronbach's alpha for an overall composite across mother-reported and father-reported peer competence scores and across 67, 80, and 100 months was .83. Consequently, that composite, representing overall peer competence across times, was used.

Effortful control, 38 - 80 months.

Batteries of tasks. Extensive behavioral batteries of effortful control tasks were administered to the children across four assessments during laboratory sessions, interspersed with other activities (Kochanska, Aksan, Penny, & Doobay, 2007). The tasks captured the following five functions of effortful control abilities: delaying, motor inhibition, go-no go, lowering voice and effortful attention. Not all functions were assessed at all ages; some tasks were repeated, and new tasks were added as permitted by the child's increasing maturity. Tasks were presented as a game rather than a directive, prohibition, or request, and the child was praised regardless of performance.

Battery at 38 months. There were ten tasks. Three *delaying* tasks involved waiting to reaching for M&M (Snack Delay), deliberately choosing a prize from a box filled with small toys (Dinky Toy), and waiting to unwrap a gift (Gift Wrap, Gift Bow). Three *motor inhibition* tasks called for slowing motor activity (drawing; walking a 6-ft line; and guiding a toy turtle slowly along a curved path to the barn). One *go-no go* task called for suppressing a response to one type of signal and initiating a response to another (a turn-taking game, (Tower)). *Lowering voice* was tapped in a whispering task. Two *effortful attention* (Stroop-like) tasks required ignoring a dominant perceptual feature of a stimulus for the sake of a subdominant feature, Day-Night (point to the day-sky square when experimenter says "night," and point to the night-sky square when experimenter says "day") and Snow-Grass (point to the white square when experimenter says "grass," and point to the green square when experimenter says "snow"; Carlson & Moses, 2001).

Battery at 52 months. There were 14 tasks. Five *delaying* tasks involved waiting to reach for an M&M, waiting to chew an M&M placed on the child's tongue, deliberately choosing a prize from a box filled with toys, and waiting to unwrap two gifts. Three *motor inhibition* tasks called for slowing fine and gross motor activity (drawing lines, walking a line, and moving a toy toward a play barn). Three *go-no go* tasks involved suppressing/initiating activity to signal (to red and green signs, and to commands given by a toy bird and toy dragon), and turn-taking while building a block tower. *Lowering voice* was assessed in a whispering task. Two *effortful attention* (Stroop-like) tasks were Day-Night and Snow-Grass.

Battery at 67 months. There were 10 tasks. Two *delaying* tasks involved waiting to waiting to unwrap two gifts. Four *motor inhibition* tasks called for slowing fine and gross motor activity (drawing circle, drawing star, walking a line, and moving a toy car following the trail (Bus-Truck)). Two *go-no go* tasks involved suppressing/initiating activity to signal (to red and green signs, and to commands given by "Simon Says.") Two *effortful attention* (Stroop-like) tasks were Day-Night/ Snow-Grass (administered simultaneously), and a tapping game (tap a stick on the table once when experimenter taps twice, and tap twice when experimenter taps once).

Battery at 80 months. There were 10 tasks. Two *delaying* tasks involved waiting to waiting to unwrap two gifts. Four *motor inhibition* tasks called for slowing fine and gross motor activity (drawing flower, drawing sun, walking a line, and moving a toy car following the trail). Two *go-no go* tasks involved suppressing/initiating activity to signal (to red and green signs, and to commands given by "Simon Says.") Two *effortful attention* (Stroop-like) tasks were a tapping game, and animal Stroop (name animal's body while ignoring the non-matching head; Wright, Waterman, Prescott, & Murdoch-Eaton, 2003).

Coding. The codes were strongly behaviorally based and required little inference. For each trial, higher score reflected better capacity for self-regulation. Reliability of coding was extremely high across all ages and across many teams. Reliabilities, kappas, ranged from .71 to 1, and alphas ranged from .81 to 1. The scores were averaged across trials, where applicable.

Data aggregation. The scores were averaged across trials, where applicable. The individual task scores were then standardized and aggregated into *composites* at 38, 52, 67 and 80 months (Cronbach's alphas, .64, .71, .73, and .66; means and standard deviations, .00 and 1 ; .00 and 1; .08 and .87; .10 and .72, respectively). The composites at 38, 52, 67 and 80 months were standardized and aggregated into one overall EC composite (Cronbach's alpha = .83). That score was used as a covariate in the analyses linking ToM and clinical symptoms, ADHD and OS symptoms, specifically.

Clinical symptoms, age 10.

Parent-reported clinical symptoms. Both parents completed Child Symptom Inventory-4 (CSI-4; Gadow & Sprafkin, 2002; Gadow, Sprafkin, & Nolan, 2001; Sprafkin, Gadow, Salisbury, Schneider, & Loney, 2002) at age 10. CSI-4 is a well-established instrument that corresponds to DSM-IV (American Psychiatric Association, 2000). Gadow and Sprafkin (2002) reported that Cronbach's αs for the specific scales in the parent checklist ranged from .74 to .94, and test-retest correlations over 4 weeks ranged from .46 to .87 (all *p*s <.001). Each item was rated from 0 (never) to 3 (very often), with higher score representing more severe symptoms (severity scores).

Data aggregation. Items were summed as recommended by the CSI-4 manual to represent the symptom severity of ADHD. One ADHD score was created for the mother and one for the father. Cronbach's αs for the ADHD scale were .91 and .92 for mother- and father-reported scores, respectively. Mother- and father-reported ADHD scores were significantly correlated, $r(75) = .72$, $p < .001$.

Second, items were summed to represent the symptom severity of CD and of ODD; then, the two scale scores were summed to represent the severity of *oppositional symptoms (OS)* for each parent. Cronbach's αs for the OS scale were .86 and .87 for mother- and father-reported scores, respectively. Mother- and father-reported OS scores were significantly correlated, $r(75) = .54$, $p < .001$

Third, for each parent, four scores were added to create an *internalizing symptoms (IS)* score: depression, generalized anxiety disorder, specific phobia, and separation anxiety, all based on the symptom severity scoring. One internalizing symptoms score was created for the mother and one for the father at each assessment. Cronbach's αs for the IS scale were .85 and .87 for mother- and father-reported scores, respectively. Mother- and father-reported IS scores were significantly correlated, $r(75) = .47$, $p < .001$.

Teacher-reported clinical symptoms. Teachers completed the teacher version of Child Symptom Inventory-4 (CSI-4; Gadow & Sprafkin, 2002; Gadow, Sprafkin, Salisbury, Schneider & Loney, 2004) at age 10. Gadow and Sprafkin (2002) reported that Cronbach's αs for the specific scales in the teacher checklist ranged from .70 to .96, and test-retest correlations over 2 weeks ranged from .47 to .88. The format of a teacher checklist was identical to that of a parent checklist. As in the parent-reported clinical symptoms, the same scales were created for the teacher-reported symptoms as well: ADHD, OS, and IS score. Cronbach's αs for the ADHD, OS, and IS scores ranged from .81 to .96. The separation anxiety scale was not included in the teacher checklist due to the nature of symptoms (e.g., "Worries that parents will be hurt or leave home and not come back). Thus, the IS scale consisted of depression, generalized anxiety disorder, and specific phobia.

Children's self-reported clinical symptoms. During the laboratory visit, the experimenter, having established good rapport with the child, administered the interactive, computerized version of Dominic-R (Shojaei et al., 2009; Valla, Bergeron, & Smolla, 2000). Dominic- R is an approximately 30-min, vignette-based, visual-auditory clinical interview instrument, appropriate for 6- to 11-year olds. Robust psychometric qualities have been established in a number of past studies, many with large samples. The vignettes depict specific clinical symptoms. Based on the child endorsing the vignettes as descriptive of him or her (yes or no), the interview produces clinical-symptom scores designed to map onto DSM-IV disorders. I used the scores for ADHD, OS (CD and ODD), and IS (depression, generalized anxiety disorder, specific phobia, and separation anxiety). The ADHD score was generated by the Dominic-R computer software program, and thus the scale reliability in the current sample is not available. The OS scores cohered, $r\ (78) = .52$, $p < .001$ and so did the IS scores, $\alpha = .83$.

RESULTS

Overview

Data were screened to ensure that assumptions were not violated. Normality was checked and transformations were applied for several variables. All structural equation modeling (SEM) analyses were conducted with Mplus (Muthen & Muthen, 2010). Missing data were addressed via maximum likelihood estimation.

When analyses involved latent variables, a series of confirmatory factor analyses were conducted first to estimate the fit of measurement models. I considered five different fit indices in evaluating the fit of SEM models: the overall model χ^2, the root mean square error of approximation (RMSEA), the comparative fit index (CFI), the Tucker-Lewis index(TLI), and the standardized root mean square residual (SRMR). Conventional guidelines suggest that a fit is acceptable if the CFI is .90 or greater, and RMSEA and SRMR are .10 or less (e.g., Watson, 2001). Researchers also suggest more stringent standards: the CFI and TLI $< .95$ are poor fit (Bentler, 1990; Hu & Bentler, 1999); the RMSEA $\leq .05$ is good fit, and between .05 and .08 is "OK" fit (Steiger & Lind, 1980); the SRMR $\leq .08$ is acceptable (Hu & Bentler, 1999). I used both sets of guidelines in evaluating the results.

To test indirect effects in mediation models, a bootstrap approach (Shrout & Bolger, 2002) was implemented. The bootstrap approach is a nonparametric resampling method proposed as an alternative to Sobel's (1982) test for estimating indirect effects. The bootstrap method has advantage over the traditional Sobel test including (1) sampling distributions of indirect effects are estimated and confidence intervals of estimates are provided; and (2) bootstrapping can be applied in a small to moderate sized sample and tends to be more powerful

(Shrout & Bolger, 2002). I performed a bias-corrected bootstrap with 5000 resamples to derive the 95% confidence intervals (CI) for the indirect effects of parent-child relationship quality on children's PT skills through EC and peer competence, and of children's early ToM skills on the clinical symptoms through children's PT skills in middle childhood.

Descriptive statistics (Table 2) and correlations among measures (available upon request) were examined. I conducted correlation analyses, Mann-Whitney U-tests, and the modified t-tests to validate my new ToM measure in both samples. I conducted SEM analyses to examine interpersonal and intrapersonal effects on individual differences in ToM skills in the Family Study sample. Lastly, I conducted SEM analyses to investigate associations between ToM skills and clinical symptoms in the Family Study sample.

Descriptive Statistics

Means and standard deviations for all variables included in the analyses are presented in Table 2. However, effortful control (EC) variables that were involved in confirmatory factor analysis (CFA) were not included in Table 2 because there were simply too many EC variables at each assessment and stages of aggregations before the composites were created. No significant effects were identified between ToM scores and demographic variables such as age, gender, or family income in the Family Study data.

Theory of Mind in Middle Childhood

One way to validate my new ToM measure was to examine heterotypic continuity of ToM skills over time. I examined the correlations among the Family Study children's performance on FBT tasks at 52 and 67 months and on the new interactive, non-FBT task at age 10. Recall that for these analyses, I used the child's overall PT skills score with both parents combined. The score was a composite score of PT errors and proficiency rating scores with both mothers and fathers, reflecting the child's overall perspective-taking skill.

The magnitude of correlation between FBT composite score at 52 months and that at 67 months indicated a robust relation, as expected ($r = .56$, $p < .001$). The correlations between FBT tasks scores at 52 and 67 months, and PT skills scores at age 10 were still significant ($r = .23$, $p = .039$; $r = .35$, $p = .002$, respectively), albeit reduced in their magnitudes.

Another way to validate the new ToM measure was to establish initial criterion validity by showing relatively poorer performance by children with ASD compared to the Family Study children on the new measure. To match age and gender of ASD participants, the Family Study boys served as a comparison sample. I used two different statistical approaches to address the question of criterion validity. Recall that for these analyses, I used four scores: 1) PT errors score with mother, 2) PT proficiency rating score with mother, 3) PT errors score with father, and 4) PT proficiency rating score with father. First, the four scores were rank ordered and a Mann-Whitney U-test was used to compare the ranks for the Family Study boys ($N = 42$) versus the participants with ASD ($N = 7$) for the scores with mothers, and the Family Study boys ($N = 39$) versus the participants with ASD ($N = 4$) for the scores with fathers. The results failed to support significant differences between the two samples, U = 125, 51, 131.50, and 72, p = ns.

Second, given the small size of the comparison sample ($N = 42$) and the novelty of the measure, I used the modified t-test procedure to gauge the statistical differences in participants' performance on the new ToM measure between the Family Study boys and participants with ASD (Crawford & Garthwaite, 2002; Crawford & Howell ,1998). The computer program available from John Crawford's website, http://homepages.abdn.ac.uk/j.crawford/pages/dept/psychom.htm, was used to obtain the modified t-statistics and 95% CI associated with estimated percentile ranks. Note that PT errors scores were reverse-computed to represent the amount of time that participants did *not* make perspective-taking errors. This was done so for the ease of graphic presentations of the PT proficiency rating and errors scores for participants with ASD (Figure 1 and 2). Thus, the higher the scores, the better the child's performance.

Table 3 presents the findings comparing ASD participants' scores to their neurotypical age- and gender-matched counterparts. The first three columns present the mean, *SD* and 95% CI for the sample mean for the comparison sample. The latter columns present each ASD participant's PT errors scores and proficiency rating relative to the respective mean scores of the comparison sample. Each column for the ASD sample presents an individual's score and its associated modified t-values. Figures 1 and 2 present the 95% CIs on each ASD participant's PT errors scores and proficiency rating with mothers and fathers, respectively.

Table 3 shows that participants with ASD performed no differently from their comparison sample in general on the PT errors; however, interestingly, two participants performed significantly better than their comparison sample, one with mother and the other with father. Figures 1 and 2 show the tight CIs around the estimates (#07 in Figure 1, #02 in Figure 2) and suggest that their performance was superior to that of their neurotypical peers. Three of the seven participants in the ASD sample performed more poorly compared to their comparison sample on the PT proficiency rating with mothers. The statistical difference was at the *p*-value of .10 using a one-tail t-test. Figure 1 shows tight CIs around the estimates. The remaining four participants' performance on the PT proficiency rating with mothers was not statistically different from their comparison sample's performance. Figure 1 shows wide CIs around the estimates and suggest that the four participants performed within the average range of their peers. In terms of PT proficiency rating with fathers, two of the four participants whose fathers participated in the session performed significantly worse than their comparison sample at the level of $p = .10$. Figure 2 shows tight CIs around the estimates. The remaining two participants' performance was not statistically different from their comparison sample's average performance. Figure 2 shows wide CIs around the estimates and suggests that the two participants performed similar to their peers.

Predictors of Individual Differences in ToM Development

Interpersonal factors and ToM.

I conducted SEM analyses with observed variables to investigate the relative influence of interpersonal factors on children's ToM skills. Recall that for these analyses, I used the separate scores, (1) the score for the child's PT skills with mother, (2) the score for the child's PT skills with father. The scores were composite scores of PT errors and proficiency rating with mothers, and of those with fathers. The results of the SEM analyses are reported in Figure 3. The model was exactly identified and thus no fit statistics were available. Whether mother-child MRO and father-child MRO and the child's peer competence uniquely predicted either children's PT skills with mothers or with fathers was examined (Figure 3). Father-child dyad MRO marginally predicted children's PT skills with mothers (p =.061). Peer competence significantly predicted children's PT skills with mothers (p =.023). Approximately, 12.5 % of the variance in children's PT skills with mothers was explained by the model, compared to only .6% of the variance in children's PT skills with fathers.

Intrapersonal factors and ToM.

Aggregating EC tasks scores. EC tasks were conceptually categorized as "hot" and "cool" tasks. Specifically, the *delaying* tasks assessed at 38 through 80 months were categorized as "hot" tasks because they tap children's self-regulatory ability in affectively charged tasks (e.g., waiting for snack while the snack is in sight). The other three functions of EC (i.e., *motor inhibition, go-no go,* and *effortful attention*) were categorized as "cool" tasks as they tap a more abstract form of self-regulation. Then, I examined correlations among the scores for "hot" and "cool" tasks respectively and aggregated the scores when appropriate (i.e., when scores were correlated). Because the aim was to create general and trait-like scores of "hot" and "cool" EC, aggregation was performed across assessments at different times. Scores that did not cohere well within a given function were dropped at this stage.

Confirmatory factor analysis estimating the EC measurement model. I conducted a series of CFAs to estimate the fit of measurement models for EC functions. I first estimated separate measurement models for each latent variable—delay function (DF), motor inhibition (MI), and go-no go and effortful attention (GN_EA)—to ensure adequate fit before testing a comprehensive three-factor EC model. The fit indices for the three-factor EC model in which DF, MI, and GN_EA were modeled as latent variables suggested that the model fit was overall acceptable (Table 4).

Structural equation models estimating the effects of EC on ToM skills. The three-factor EC model was applied to estimate the effects of EC in toddler and preschool years on children's ToM skills in middle childhood (see Figure 4.). Children's PT skills at age 10 were modeled as observed endogenous variables and the three previously tested latent EC variables—DF, MI, and

GN_EA—were examined as predictors of PT skills. Recall that for these analyses, I used the separate scores, (1) the score for the child's PT skills with mother and (2) the score for the child's PT skills with father (the composites of PT errors and proficiency rating with mother, and of those with father). Fit indices are reported in Table 4. The chi-square test was significant (p = .047) and TLI (.887) was slightly outside the desired range for the model; however, considered together, the global fit indices suggest adequate fit.

MI significantly predicted children's PT skills with mothers (p = .002), controlling for DF and GN_EA. None of the other EC functions uniquely predicted children's PT skills with fathers. The model explained 17% of the variance in children's PT skills with mothers and 8.3% in skills with fathers.

Interpersonal and intrapersonal factors and ToM.

In the final, comprehensive analyses, my goal was to examine the two sets of predictors of ToM at age 10 – the interpersonal and intrapersonal factors – simultaneously, in a conceptually informed overall model. I conducted SEM analyses to investigate direct effects of the interpersonal factors (i.e., mother-child and father-child dyad MRO at 15 – 80 months, and peer competence 67 – 80 months) and intrapersonal factors (i.e., EC skills at 38 – 80 months) on ToM skills at age 10, and indirect effects of mother-child and father-child MRO through the hypothesized mediators (i.e., EC skills and peer competence) on ToM skills at age 10. Recall that for these analyses, I used the separate scores, (1) the score for the child's PT skills with mother and (2) the score for the child's PT skills with father (the composites of PT errors and proficiency rating with mothers, and of those with fathers).

The mediational SEM model in this section was informed by robust existing research evidence and specified based on that research. Consequently, early parent-child relationship was modeled as influencing the subsequent development of EC (e.g., Eisenberg et al., 2005; Kochanska, Murray, & Harlan, 2000; Kochanska & Kim, 2014; Power & Chapieski, 1986), and ToM (e.g., Meins et al., 1998; Symons & Clark, 2000); and EC was modeled as predicting ToM (e.g., Carlson et al., 2004, Sabbagh et al., 2006).

I first ran a set of preliminary analyses to establish that the hypothesized mediators were significantly associated with children's PT skills. This was done to identify more parsimonious models. As a result, DF was omitted because it was not significantly associated with children's PT skills with either parent in the preliminary analyses.

The final model tested whether mother-child and father-child dyad MRO, peer competence, and two latent EC variables (MI, GN_EA) uniquely predicted children's PT skills with mothers or with fathers, and whether EC variables and peer competence represent mechanisms through which MRO influences PT skills (Figure 5). Fit indices are reported in Table 4. TLI was outside the desired range (.885). However, other indices were adequate, and thus the model was retained. Peer competence marginally predicted PT skills with mothers (p = .059). GN_EA marginally predicted PT skills with fathers (p = .056). Interestingly father-child MRO negatively predicted children's PT skills with fathers at a marginal level (p = .070). None of the tested indirect effects were present as indicated by the bias-corrected bootstrapped CIs. The amounts of variance explained by the model in Figure 5 were 21% and 16.3% for children's PT skills with mothers and fathers, respectively.

ToM and Clinical Symptoms

I conducted SEM analyses to investigate whether ToM scores are associated with clinical symptom scores (ADHD, OS, and IS). An SEM model was estimated to examine whether ToM skills in kindergarten years (the observed variable aggregating ToM skills in FBTs across 52 and 67 months) or in middle childhood (the observed variable aggregating PT skills across mothers and fathers) would be uniquely associated with children's clinical symptoms at age 10 (a latent variable of mother, father, teacher, and child reports). Recall that for these analyses, I used the child's overall PT skills score at age 10 (middle childhood) with both parents combined (the composite of PT errors and proficiency rating scores across mothers and fathers, reflecting the child's overall perspective-taking skill). Based on the previous findings that EC was associated with both symptoms of ADHD, and with OS (Fahie & Symons, 2003; Hughes et al., 1998, 2000), the EC composite score (see p. 36) was included as a covariate in models predicting ADHD and OS.

ToM and ADHD. An ADHD measurement model in which mother, father, teacher, and child report reflect an ADHD latent variable was estimated first. The fit indices for the model suggested that the model fit was excellent (Table 5). Next, the larger structural model depicted in Figure 6 was tested. The fit indices of the SEM model are presented in Table 5 and suggest adequate fit to the data. Early ToM skills during kindergarten years significantly predicted children's ADHD symptoms at age 10 (p = .039), controlling for concurrent ToM skills, also at age 10. In contrast, ToM skills at age 10 were not associated with ADHD symptoms when controlling for early ToM skills. The model accounted for 8.9% of the variance in children's ADHD symptoms.

ToM and oppositional symptoms (OS). An OS measurement model in which mother, father, teacher, and child report reflect an OS latent variable was estimated. However, the fit indices for the model suggested a poor fit to the data (Table 5). Re-specification of the model resulted in dropping teacher report of OS symptoms because of its non-significant and low factor loading ($\lambda = .07$, $p = .653$). The re-specified OS measurement model in which mother, father, and child report reflect an OS latent variable was estimated. The re-specified OS model was exactly identified, and thus no fit indices were available. The component fit of the model was examined and all factor ladings were significant (and exceeded .30) in the expected directions. Next, the larger structural model depicted in Figure 7 was tested. The fit indices of the SEM model are presented in Table 5 and suggest adequate fit to the data. Early ToM skills in kindergarten years negatively predicted children's OS symptoms at age 7 ($p = .066$) at a marginal level accounting for ToM skills, also at age 10. The amount of variance in children's OS symptoms explained by the model was 5.6 %.

ToM and internalizing symptoms (IS). An IS measurement model in which mother, father, teacher, and child report reflect an IS latent variable was estimated. The fit indices suggested that the model fit was excellent (Table 5). Next, the larger structural model depicted in Figure 8 was tested, and fit indices are presented in Table 5 which suggest adequate fit to the data. ToM skills in middle childhood were uniquely associated with children's IS symptoms at age 10 ($p = .004$) controlling for early ToM skills. Further, early ToM skills indirectly and negatively predicted children's IS symptoms at age 10 through concurrent ToM skills, 95% CI [-.20, -.02]. The direct effect of early ToM skills on IS symptoms was not significant when controlling for concurrent ToM skills, suggesting it primarily influences IS symptoms via ToM skills later in development. Approximately, 13.8 % of the variance in children's IS symptoms was explained by the model.

DISCUSSION

The overarching goal of my dissertation was to examine ToM skills in middle childhood. Largely, I had three aims. First, I developed and validated a novel interactive ToM measure highlighting perspective-taking skills during a dynamically flowing social interaction. Second, I comprehensively examined interpersonal and intrapersonal factors as predictors of individual differences in ToM skills in middle childhood. Third, I examined children's ToM skills in the context of a wide range of clinical symptoms. Below, I will discuss the findings of each aim.

The First Aim: A New ToM Measure

With regard to the first aim, I used three statistical methods to validate the new measure. The results were mixed. The first method was to examine the correlations between FBT tasks at 52 and 67 months and the newly developed, non-FBT task at age 10 to establish heterotypic continuity. I hypothesized that children's performance in ToM tasks at 52 and 67 months would be modestly to moderately associated with their ToM performance at the age of 10. They were indeed significantly correlated, with the correlations medium in size, supporting conceptual coherence between the two types of measures.

The second and third method was to establish initial criterion validity, expecting that children with ASD would perform more poorly than their typically developing peers on the new measure. However, a foremost, key limitation for those comparisons was the low number of children with ASD (*N*=7), which should be kept in mind as those data are discussed. The second method, Mann-Whitney U-test did not support the hypothesis, in that there were no discernible group differences in either PT errors score or in PT proficiency rating score. The third method, the modified t-test, partially supported the hypothesis, in that some children with ASD demonstrated significantly lower quality of perspective-taking proficiency skills than their typically developing peers; however, the test failed to find children with ASD making more perspective-taking errors compared to their counterparts.

Whereas the overall mixed results of the first aim may appear to compromise the robustness of the new measure's validity, the results were still encouraging, especially regarding heterotypic continuity. For typically developing children from a community sample, the new measure – despite being entirely different in format and requirements, and administered 4 to 5 years later – successfully captured behavioral aspects of ToM skills, and it provided conceptually coherent data over time. It was an encouraging result, supporting my view of perspective-taking skills, embedded in the flow of a social interaction, and observationally coded, as reflecting a ToM skill in a general sense.

The mixed findings regarding criterion validity -- the failure to find group differences on PT scores, and individually, finding that that only some children with ASD performed worse than their typically developing peers -- were unexpected. Perhaps our coding system was not sensitive enough to capture the entire range of children's performance. The Family Study children's relatively weak performance on the new measure, specifically on the PT proficiency, is relevant, and may inform the null findings. The means for PT proficiency rating for the typically developing boys in Family Study were around 2, with both parents (M = 2.29, SD = .92 with mothers; M = 2.36, SD = .96 with fathers). The proficiency rating score of 2 indicates overall poor quality of a child's ability to take parents' perspectives during the task. The fact that typically developing children performed poorly on average may have contributed to the null findings.

As another, related possibility, the coding in our new measure might not have been appropriately calibrated for children with ASD. That is, the proficiency scale of 1 to 5 may have been too restrictive to capture the lowest range of those children's ToM skills. About a half of children with ASD received the lowest score of the proficiency rating (1). It may have been the case that those who received a 1 received that score merely because it was the lowest score on the scale. In other words, there might have been a qualitative difference in the same score of 1 used for performance of Family Study children and that of children with ASD, with the latter performing worse. If so, the range of the current scale of 1 to 5 may be the limitation. Taken together, a new, improved version, with finer distinctions among the codes, particularly at the lower end of the scale, should aim for better-distributed scores.

And finally, there were no group differences in PT errors scores and interestingly, two boys with ASD made significantly fewer PT errors compared to the Family Study children. However, it may be hasty to conclude that the results suggest comparable PT skills between the Family Study children and the participants with ASD, or even better PT skills of the two boys with ASD compared to their typically developing peers. The premise for the PT errors score was to quantitatively assess children's ability to self-correct their errors by getting feedback from their partners and to utter more correct statements. In other words, PT errors scores are the microscopically measured quantity indicator of PT skills; and it is expected to moderately correlate with the globally measured quality indictor of PT skills, PT proficiency rating. In Family Study, the two scores were moderately correlated with mothers $r(42) = -.60$, $p < .001$, and with fathers $r(39) = -.50$, $p = .001$. Therefore, the pattern of the correlations was consistent with expectations. In the ASD sample, however, the correlations between the two scales ($r(7)$ = -

.84, p = .019; $r(4)$ = -.94, p = .065, with mothers and fathers, respectively) suggested stronger associations than in the Family Study sample. This may indicate that the two scales functioned differently in the ASD sample from the Family Study sample. Of course, the reliability of correlations in the ASD sample is very limited due to the small sample size (N = 7 with mothers, N= 4 with fathers, and among those 4 with both mothers and fathers)

The Second Aim: Interpersonal and Intrapersonal Factors as Predictors of Individual Differences in ToM

The results of the analyses addressing the second aim were overall modest, and not entirely consistent. However, several significant and anticipated findings have emerged, along with several promising trends that may provide productive leads for future research. There was some evidence that, as expected, both interpersonal *and* intrapersonal factors would predict individual differences in ToM skills in middle childhood.

In the interpersonal model that examined the child's relationships with mothers, fathers, and peers, peer competence significantly predicted children's ToM skills with mothers at age 10. Peer competence uniquely predicted children's ToM skills with mothers, above and beyond mother-child relationship quality and father-child relationship quality in the interpersonal model. The results support the experiential view of ToM development, such that children's diverse social experiences continue to influence their evolving ToM skills over time as their social world expands. The results contribute to the literature by suggesting that peer relationships should be considered as a critical factor in children's social functioning beyond the influence of parents, perhaps starting in middle childhood.

In the intrapersonal model that examined "hot" and "cool" forms of EC, one of the "cool" factors, motor inhibition, significantly predicted children's ToM skills with mothers. Whereas the current literature finds the strongest association between children's performance on conflict-inhibition types of tasks and their ToM skills, specifically FBTs, in this study, it was motor inhibition that significantly predicted children's ToM skills in our new non-FBT task. In the intrapersonal factor model, motor inhibition was the only significant predictor of children's ToM skills with mothers. On the one hand, as discussed in Chapter 1, FBTs focus on assessing the conceptual aspect of ToM skills: Whether one *understands* that others will behave according to their own internal perspectives despite being in fact mistaken. Consequently, this calls for the skill assessed in conflict-inhibition types of tasks. Those tasks require that children inhibit dominant responses (e.g., their own view of the reality) and activate subdominant responses (e.g., a mistaken view of someone else). On the other hand, in the present study, the new task assessed the behavioral aspect of ToM skills: Whether children uttered more correct perspective-taking statements and fewer error-ridden statements (to achieve a high perspective-taking score). I speculate that children had to *deliberately* inhibit the utterance of erroneous statements (e.g., "move the red piece this way") and had to utter correct statements instead. The results contribute to the current literature by suggesting that another dimension of EC, the ability to slow or inhibit specifically motor activity, may be associated with ToM skills that are assessed by non-FBTs.

As for the promising leads, several associations that fell short of significance were also present. For example, father-child relationship quality *marginally* predicted children's ToM skills with mothers, but mother-child relationship quality did not emerge as a significant predictor. This is inconsistent with the previous studies, in which significant associations were found between maternal parenting characteristics and children's ToM skills (e.g., Meins et al., 2002; Ontai & Thompson, 2008). Imaginably, the models explored in the present study may have failed to find a significant contribution of either parent-child relationship quality due to their shared variance with children's ToM skills. Perhaps other statistical approaches such as examining mother-child and father-child relationship quality in separate models, or modeling parent-child relationship quality as one latent variable may have found a unique contribution of parent-child relationship quality in predicting children's ToM skills.

Although the literature has alluded to a developmental pathway (early parent-child relationship affecting children's EC skills, Eisenberg et al., 2005; Kochanska & Kim, 2014; Kochanska et al., 2000; Power & Chapieski, 1986, and in turn, children's EC skills predicting their ToM skills, Carlson et al., 2004; Sabbagh et al., 2006), no study has comprehensively examined such model. Consequently, I believe that testing a mediation model that systematically examined the associations among interpersonal factors, intrapersonal factors, and children's ToM skills in middle childhood is a contribution of this work. Unfortunately, no mediation effects, but only marginal direct effects, were found. In that comprehensive mediation model, peer competence marginally predicted children's ToM skills with mothers, and another "cool" form of EC, the capacity to perform go-no go and effortful attention tasks, marginally predicted children's ToM skills with fathers.

Somewhat inconsistent findings across the different models should be noted. First, whereas motor inhibition (MI) significantly predicted children's ToM skills with mothers in the intrapersonal factors model, MI did not emerge as a significant predictor in the comprehensive mediation model. In that comprehensive model, peer competence marginally predicted children's

ToM skills with mothers, somewhat consistent with the results of the interpersonal factors model. The results may imply that peer competence and MI are both associated with children's ToM skills with mothers when examined separately; however, peer competence is more of a unique predictor of children's ToM skills with mothers than any other factors examined, albeit at a marginal level.

Second, no EC skills significantly predicted children's ToM skills with fathers in the intrapersonal factors model. However, go-no go and effortful attention skills (GN_EA) marginally predicted children's ToM skills with fathers in the mediation model. Perhaps, the inclusion of interpersonal factors and the exclusion of "hot" EC skills in the mediation model helped GN_EA marginally explain unique variance of ToM skills with fathers whereas three EC skills examined in the intrapersonal factor model overlapped too much among them in accounting for ToM skills with father.

Another marginally significant finding that father-child relationship negatively predicted children's ToM skills with fathers must be addressed. The result suggests an inconsistent mediation model (Davis, 1985), in which a suppression effect is present. MacKinnon, Krull, and Lockwood (2000) explain suppression as a situation in which the inclusion of a third variable strengthens an association between an independent variable and dependent variable, and the signs are opposite for the direct effect and indirect effect. The univariate association between father-child relationship quality and children's ToM skills with fathers was smaller than the association with mediating variables ($\beta = -.02$, $\beta = -.31$, respectively with fathers). Also, the direct effect between father-child relationship quality and children's ToM skills was in a negative direction, and the total indirect effect between the two variables was indeed in a positive direction. The inconsistent pathways are suggested to cancel each other out, resulting in the zero total effect of father-child relationship quality on children's ToM skills with fathers, consistent with the univariate analyses results (MacKinnon et al., 2000).

The Third Aim: Children's ToM Skills in the Context of Clinical Symptoms

The analyses that addressed the links between ToM and children's clinical symptoms were productive and produced expected and interesting findings. The results of the third aim supported the hypothesis that ToM skills are broadly associated with a range of clinical symptoms, beyond symptoms of ASD. I simultaneously examined ToM skills in kindergarten years and at age 10 (as the child's overall PT skills deployed with both parents in my new task) and their associations with clinical symptoms -- ADHD, OS, and IS symptoms -- in the three respective incremental models. Children's clinical symptoms reported by self, parents, and teachers were successfully modeled as latent variables, which is another contribution of this study.

Consistent with the literature, I found that earlier ToM skills were associated with ADHD and OS symptoms in their respective models in the expected direction. I hypothesized that the strong associations between young children's ADHD and OS symptoms and ToM skills found in the literature may be due to the ceiling effects of the existing ToM measures or children's compromised EC skills that commonly co-occur with ADHD and OS symptoms. However, I found that early ToM skills incrementally predicted ADHD symptoms at age 10, above and beyond children's EC skills (recall EC was covaried) and the concurrent ToM skills. Similar results were found for OS symptoms as well: Early ToM skills marginally predicted ToM skills at age 10 above and beyond children's EC skills and the concurrent ToM skills.

The results imply that young children's compromised ToM skills may be a warning sign, indicating an early risk for developing ADHD or OS symptoms later in life. They also inform a potential intervention. Perhaps enhancing false-belief understanding in young children may help reduce their risks of developing ADHD or OS symptoms.

Of course, the inferences in the current study are limited in how much they reveal about a more developmentally complete picture. For example, ToM skills and ADHD or OS symptoms may bi-directionally interact over time, or EC skills may directly or indirectly impact the clinical symptoms. To address such questions would require analyses using a cross-lagged model. Nonetheless, the association between early ToM skills and later ADHD was unique and significant, and that between early ToM skills and later OS symptoms was marginally significant. Thus, further research is warranted to elucidate a comprehensive developmental pathway that involves ToM skills, ADHD or OS symptoms, and EC skills.

Surprisingly, there has been hardly any research on the links between children's ToM skills and internalizing symptoms. Consequently, the current study is novel and the results are very promising. I found that ToM skills at age 10 were concurrently negatively associated with children's IS symptoms. Children who showed poorer overall perspective taking abilities in our new interactive task had more internalizing symptoms. Moreover, concurrent ToM abilities mediated the association between early ToM skills and children's IS symptoms at age 10, suggesting a causal developmental path from early ToM difficulties to future risk for anxiety and depression.

Future research should address the nature of the concurrent links between ToM and internalizing symptoms in middle childhood. It may be the case that the compromised ToM skills may interfere with children's current social functioning and result in social exclusion and rejection. Those social experiences, in turn, may lead to depression, anxiety, and other IS symptoms. It is also possible that increased internalizing symptoms may interfere with children's social cognitive ability as well, for example, due to an increased self-focus typical for depression (Lee, Harkness, Sabbagh, & Jacobson, 2005), although the finding that early ToM skills predicted internalizing symptoms through concurrent ToM skills supports the former case. It would be important to examine whether concurrent associations between ToM skills and internalizing symptoms occur also at other developmental periods (e.g., kindergarten years, adolescent years). As well, it will be important to determine whether ToM skills are uniquely associated with specific internalizing symptoms (e.g., depression, anxiety, separation anxiety, etc.). Regardless, the findings are a unique and novel contribution to the field, demonstrating that children's social cognitive processes may be involved in the development of internalizing symptoms in children.

Limitations

To the best of my knowledge, this is the first study that has assessed ToM skills via an interactive task with a social partner, integrated interpersonal and intrapersonal factors to predict ToM skills in middle childhood, and examined associations between ToM skills and a broad range of clinical symptoms in typically developing children. Despite its innovation, however, the study is subject to several limitations.

First, our samples were relatively small in size and homogeneous in composition. Although statistical techniques such as the modified t-test procedure or a bootstrapping method were used to address relatively small sample sizes in the current study, replication with a larger sample is necessary. Generally speaking, the Family Study children included in the current study were a well-functioning, non-clinical, community sample, and their CSI-4 scores were in the normative range. Thus, the findings regarding the associations between the ToM skills and clinical symptoms may not apply to samples of children with elevated clinical scores. Therefore, a replication with clinical or at-risk samples is also recommended. As well, participants with ASD in the current study were all male. Although males are four times more likely to be diagnosed with ASD than females (Kogan et al., 2009), and thus female participants are harder to recruit, effort should be made to include female participants with ASD in future studies.

Second, the new ToM measure had some limitations as well. The following two points were discussed earlier as a potential explanation of the compromised validity of the new measure: (1) the possible range restriction on the PT proficiency rating, and (2) the possibility that PT error scores may not have functioned in the consistent way across the Family Study children and children with ASD. Another limitation of the new ToM measure is its reliance on children's cognitive and expressive language ability. This was inevitable, despite our efforts to focus the measure on behavioral and interactive ToM skills rather than on their conceptual aspect. For a child to guide the puzzle task successfully, the child had to correctly perceive the puzzle and deliver verbal instructions to his or her partner. The demands on children's perceptual and linguistic abilities may have affected children's performance. Thus, effects of those abilities cannot be ruled out. The literature has long shown that ToM skills are strongly associated with verbal skills (e.g., Happe, 1995). Further effort should continue to be made to develop a ToM measure that can be validly used across individuals with a wide range of verbal abilities.

Third, data were not analyzed in a way that maximized the benefit of their longitudinal nature. For example, parent-child relationship quality and peer competence were modeled as observed variables (an overall composite score across assessments). We had made that decision because the main goal was to investigate the relative effects of the overall, cumulative qualities of different interpersonal relationships on children's ToM skills. However, future studies may employ techniques such as hierarchical linear modeling to help answer other important questions, for example, whether changes in the quality of children's interpersonal relationships over time impact children's ToM skills. From the intervention perspective, it would be also extremely informative to examine whether early versus later parent-child relationships have more impact on children's ToM skills.

Conclusion

This research has implications for both basic and translational research. First, ToM may be a multi-faceted construct that largely consists of conceptual and behavioral dimensions. ToM is a source of pragmatic skills deployed in social interactions with others. To date, the literature has largely focused on the conceptual aspects of ToM, entirely ignoring its behavioral aspects. Thus, more attention is warranted for the behavioral dimension of the construct. This work has generated a new ToM task, entirely embedded in the flow of social interaction.

Second, the development of ToM should be studied not only in early years of the child's life, as has been mostly the case in the extant research, but well beyond early childhood. The current study showed that interpersonal factor and intrapersonal factor predicted ToM skills in middle childhood. Also, the analyses that included ToM skills in both early and middle childhood showed promising and potentially complex links between ToM skills and children's broadly-ranging behavioral problems.

Third, ToM skills are influenced by both interpersonal and intrapersonal factors. The findings have particular implications for the treatment literature. An intervention program that targets both interpersonal and intrapersonal factors, including aspects of peer relationship and child temperamental traits (e.g., effortful control) to improve ToM skills may in turn promote overall social communicative skills.

Fourth, ToM skills are associated with a broader range of clinical symptoms beyond ASD symptoms, the only set of disorders extensively studied to date in this context. Most, if not all, clinical symptoms interfere with individuals' social functioning. The study of associations between ToM, a core skill that facilitates one's navigation in the social world and multiple clinical symptoms, is certainly a promising endeavor.

Table 1

Summary of behavioral and reported measures of theory of mind, parent-child relationship, peer competence, temperament, and clinical symptoms

Measures	Data Type	Age of Assessment (In Months)							
		15	25	38	52	67	80	100	123
Theory of Mind									
FBT	B				X	X			
Interactive Task	B								X
Parent-Child Relationship									
MRO	B	X	X	X	X	X	X		
Peer Competence									
HBQ	MR, FR					X	X	X	
Temperament									
Effortful Control	B			X	X	X	X		

Table 1 (Continued)

Measures	Data Type	Age of Assessment (In Months)							
		15	25	38	52	67	80	100	123
Clinical Symptoms									
CSI-4	MR, FR, TR								X
Dominic-R	CR								X

Note. FBT = False-Belief Task; MRO = Mutually Responsive Orientation; HBQ = MacArthur's Health Behavior Questionnaire; CSI-4 = Child Symptom Inventory-4.

Data Type: B = Behavioral; MR= Mother Report; FR= Father Report; TR = Teacher Report; CR= Child Self-Report

Table 2

Means, standard deviations, and sample sizes for all variables included in the analyses

	N	Mean (SD)	Range
False-Belief Task at 52 months	98	3 (2.64)	0 – 8
False-Belief Task at 67 months	90	6.27 (2.50)	1 – 11
Early Theory of Mind skills	98	-.01 (.90)	-1.62 – 1.90
Overall PT Skills with Both Parents at age 10	78	-.02 (.71)	-1.57 – 2.08
PT Skills with Mothers at age 10	77	.00 (.87)	-1.69 – 1.77
PT Skills with Fathers at age 10	73	.00 (.87)	-2.21 – 2.39
Overall MRO across 15 to 80 months	100	3.08 (.29)	2.07 – 3.59
Mother-Child Dyad MRO across 15 to 80 months	100	3.14 (.33)	1.85 – 3.83
Father-Child Dyad MRO across 15 to 80 months	100	3.05 (.31)	1.99 – 3.63
Peer Competence across 67 to 80 months	91	-.14 (.16)	-.62 – .22
EC composite across 38 to 80 months	99	.00 (.81)	-2.84 – 1.57
Parent-Reported ADHD at age 10	81	13.72 (7.69)	.50 – 32
Parent-Reported OS at age 10	81	6.57 (4.07)	0 – 21
Parent-Reported IS at age 10	81	8.24 (5.07)	2 – 22.50
Teacher-Reported ADHD at age 10	67	7.94 (9.42)	0 – 36
Teacher-Reported OS at age 10	67	1.39 (2.65)	0 – 12
Teacher-Reported IS at age 10	67	3.31 (3.18)	1 – 15
Child-Reported ADHD at age 10	78	1.59 (1.91)	0 – 8
Child-Reported OS at age 10	78	2.12 (2.59)	0 – 13

Table 2 (Continued)

	N	Mean (SD)	Range
Child-Reported IS at age 10	78	9.12 (8.49)	0 – 41

Note. PT = Perspective Taking; MRO = Mutually Responsive Orientation; MI = Motor Inhibition; EC = Effortful Control; OS = Oppositional Symptoms; IS = Internalizing Symptoms; ADHD = Attention Deficit Hyperactivity Disorder

Early Theory of Mind skills, Overall PT Skills, PT skills with mothers, PT skills with fathers, MI Function, and EC composite are aggregates of standardized variables. Means, standard deviations, and ranges for the remaining variables are not standardized and represent actual scores.

Table 3

Descriptive statistics for the matched comparison sample and ASD participants' scores and associated modified t-values

	Family Study Sample, Boys only				ASD Sample						
	Mean	SD	95% CI Mean		#01	#02	#03	#04	#05	#06	#07
			LL	UL	Score[b]/ t-value	Score/ t-value	Score/ t-value	Score/ t-value	Score/ t-value	Score/ t-value	Score/ t-value
PT Errors	.66	.15	.61	.70	.53/	.78/	.72/	.67/	.57/	.77/	.87/
with M[a]					-.85	.80	.40	.07	-.59	.73	**1.4**[+]
PT Errors	.69	.15	.65	.74	n/a	.90/	.62/	.65/	.58/	n/a	n/a
with F[a]						**1.4**[+]	-.49	-.29	-.76		
PT Pro.	2.29	.92	2.01	2.57	1/	3/	1/	2/	1/	3/	3/
with M					**-1.39**[+]	.76	**-1.39**[+]	-.31	**-1.39**[+]	.76	.76
PT Pro.	2.36	.96	2.06	2.66	n/a	3/	1/	2/	1/	n/a	n/a
with F						.66	**-1.40**[+]	-.37	**-1.40**[+]		

Table 3 (Continued)

Note. LL = Lower Limit, UL = Upper Limit PT= Perspective Taking, Pro. = Proficiency, M = Mothers, F= Fathers, n/a = no father data

+ $p < .10$

[a] PT Errors are computed to represent the amount of no-error occurrence during the 5-minute episode.

[b] Scores for PT proficiency ranges from 1 to 5 and for PT errors ranges from .00 to 1.00.

Table 4
Goodness-of-fit indices for EC measurement model and SEM models

Model	Df	χ^2	RMSEA	CFI	TLI	SRMR
Three-factor EC model	17	23.16[a]	.060	.965	.943	.050
Figure 4	38	40.32*[a]	.071	.932	.887	.068
Figure 5	19	26.77	.064	.951	.885	.052

Note. EC = Effortful Control

[a] Chi-square test statistics that are robust to non-normality

* $p < .05$

Table 5
Goodness-of-fit indices for clinical symptom measurement models and SEM models.

Model	Df	χ^2	RMSEA	CFI	TLI	SRMR
ADHD measurement	2	2.13	.028	.999	.996	.026
ADHD SEM model	11	14.83	.059	.973	.948	.057
OS measurement	2	5.61	.149	.912	.736	.076
OS SEM model	6	7.25	.046	.984	.960	.044
IS measurement	2	2.13	.028	.997	.991	.033
IS SEM model	8	9.82	.048	.968	.941	.053

Note. ADHD = Attention Deficit Hyperactivity Disorder, OS = Oppositional Symptoms, IS = Internalizing Symptoms

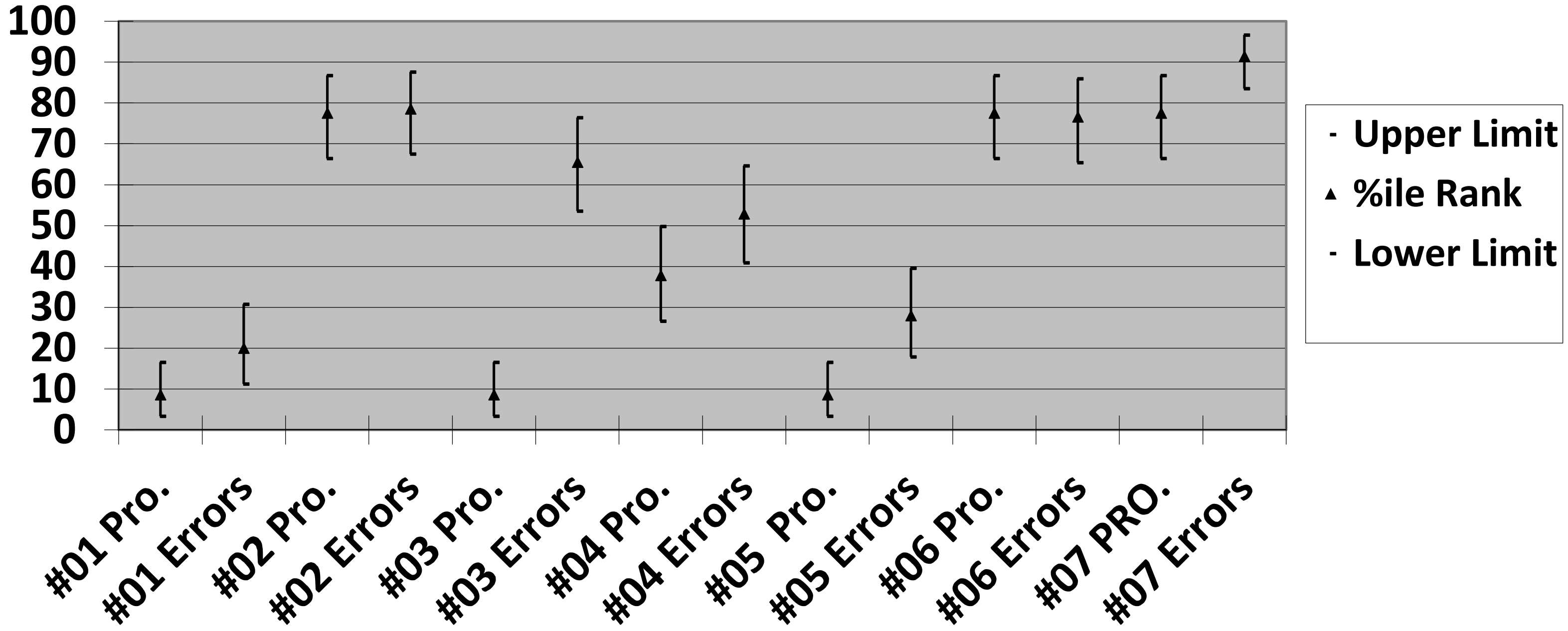

Figure 1. 95% confidence intervals for scores from the ASD sample on perspective-taking proficiency rating and errors with mothers.

Note. Pro. = Proficiency

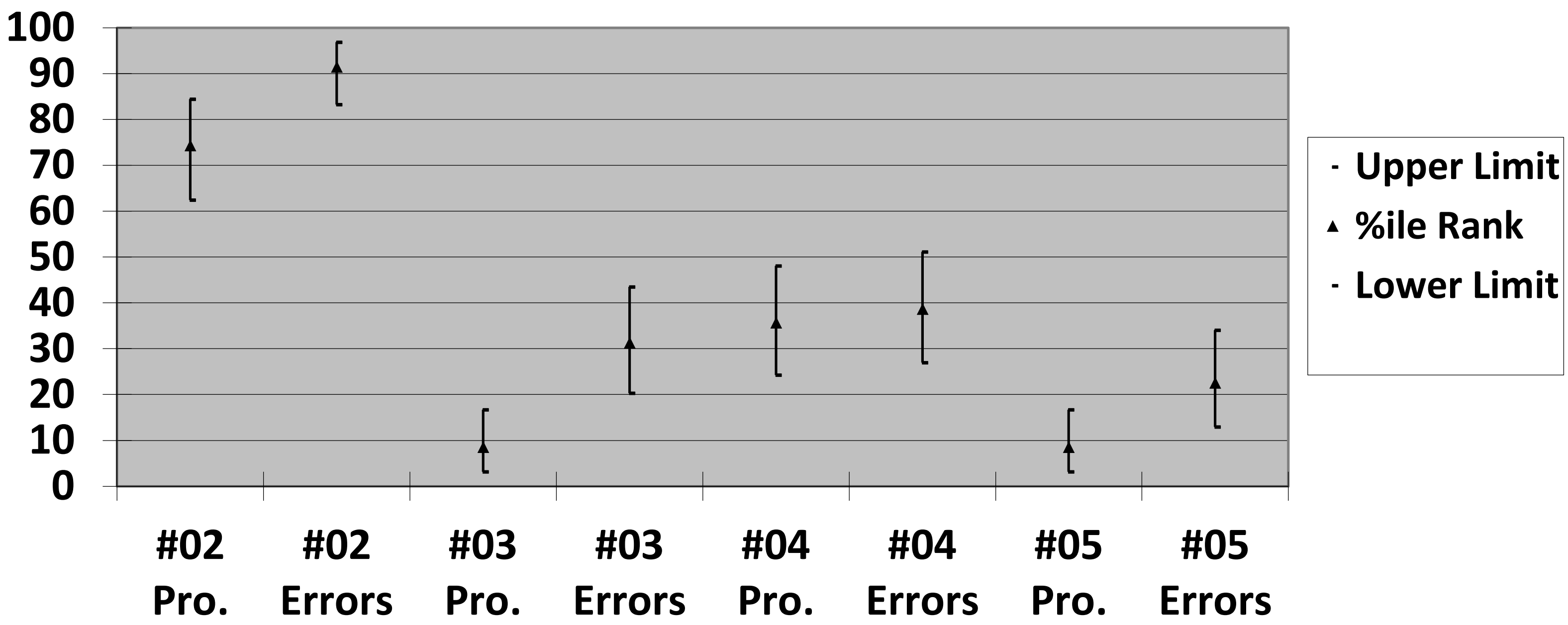

Figure 2. 95% confidence intervals for scores from the ASD sample on perspective-taking proficiency rating and errors with fathers.

Note. Pro. = Proficiency

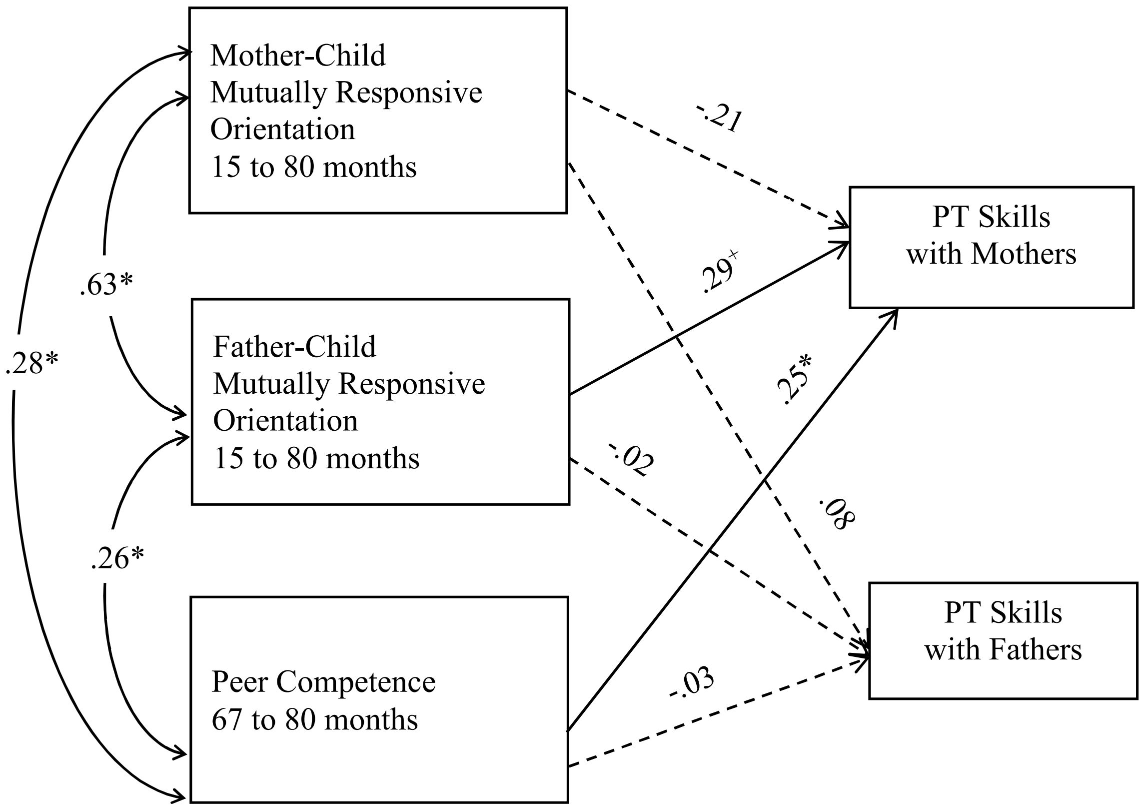

Figure 3. A structural equation model estimating the effects of interpersonal factors- mother-child MRO (15 – 80 months), father-child MRO (15 – 80 months), and peer competence (67 – 80 months) children's perspective-taking skills with each parent (age 10). Coefficients are standardized maximum likelihood estimates. Solid lines represent significant effects (*$p < .05$) and marginal effects ($^{+}\ p < .10$), and dashed lines represent non-significant effects.

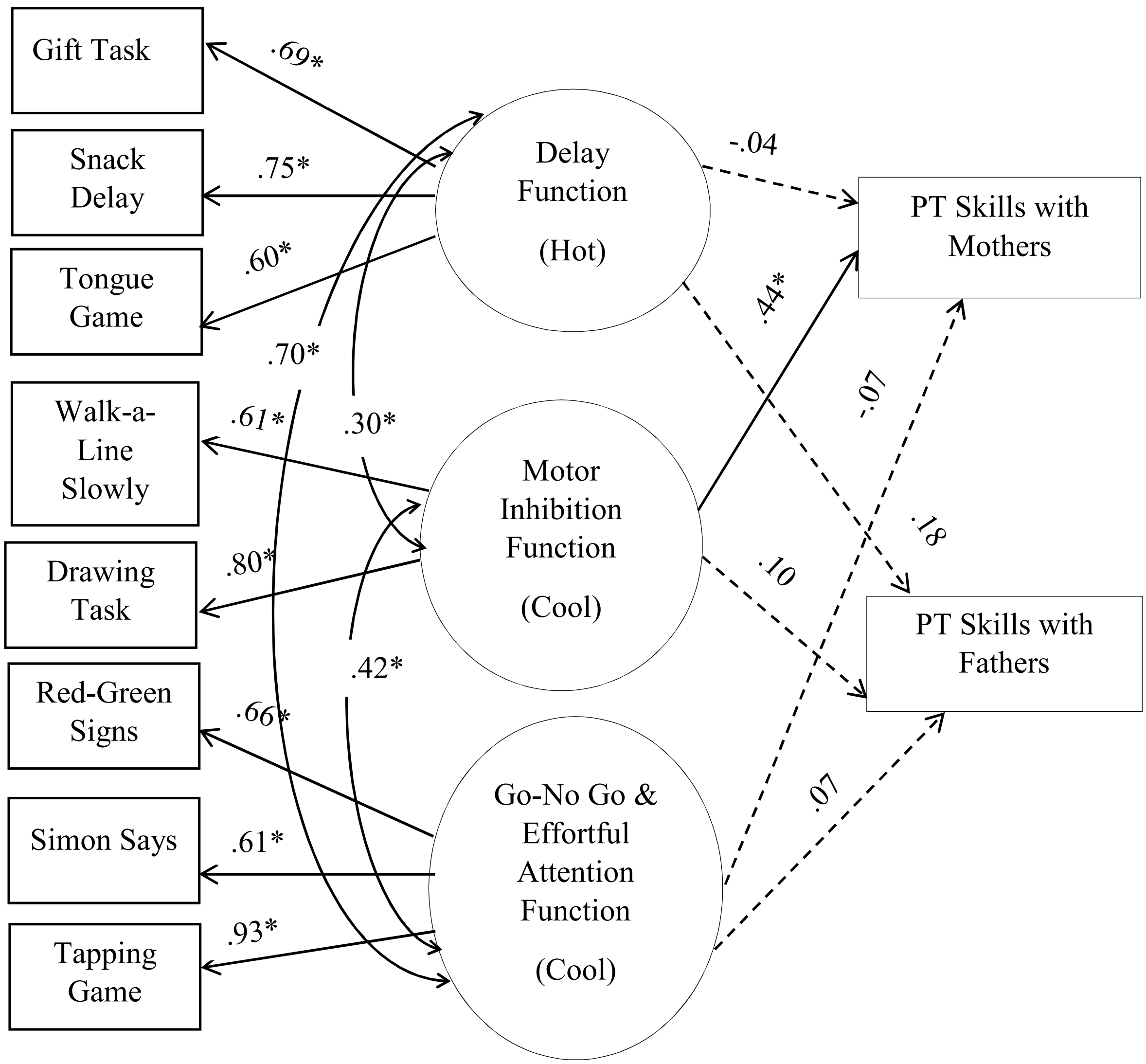

Figure 4. A structural equation model estimating the effects of "hot" and "cool" EC functions (38 – 80 months) on children's perspective-taking skills with each parent (age 10). Coefficients are standardized maximum likelihood estimates. Solid lines represent significant effects (*$p < .05$), and dashed lines represent non-significant effects.

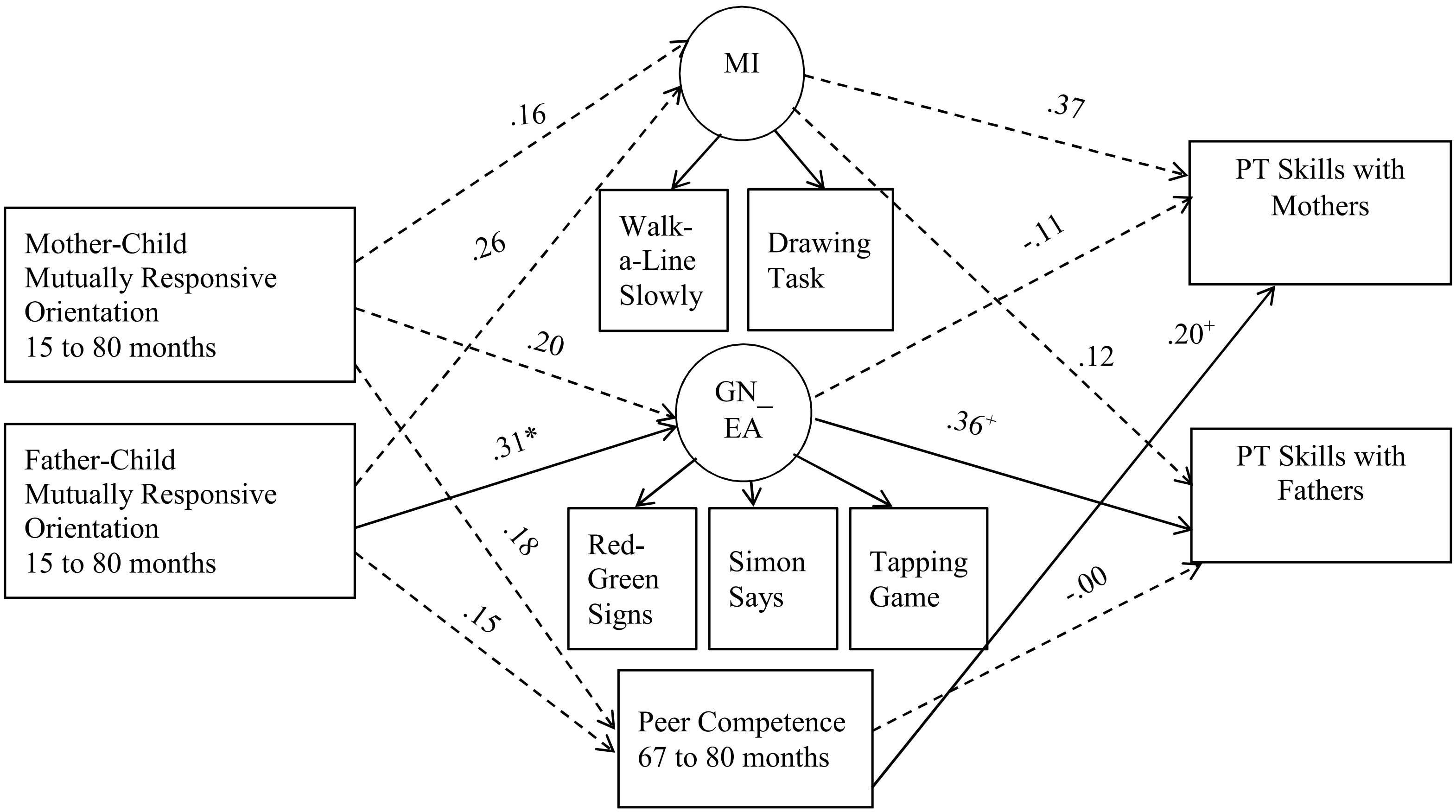

Figure 5. A structural equation model estimating the effects of interpersonal factors, mother-child MRO (15 – 80 months), father-child MRO (15 – 80 months), and peer competence (67 – 80 months), and intrapersonal factors, motor inhibition function (MI, 38-80 months) and go-no go and effortful attention function (GN_EA, 52-80 months), on children's perspective-taking skills with each parent (age 10). Direct paths from MRO to PT skills are omitted for the ease of presentation. Coefficients are standardized maximum likelihood estimates. Solid lines represent significant effects (*$p < .05$) and marginal effects ($^{+}$ $p < .10$), and dashed lines represent non-significant effects.

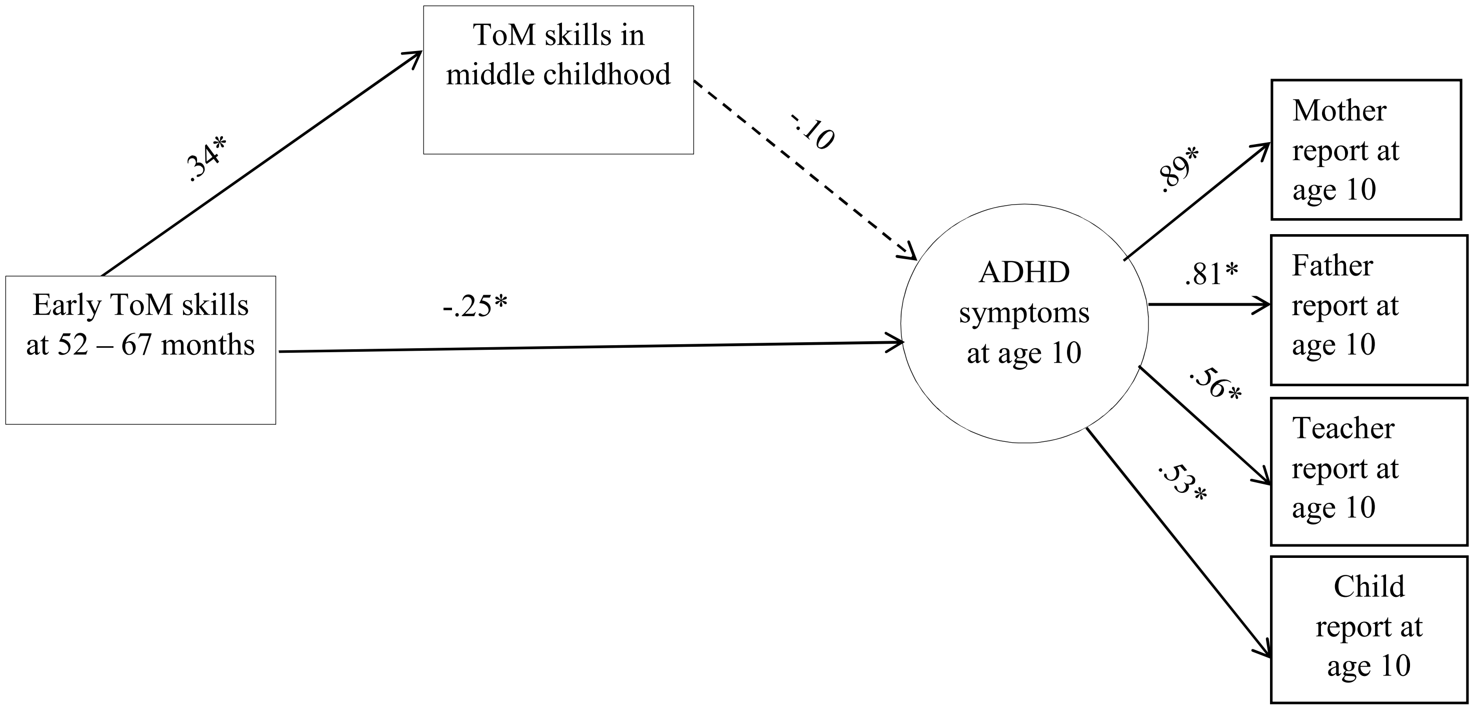

Figure 6. ADHD SEM model. A structural equation model estimating the unique effects of children's ToM skills in kindergarten years (52 – 67 months) and middle childhood (age 10) on children's ADHD symptoms (age 10). Coefficients are standardized maximum likelihood estimates. Effortful control is included in this model as a covariate with all the variables (not depicted). Solid lines represent significant effects (*$p < .05$), and dashed lines represent non-significant effects.

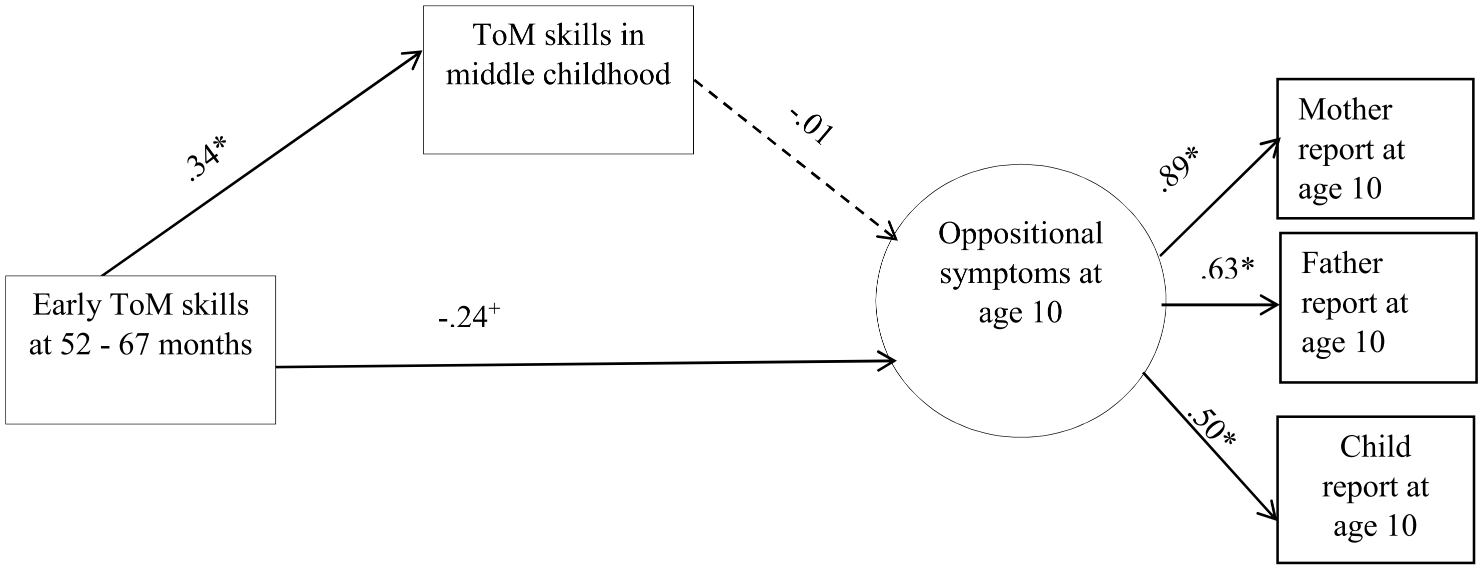

Figure 7. OS SEM model. A structural equation model estimating the unique effects of children's ToM skills in kindergarten years (52 – 67 months) and middle childhood (at age 10) on children's OS symptoms (age 10). Coefficients are standardized maximum likelihood estimates. Effortful control is included in this model as a covariate with all the variables (not depicted). Solid lines represent significant effects (*$p < .05$) and marginal effect ($^{+}$ $p < .10$), and dashed lines represent non-significant effects.

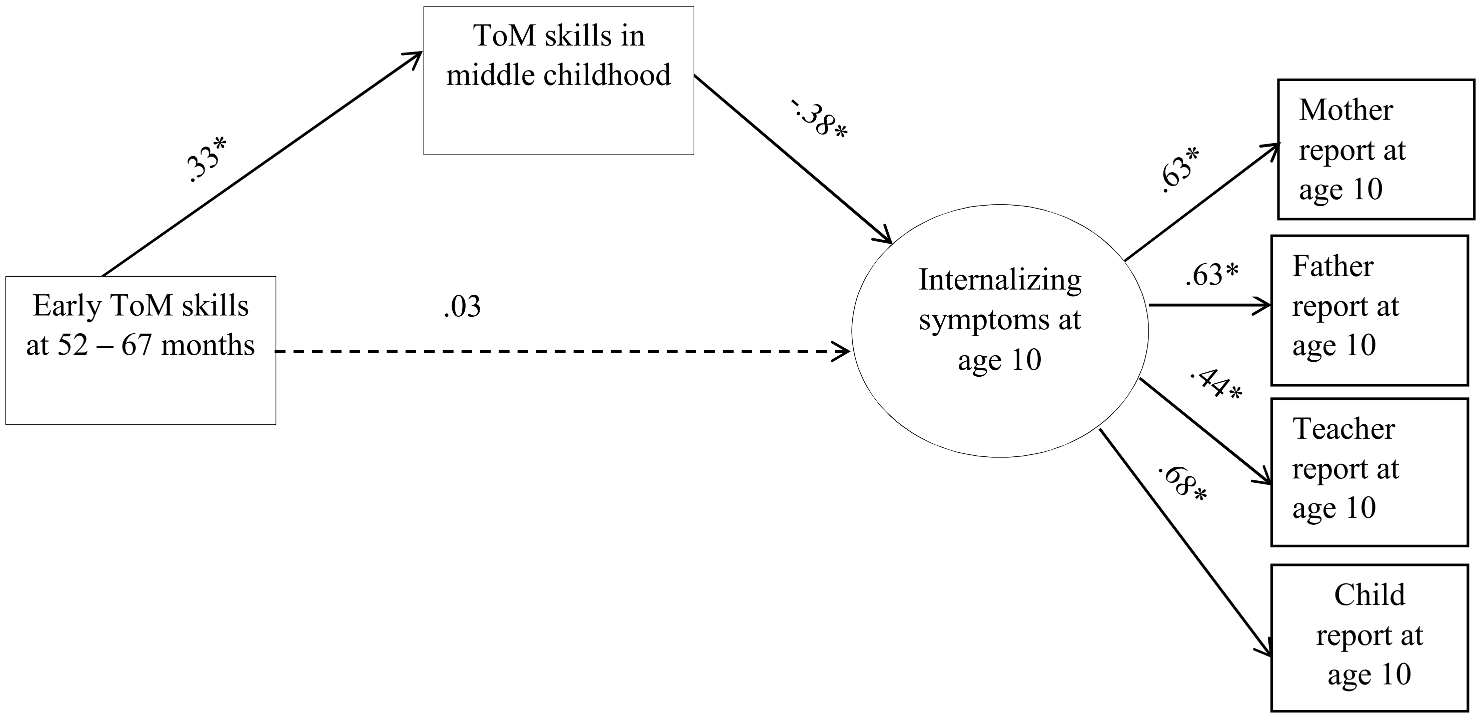

Figure 8. IS SEM model. A structural equation model estimating the unique effects of children's ToM skills in kindergarten years (52 – 67 months) and middle childhood on children's IS symptoms (age 10). Coefficients are standardized maximum likelihood estimates. Solid lines represent significant effects (*$p < .05$), and dashed lines represent non-significant effects.

APPENDIX

ToM Coding Descriptions

E = Experimenter; P = Parent; C = Child; G=Guider (the one who gives instructions in Object Assembly); R= Receiver (the one who is guided to complete a puzzle set in Object Assembly)

Perspective-taking errors.

Unclear physical direction (GU). A GU code is appropriate when G's instruction about physical directions is ambiguous without further clarification because G lacks the perspective-taking skills. Coders give a GU code when: (1) G gives physical directions from G's perspective. That is, when G said, "move it to the right," it means 'move it to G's right', (2) G uses referential language such as "move it over *there*," "turn *this (or that)* way," "put it *here,*" "flip it *over*," "turn it *over*," etc., and (3) G uses *vague* directions in giving physical directions. G uses words indicating specific places in a puzzle (e.g., a corner, side) but uses those words without a context. For example, a GU code is appropriate, when G said, "put that on the corner," but where 'the corner' is not established between G and R. (e.g., put it somewhere, no, wrong)

Visual information (GV). Coders give a GV code when G uses visual information to direct R. Coders consider a GV code when G gives directions that would only make sense to R if R could see his or her surroundings. Examples of visual information include: (1) G names colors of puzzle pieces (e.g., "pick a red piece"), (2) G refers puzzle pieces in a way that makes sense to G because G can see his or her surroundings. For example, a GV code is appropriate when G says, "grab the piece that is the furthest from you." This direction only makes sense to G because G can visually locate the furthest piece from R. However, it is not clear to R who cannot see his or her surroundings, and (3) G gestures as if R can see him or her. However, note that coders **do not** give a GV code for G's gestures if G's directions are clear otherwise. For example, a GV code **is not** appropriate when G gestures (i.e., point to R's left) while saying, "move the piece to your left."

Overall ratings of child perspective-taking proficiency.

A global ToM skills captures G's quality of perspective-taking ability, and thus it is the **quality** of G's ToM skills, not the quantity. Coders assign high global codes when G's ToM strategies efficiently guide R as R attempts to complete a puzzle set. The global ToM score captures G's proficiency in taking R's perspective in completing a puzzle set. In addition, coders should pay attention to how efficiently G works with R. Coders need to integrate their overall impressions of G into this global rating.

1. G's quality of ToM skills is very poor in whatever strategy G is employing. G is not picking up a cue from R's performance that G's strategy is not working for R to accomplish the task. In essence, G's instruction interferes with R's task and R is often confused solely because of G's instruction quality. G is impatient and lacks understanding of what R needs and how R feels. G is oblivious to the fact that G's strategy is not efficient and is the cause of the dyad's struggle.

2. G's quality of ToM skills is poor. G sees that R is unable to make progress on the task because of G's inefficient ToM strategy, but G is not quick to adapt in response. A difference between a 1 and 2 is G's attempt to change his or her strategy in ToM in response to R's performance. However, in a code of 2, G's attempt to adapt is neither sustained nor successful. Because G is not able to employ a better strategy to guide R but is aware of his or her inability, G's directions may be vague or passive.

3. G's quality of ToM skills is overall acceptable. G reasonably guides R through the task. However, G does not always employ efficient ToM strategy. G may employ inefficient ToM strategy for some time. However, G is able to sense that his or her strategy is not working and fixes his or her approach accordingly. Also, G may make several mistakes but is able to self-correct them. Thus, G's performance may improve as the dyad works through the task.

4. G's quality of ToM skills is overall efficient in making progress in the completion of the puzzle. G promptly self-corrects his or her mistakes when committed and G's instructions, for the most part, are clearly delivered to R and greatly contribute to the task completion. An overall impression in a code of 4 is that G is clearly engaged to "see" things from R's perspective.

5. G's quality of ToM skills is excellent. Inefficient strategies are rarely employed, and thus instances of self-corrections are scarcely observed. G rarely causes confusion in R because G is exceptionally careful in delivering his or her directions and monitors his or her own behavior carefully to avoid mistakes due to differences in perspectives. G is highly flexible and accommodating in working with R.

REFERENCES

Ainsworth, M.D.S., Bell, S.M., & Stayton, D.J. (1971). Individual differences in strange situation behavior of one year olds. In H.R. Schaffer (Ed.), The origins of human social relationships (pp.17-52). New York: Academic Press.

Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation.* Hillsdale, NJ: Erlbaum.

Allan, P.N., & Lonigan, C. J. (2011). Examining the dimensionality of effortful control in preschool children and its relation to academic and socioemotional indicators. *Developmental Psychology*, *47,* 905-915. doi:10.1037/a0023748

American Psychiatric Association. (2013) *Diagnostic and statistical manual of mental disorders* (5^{th} ed.). Washington, DC: Author.

American Psychiatric Association. (2000) *Diagnostic and statistical manual of mental disorders,* (4^{th} ed., text rev.). Washington, DC: Author.

Banerjee, R., Watling, D., & Caputi, M. (2011). Peer relations and the understanding of faux pas: Longitudinal evidence for bidrectional associations. *Child Development*, *82,* 1887-1906. doi:10.1111/j.1467-8624.2011.01669.x

Baron-Cohen, S. (1989). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 30,* 285-297. doi: 10.1111/j.1469-7610.1989.tb00241.x

Baron-Cohen (1997). *Mindblindness: An essay on autism and theory of mind.* Boston: MIT Press/Bradford Books.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 38,* 813-822. doi: 10.1111/j.1469-7610.1997.tb01599.x

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"?. *Cognition*, *21*, 37-46. doi:http://dx.doi.org.proxy.lib.uiowa.edu/10.1016/0010-0277(85)90022-8

Baron-Cohen, S., O'Riordan, M., Stone, V., Jones, R., & Plaisted, K. (1999). Recognition of faux pas by normally developing children and children with Asperger syndrome or high functioning autism. *Journal of Autism & Developmental Disorders, 29,* 407-418.doi:10.1023/A:1023035012436

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42,* 241-251. doi:10.1111/1469-7610.00715

Bartsch, K. & Wellman, H. (1995). *Children talk about the mind.* New York: Oxford University Press.

Belsky, J., Garduque, L., & Hrncir, E. (1984). Assessing performance, competence, and executive capacity in infant play: Relations to home environment and security of attachment. *Developmental Psychology, 20,* 406-417. doi:10.1037/0012-1649.20.3.406

Belsky, J., & Pluess, M. (2012). Differential susceptibility to long-term effects of quality of child care on externalizing behavior in adolescence? *International Journal of Behavioral Development*, *36,* 2-10. doi:10.1177/0165025411406855

Benson, J., & Sabbagh, M. (2010). Theory of mind and executive functioning: A developmental neuropsychological approach. In P. Zelazo, M. Chandler, & E. Crone (Eds.), *Developmental social cognitive neuroscience* (pp.63-80). New York, NY: Psychology Press.

Benson, J. E., Sabbagh, M. A., Carlson, S. M., & Zelazo, P. D. (2012). Individual differences in executive functioning predict preschoolers' improvement from theory-of-mind training. *Developmental Psychology*, *49,* 1615-1627. doi: http://dx.doi.org/10.1037/a0031056

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological bulletin*, *107*, 238. doi: http://dx.doi.org/10.1037/0033-2909.107.2.238

Boyce, W. T., Essex, M. J., Woodward, H. R., Measelle, J.R., Ablow, J. C., & Kupfer, D. J., (2002). The confluence of mental, physical, social, and academic difficulties in middle childhood. I: Exploring the "head waters" of early life morbidities. *Journal of the American Academy of Child and Adolescent Psychiatry*, *41,* 580-587. doi: 10.1097/00004583-200205000-00016

Bravo, G., & Potvin, L. (1991). Estimating the reliability of continuous measures with cronbach's alpha or the intraclass correlation coefficient: Toward the integration of two traditions. *Journal of Clinical Epidemiology, 44*, 381-390. doi: http://dx.doi.org/10.1016/0895-4356(91)90076-L

Brock, L. L., Rimm-Kaufman, S. E., Nathanson, L., Nathanson, L., & Grimm, K. J. (2009). The contributions of 'hot' and 'cool' executive function to children's academic achievement, learning-related behaviors, and engagement in kindergarten. *Early Childhood Research Quarterly, 24,* 337-349. doi: http://dx.doi.org/10.1016/j.ecresq.2009.06.001

Brüne, M., Abdel-Hamid, M., Lehmkämper, C., & Sonntag, C. (2007). Mental state attribution, neurocognitive functioning, and psychopathology: what predicts poor social competence in schizophrenia best?. *Schizophrenia research*, *92*, 151-159. doi: http://dx.doi.org.proxy.lib.uiowa.edu/10.1016/j.schres.2007.01.006

Buitelaar, J. K., van der Wees, M., Swaab-Bameveld, H., & van der Gaag, R. G. (1999). Theory of mind and emotion-recognition functioning in autistic spectrum disorders and in psychiatric control and normal children. *Development and Psychopathology, 11*, 39-58. doi: http://dx.doi.org/10.1017/S0954579499001947 10208355.

Caputi, M., Lecce, S., Pagnin, A., & Banergee, R. (2012). Longitudinal effects of theory of mind on later peer relations: The role of prosocial behavior. *Developmental Psychology*, *48,* 257-270. doi:10.1037/a0025402

Carlson, S. M., Mandell, D. J., & Williams, L. (2004). Executive function and theory of mind: Stability and prediction from ages 2 to 3. *Developmental Psychology, 40,* 1105-1122. doi:10.1037/0012-1649.40.6.1105

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development, 72,* 1032-1053. doi:10.1111/1467-8624.00333

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? *Infant and Child Development, 11,* 73-92. doi:10.1002/icd.298

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87,* 299-319. doi:10.1016/j.jecp.2004.01.002

Caspi, A., & Shiner, R. L. (2006). Personality development. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (6th ed., pp. 300–365). New York, NY: Wiley.

Cassidy, K. W., Werner, R. S., Rourke, M., Zubernis, L. S., & Balaraman, G. (2003). The Relationship between psychological understanding and positive social behaviors. *Social Development, 12,* 198-221. doi:10.1111/1467-9507.00229

Centers for Disease Control and Prevention, Child Development. (2015). *Positive parenting tips*. Retrieved from http://www.cdc.gov/ncbddd/childdevelopment/positiveparenting/index.html

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, *15,* 481-498. doi:10.1016/S0885-2014(01)00037-5

Charman, T., Carroll, F., & Sturge, C. (2001). Theory of mind, executive function, and social competence in boys with ADHD. *Emotional and Behavioral Difficulties, 6,* 31-49. doi:10.1177/1363275201006001004

Cole, K., & Mitchelle, P. (1998). Family background in relation to deceptive ability and understanding of the mind. *Social Development, 7,* 181-197. doi:10.1111/1467-9507.00061

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia, 40,* 1196-1208. doi:10.1016/S0028-3932(01)00224-X

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *Clinical Neuropsychologist, 12,* 482-486. doi:10.1076/clin.12.4.482.7241

Davis, M.D. (1985). The logic of causal order. In J. L. Sullivan and R. G. Niemi (Eds.), *Sage university paper series on quantitative applications in the social sciences.* Beverly Hills, CA: Sage Publications.

Davis, H. L., & Pratt, C. (1995). The development of children's theory of mind: The working memory explanation. *Australian Journal of Psychology, 47,* 25-31. doi:10.1080/00049539508258765

Dunn, J., Brown, J., Slomkowski, C., Tesla, C., & Youngblade, L. (1991). Young children's understanding of other people's feelings and beliefs: Individual differences and their antecedents. *Child Development, 62,* 1352-1366. doi:10.2307/1130811

Dyck, M. J., Ferguson, K., & Shochet, I. M. (2001). Do autism spectrum disorders differ from each other and from non-spectrum disorders on emotion recognition tests? *European Child & Adolescent Psychiatry, 10,* 105-116. doi:10.1007/s007870170033

Eisenberg, N., Zhou, Q., Spinrad, T. L., Valiente, C., Fabes, R. A., & Liew, J. (2005). Relations among positive parenting, children's effortful control, and externalizing problems: A three-wave longitudinal study. *Child Development, 76,* 1055-1071. doi:10.1111/j.1467-8624.2005.00897.x

Essex, M. J., Boyce, W. T., Goldstein, L. H., Armstrong, J. M., Kraemer, H. C., & Kupfer, D. J. (2002). The confluence of mental, physical, social, and academic difficulties in middle childhood. II: Developing the MacArthur health and behavior questionnaire. *Journal of the American Academy of Child and Adolescent Psychiatry, 41,* 588-603. doi: 10.1097/00004583-200205000-00017

Fahie, C. M., & Symons, D. K. (2003). Executive functioning and theory of mind in children clinically referred for attention and behavior problems. *Journal of Applied Developmental Psychology, 24,* 51-73. doi:10.1016/S0193-3973(03)00024-8

Filippova, E., & Astington, J. W. (2008). Further development in social reasoning revealed in discourse irony understanding. *Child Development, 79,* 126-138. doi:10.1111/j.1467-8624.2007.01115.x

Flavell, J. H. (1991).What young children know about the mind. *PsycCRITIQUES*, *36,* 741-742. doi: 10.1037/030125

Fodor, J. A. (1983). *The Modularity of Mind,* Cambridge, MA: MIT Press.

Fonagy, P., Redfern, S., & Charman, T. (1997). The relationship between belief--desire reasoning and a projective measure of attachment security (SAT). *British Journal of Developmental Psychology, 15,* 51-61. doi:10.1111/j.2044-835X.1997.tb00724.x

Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology, 9,* 679-700. doi:10.1017/S0954579497001399

Frye, D., Zelazo, P. D., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10,* 483-527. doi:10.1016/0885-2014(95)90024-1

Gadow, K. D., Sprafkin, J., Salisbury, H., Schneider, J., & Loney, J. (2004). Further validity evidence for the teacher version of the child symptom inventory-4. *School Psychology Quarterly*, *19,* 50-71. doi:10.1521/scpq.19.1.50.29408

Gadow, K. D., & Sprafkin, J. (2002). *Child Symptom Inventory-4: Screening and norms manual.* Stony Brook, NY: Checkmate Plus.

Gadow, K. D., Sprafkin, J., & Nolan, E. E. (2001). DSM-IV Symptoms in community and clinic preschool children. *Journal of the American Academy of Child and Adolescent Psychiatry, 40,* 1383-1392. doi:10.1097/00004583-200112000-00008

Ghiassi, V., Dimaggio, G., & Brüne, M. (2010). Dysfunctions in understanding other minds in borderline personality disorder: A study using cartoon picture stories. *Psychotherapy Research, 20,* 657-667. doi:10.1080/10503307.2010.501040

Goldstein, T. R., & Winner, E. (2010). Engagement in role play, pretense, and acting classes predict advanced theory of mind skill in middle childhood. *Imagination, Cognition and Personality, 30,* 249-258. doi:10.2190/IC.30.3.c

Gopnik, A., & Meltzoff, A. (1997). *Words, thoughts, and theories.* Cambridge, MA: MIT Press.

Guajardo, N. R., Snyder, G., & Petersen, R. (2009). Relationships among parenting practices, parental stress, child behaviour, and children's social-cognitive development. *Infant and Child Development, 18,* 37-60. doi:10.1002/icd.578

Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped and normal children and adults. *Journal of Autism & Developmental Disorders, 24,* 129-154. doi:10.1007/BF02172093

Happé, F. G. E. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66,* 843–855. doi: 10.1111/j.1467-8624.1995.tb00909.x

Happé, F. G. E., & Frith, U. (1996). Theory of mind and social impairment in children with conduct disorder. *British Journal of Developmental Psychology, 14*, 385-398. doi:10.1111/j.2044-835X.1996,tb00713.x

Harkness, K. L., Sabbagh, M. A., Jacobson, J. A., Chowdrey, N. K., & Chen, T. (2005). Enhanced accuracy of mental state decoding in dysphoric college students. *Cognition and Emotion, 18*, 999-1025. doi:10.1080/02699930541000110

Harris, P. L., Johnson, C. N., Hutton, D., Andrews, G., & Cooke, T. (1989). Young children's theory of mind and emotion. *Cognition and Emotion, 3*, 379-400. doi:10.1080/02699938908412713

Hartup, W. W. (1979). The social worlds of childhood. *American Psychologist*, *34*, 944–950. doi:10.1037/0003-066X.34.10.944

Hartup, W. W. (2005). Peer interaction: What causes what? *Journal of Abnormal Child Psychology, 33*, 387-394. doi:10.1007/s10802-005-3578-0

Hogrefe, G., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development, 57*, 567-582. doi:10.2307/1130337

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: Age-related changes and individual differences. *Developmental Neuropsychology, 28*, 617-644. doi:10.1207/s15326942dn2802_4

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*, 1-55. doi: 10.1080/10705519909540118

Hughes, C. (1998). Executive function in preschoolers: Links with theory of mind and verbal ability. *British Journal of Developmental Psychology, 16*, 233-253. doi:10.1111/j.2044-835X.1998.tb00921.x

Hughes, C., Deater-Deckard, K., & Cutting, A. L. (1999). 'Speak roughly to your little boy'? Sex differences in the relations between parenting and preschoolers' understanding of mind. *Social Development, 8*, 143-160. doi:10.1111/1467-9507.00088

Hughes, C., Dunn, J., & White, A. (1998). Trick or treat?: Uneven understanding of mind and emotion and executive dysfunction in "hard-to-manage" preschoolers. *Journal of Child Psychology and Psychiatry, 39*, 981-994. doi:10.1111/1469-7610.00401

Hughes, C., White, A., Sharpen, J. and Dunn, J. (2000), Antisocial, angry, and unsympathetic: "Hard-to-manage" preschoolers' peer problems and possible cognitive influences. *Journal of Child Psychology and Psychiatry, 41*, 169–179. doi: 10.1111/1469-7610.00558

Inoue, Y., Tonooka, Y., Yamada, K., & Kanba, S. (2004). Deficiency of theory of mind in patients with remitted mood disorder. *Journal of Affective Disorders, 82*, 403-409. doi:10.1016/j.jad.2004.04.004

Joiner Jr., T.E., (2002). Depression in its interpersonal context. In Gotlib, I.H., & Hammen, C.L. (Eds.), *Handbook of depression* (pp. 295–313). New York, NY: Guilford Press.

Kim, S., Nordling, J. K., Yoon, J. E., Boldt, L. J., & Kochanska, G. (2013). Effortful control in "hot" and "cool" tasks differentially predicts children's behavior problems and academic performance. *Journal of Abnormal Child Psychology, 41*, 43-56. doi:10.1007/s10802-012-9661-4

Klin, A. (2000). Attributing social meaning to ambiguous visual stimuli in higher functioning autism and Asperger syndrome: The social attribution task. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*, 831-46. doi: 10.1111/1469-7610.00671

Kochanska, G. (1997). Mutually responsive orientation between mothers and their young children: Implications for early socialization. *Child Development, 68*, 94-112. doi:10.2307/1131928

Kochanska, G. (2002). Committed compliance, moral self, and internalization: A mediational model. *Developmental Psychology, 38*, 339-351. doi:10.1037/0012-1649.38.3.339

Kochanska, G., & Aksan, N. (2006). Children's conscience and self-regulation. *Journal of Personality, 74*, 1587-1617. doi:10.1111/j.1467-6494.2006.00421.x

Kochanska, G., Aksan, N., Penney, S. J., & Doobay, A. F. (2007). Early positive emotionality as a heterogenous trait: Implications for children's self-regulation. *Journal of Personality and Social Psychology, 93*, 1054-1066. doi:10.1037/0022-3514.93.6.1054

Kochanska, G., & Kim, S. (2014). A complex interplay among the parent–child relationship, effortful control, and internalized, rule-compatible conduct in young children: Evidence from two studies. *Developmental Psychology, 50,* 8-21. doi: http://dx.doi.org/10.1037/a0032330

Kochanska, G., Kim, S., Boldt, L. J., & Yoon, J. E. (2013). Children's callous-unemotional traits moderate links between their positive relationships with parents at preschool age and externalizing behavior problems at early school age. *Journal of Child Psychology and Psychiatry, 54,* 1251-1260. doi:10.1111/jcpp.12084

Kochanska, G., Murray, K. T., & Harlan, E. T. (2000). Effortful control in early childhood: Continuity and change, antecedents, and implications for social development. *Developmental Psychology, 36*, 220-232. doi:10.1037/0012-1649.36.2.220

Kogan, M. D., Blumberg, S. J., Schieve, L. A., Boyle, C. A., Perrin, J. M., Ghandour, R. M., Singh, G.K., Strickland, B.B., Trevathan, E., & van Dyck, P. C. (2009). Prevalence of parent-reported diagnosis of autism spectrum disorder among children in the US, 2007. *Pediatrics*, *124*, 1395-1403. doi:10.1542/peds.2009-1522

Kovan, N. M., Chung, A. L., & Sroufe, L. A. (2009). The intergenerational continuity of observed early parenting: A prospective, longitudinal study. *Developmental Psychology, 45*, 1205-1213. doi:10.1037/a0016542

LaBounty, J., Wellman, H. M., Olson, S., Lagattuta, K., & Liu, D. (2008). Mothers' and fathers' use of internal state talk with their young children. *Social Development, 17*, 757-775. doi:10.1111/j.1467-9507.2007.00450.x

Lee, L., Harkness, K. L., Sabbagh, M. A., & Jacobson, J. A. (2005). Mental state decoding abilities in clinical depression. *Journal of Affective Disorders*, *86*, 247-258. doi: http://dx.doi.org/10.1016/j.jad.2005.02.007

Leslie, A. M. (1987). Pretense and representation: The origins of “theory of mind.” *Psychological Review*, *94*, 412–426. doi: 10.1037/0033-295X.94.4.412

Liew, J. (2012). Effortful control, executive functions, and education: Bringing self-regulatory and social-emotional competencies to the table. *Child development perspectives*, *6*, 105-111. doi: http://dx.doi.org.proxy.lib.uiowa.edu/10.1111/j.1750-8606.2011.00196.x

Lewis, C., Huang, Z., & Rooksby, M. (2006). Chinese preschoolers' false belief understanding: Is social knowledge underpinned by parental styles, social interactions or executive functions? *Psychologia: An International Journal of Psychology in the Orient, 49*, 252-266. doi:10.2117/psysoc.2006.252

MacKinnon, D. P., Krull, J. L., & Lockwood, C. M. (2000). Equivalence of the mediation, confounding and suppression effect. *Prevention Science*, *1*, 173-181. doi: 10.1023/A:1026595011371

McDowell, D. J., & Parke, R. D. (2009). Parental correlates of children’s peer relations: An empirical test of a tripartite model. *Developmental Psychology, 45,* 224–235. doi:10.1037/a0014305

McDowell, D. J., Parke, R. D., & Wang, S. J. (2003). Differences between mothers’ and fathers’ advice-giving style and content: Relations with social competence and psychological functioning behavior in middle childhood. *Merrill Palmer Quarterly, 49,* 55–77. doi:10.1353/mpq.2003.0004

McHughe, L., Stewart, I., & Hooper, N. (2012). A contemporary functional analytic account of perspective taking. In L. McHugh, & I. Stewart (Eds.), *The self and perspective taking* (pp. 55-71). Oakland, CA: Context Press.

Meins, E. (1997). *Security of attachment and the social development of cognition.* Hove: Psychology Press.

Meins, E. (1998). The effects of security of attachment and maternal attribution of meaning on children's linguistic acquisitional style. *Infant Behavior & Development, 21*, 237-252. doi:10.1016/S0163-6383(98)90004-2

Meins, E., Fernyhough, C., Johnson, F., & Lidstone, J. (2006). Mind-mindedness in children: Individual differences in internal-state talk in middle childhood. *British Journal of Developmental Psychology, 24*, 181-196. doi:10.1348/026151005X80174

Meins, E., Fernyhough, C., Russell, J., & Clark-Carter, D. (1998). Security of attachment as a predictor of symbolic and mentalising abilities: A longitudinal study. *Social Development, 7*, 1-24. doi:10.1111/1467-9507.00047

Meins, E., Fernyhough, C., Wainwright, R., Gupta, M. D., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development, 73*, 1715-1726. doi:10.1111/1467-8624.00501

Meins, E., & Russell, J. (1997). Security and symbolic play: The relation between security of attachment and executive capacity. *British Journal of Developmental Psychology, 15*, 63-76. doi:10.1111/j.2044-835X.1997.tb00725.x

Meltzoff, A. N. (1995). Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children. *Developmental Psychology, 31*, 838-850. doi:10.1037//0012-1649.31.5.838

Minter, M., Hobson, R. P., & Bishop, M. (1998). Congenital visual impairment and 'theory of mind'. *British Journal of Developmental Psychology, 16*, 183-196. doi:10.1111/j.2044-835X.1998.tb00918.x

Moor, B. G., Op de Macks, Zdeňa A., Güroğlu, B., Rombouts, Serge A. R. B., Van der Molen, Maurits W., & Crone, E. A. (2012). Neurodevelopmental changes of reading the mind in the eyes. *Social Cognitive and Affective Neuroscience, 7*, 44-52. doi:10.1093/scan/nsr020

Muthen, L. K., & Muthen, B. O. (2010). *MPLUS user's guide (6th ed.)*. Los Angeles, CA: Muthen & Muthen.

NICHD Early Child Care Research Network. (2005). Predicting individual differences in attention, memory, and planning in first graders from experiences at home, child care, and school. *Developmental Psychology*, *41*, 99. doi: 10.1037/0012-1749.41.1.99

Nijmeijer, J. S., Minderaa, R. B., Buitelaar, J. K., Mulligan, A., Hartman, C. A., & Hoekstra, P. J. (2008). Attention-deficit/hyperactivity disorder and social dysfunctioning. *Clinical Psychology Review, 28*, 692-708. doi:10.1016/j.cpr.2007.10.003

Novak, G. (2012). Precursors to perspective taking. In L. McHugh, & I. Stewart (Eds.), *The self and perspective taking* (pp. 73-89). Oakland, CA: Context Press.

Ontai, L. L., & Thompson, R. A. (2008). Attachment, parent-child discourse and theory-of-mind development. *Social Development, 17*, 47-60. doi:10.1111/j.1467-9507.2007.00414.x

Parke, R. D., & Buriel, R. (2006). Socialization in the family: Ethnic and ecological perspectives. In W. Damon, R. Lerner, & N. Eisenberg (Eds.), *Handbook of Child Psychology. Vol. 3. Social, Emotional, and Personality Development* (6th ed., pp. 95–138). New York, NY: Wiley.

Parke, R. D., & O’Neil, R. (1999). Social relationships across contexts: Family-peer linkages. In W. L. Collins, & B. Laursen. (Eds.), *Relationships as developmental contexts, The Minnesota symposia on child psychology, Volume 30* (pp. 211-239). Hillsdale, NJ: Lawrence Erlbaum Associates

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: Clues to developmental primacy. *Developmental Psychology*, *43*, 974. doi: 10.1037/0012-1649.43.4.974

Pellicano, E., Maybery, M., Durkin, K., & Maley, A. (2006). Multiple cognitive capabilities/deficits in children with an autism spectrum disorder: "Weak" central coherence and its relationship to theory of mind and executive control. *Development and Psychopathology, 18*, 77-98. doi:10.1017/S0954579406060056

Perner, J. (1991). *Understanding the representational mind.* Cambridge, MA: MIT Press.

Perner, J., Kain, W., & Barchfeld, P. (2002). Executive control and higher-order theory of mind in children at risk of ADHD. *Infant and Child Development, 11*, 141-158. doi:10.1002/icd.302

Perner, J., & Lang, B. (1999). Development of theory of mind and executive control. *Trends in Cognitive Sciences, 3*, 337-344. doi:10.1016/S1364-6613(99)01362-5

Perner, J., Ruffman, T., & Leekam, S. R. (1994). Theory of mind is contagious: You catch it from your sibs. *Child Development, 65*, 1228-1238. doi:10.2307/1131316

Perner, J., & Wimmer, H. (1985). "John thinks that Mary thinks that…": Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology, 39*, 437-471. doi:10.1016/0022-0965(85)90051-7

Peterson, C. C., & Siegal, M. (2002). Mindreading and moral awareness in popular and rejected preschoolers. *British Journal of Developmental Psychology, 20*, 205-224. doi:10.1348/026151002166415

Peterson, C. C., & Slaughter, V. (2009). Theory of Mind (ToM) in children with autism or typical development: Links between eye-reading and false belief understanding. *Research in Autism Spectrum Disorders, 3*, 462-473. doi:10.1016/j.rasd.2008.09.007

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development, 83*, 469-485. doi:10.1111/j.1467-8624.2011.01728.x

Peterson, C. C., Wellman, H. M., & Slaughter, V. (2012). The mind behind the message: Advancing theory-of-mind scales for typically developing children, and those with deafness, autism, or Asperger syndrome. *Child Development, 83*, 469-485. doi:10.1111/j.1467-8624.2011.01728.x

Power, T. G., & Chapieski, M. L. (1986). Childrearing and impulse control in toddlers: A naturalistic investigation. *Developmental Psychology, 22*, 271-275. doi:10.1037/0012-1649.22.2.271

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1,* 515-526. doi: 10.1017/S0140525X00076512

Prior, M., Dahlstrom, B., & Squires, T. L. (1990). Autistic children's knowledge of thinking and feeling states in other people. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 31,* 587-601. doi: 10.1111/j.1469-7610.1990.tb00799.x

Putnam, S. P., Gartstein, M. A., & Rothbart, M. K. (2006). Measurement of fine-grained aspects of toddler temperament: The early childhood behavior questionnaire. *Infant Behavior & Development, 29*, 386-401. doi:0.1016/j.infbeh.2006.01.004

Razza, R. A., & Blair, C. (2009). Associations among false-belief understanding, executive function, and social competence: A longitudinal analysis. *Journal of Applied Developmental Psychology, 30*, 332-343. doi:10.1016/j.appdev.2008.12.020

Ronald, A., Viding, E., Happé, F., & Plomin, R. (2006). Individual differences in theory of mind ability in middle childhood and links with verbal ability and autistic traits: A twin study. *Social Neuroscience, 1*, 412-425. doi:10.1080/17470910601068088

Rothbart, M. K. (1989). Temperament and development. In J.E. Kohnstamm, J.E. Bates & M.K. Rothbart (Eds.), *Temperament in childhood* (pp. 187-247). New York, NY: John Wiley.

Rothbart, M. K., & Ahadi, S. A. (1994). Temperament and the development of personality. *Journal of Abnormal Psychology, 103*, 55-66. doi:10.1037/0021-843X.103.1.55

Rothbart, M. K., & Bates, J. E. (1998). Temperament. In W. Damon (Series Ed.) and Eisenberg N. (Vol. Ed.), *Handbook of child psychology: Vol. 3, Social emotional and personality development* (5th ed., pp. 105-176). New York, NY: Wiley.

Rothbart, M. K., Chew, K. H., & Gartstein, M. A. (2001). Assessment of temperament in early development. In S. L. Twarog & Z. P. Sanford (Eds.), *Biobehavioral assessment of the infant* (pp. 190-208). New York, NY: Guilford Press

Rueda, C. (2012). Effortful control. In M. Zentner & R. Shiner (Eds.), *Handbook of temperament* (pp. 145-167). New York, NY: Guilford Press.

Ruffman, T., Slade, L., Devitt, K., & Crowe, E. (2006). What mothers say and what they do: The relation between parenting, theory of mind, language and conflict/cooperation. *British Journal of Developmental Psychology, 24*, 105-124. doi:10.1348/026151005X82848

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The development of executive functioning and theory of mind: A comparison of Chinese and U.S. preschoolers. *Psychological Science, 17*, 74-81. doi:10.1111/j.1467-9280.2005.01667.x

Samson, A. C., Lackner, H. K., Weiss, E. M., & Papousek, I. (2012). Perception of other people's mental states affects humor in social anxiety. *Journal of Behavior Therapy and Experimental Psychiatry, 43*, 625-631. doi:10.1016/j.jbtep.2011.08.007

Selman, R. L. (1980). *The growth of interpersonal understanding.* New York: Academic Press.

Sharp, C. (2008). Theory of Mind and conduct problems in children: Deficits in reading the "emotions of the eyes". *Cognition and Emotion, 22*, 1149-1158. doi:10.1080/02699930701667586

Sharp, C., & Fonagy, P. (2008). The parent's capacity to treat the child as a psychological agent: Constructs, measures and implications for developmental psychopathology. *Social Development, 17*, 737-754. doi:10.1111/j.1467-9507.2007.00457.x

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology, 26*, 978-986. doi:10.1037/0012-1649.26.6.978

Shojaei, T., Wazana, A., Pitrou, I., Gilbert, F., Bergeron, L., Valla, J., & Kovess-Masfety, V. (2009). Psychometric properties of the Dominic Interactive in a large French sample. *Canadian Journal of Psychiatry, 54*, 767-776. Retrieved from http://publications.cpa-apc.org/browse/sections/0

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychological Methods*, *7*, 422–445. doi:10.1037//1082-989X.7.4.422

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86*, 420-428. doi:10.1037/0033-2909.86.2.420

Slade, A. (1987). A longitudinal study of maternal involvement and symbolic play during the toddler period. *Child Development,* 58, 367-375. doi: http://dx.doi.org/10.2307/1130513

Slade, L., & Ruffman, T. (2005). How language does (and does not) relate to theory of mind: A longitudinal study of syntax, semantics, working memory and false belief. *British Journal of Developmental Psychology, 23*, 117-141. doi:10.1348/026151004X21332

Slaughter, V., Dennis, M. J., & Pritchard, M. (2002). Theory of mind and peer acceptance in preschool children. *British Journal of Developmental Psychology, 20,* 545-564. doi:10.1348/026151002760390945

Smiley, P. A. (2001). Intention understanding and partner-sensitive behaviors in young children's peer interactions. *Social Development*, *10*, 330-354. doi: 10.1111/1467-9507.00169

Smith, R., & LaFreniere, P. J. (2009). Development of children's ability to infer intentions from nonverbal cues. *The Journal of Social, Evolutionary, and Cultural Psychology, 3*, 315-327. doi: http://dx.doi.org/10.1037/h0099313

Sobel, M. E. (1982). Asymptotic confidence intervals for indirect effects in structural equation models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290-312). Washington, DC: American Sociological Association.

Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science*, *10,* 199–204. doi: 10.1111/j.1467-7687.2007.00580.x

Sprafkin, J., Gadow, K. D., Salisbury, H., Schneider, J., & Loney, J. (2002). Further evidence of reliability and validity of the child symptom inventory-4: Parent checklist in clinically referred boys. *Journal of Clinical Child and Adolescent Psychology, 31*, 513-524. doi:10.1207/153744202320802188

Sroufe, L.A., Egeland, B, & Carlson, E. A. (1999). One social world: The integrated development of parent-child and peer relationships. In W. L. Collins, & B. Laursen. (Eds.), *Relationships as developmental contexts, The Minnesota symposia on child psychology, Volume 30* (pp. 241-261). Hillsdale, NJ: Lawrence Erlbaum Associates

Sroufe, L.A., Egeland, B, Carlson, E. A., & Collins, W.A. (2005). *The development of the person: The Minnesota Study of Risk and Adaptation from Birth to Adulthood.* New York, NY: Guilford Press

Steiger , J. H. & Lind , J. C. (1980, May). *Statistically-based tests for the number of common factors.* Paper presented at the annual spring meeting of the psychometric society, Iowa City, IA.

Sulik, M. J., Huerta, S., Zerr, A. A., Eisenberg, N., Spinrad, T. L., Valiente, C., . . . Taylor, H. B. (2010). The factor structure of effortful control and measurement invariance across ethnicity and sex in a high-risk sample. *Journal of Psychopathology and Behavioral Assessment, 32*, 8-22. doi:10.1007/s10862-009-9164-y

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30*, 395-402. doi:10.1037/0012-1649.30.3.395

Suway, J. G., Degnan, K. A., Sussman, A. L., & Fox, N. A. (2012). The relations among theory of mind, behavioral inhibition, and peer Interactions in early childhood. *Social Development*, *21,* 331–342. doi: 10.1111/j.1467-9507.2011.00634.x

Symons, D. K., & Clark, S. E. (2000). A longitudinal study of mother–child relationships and theory of mind in the preschool period. *Social Development, 9*, 3-23. doi:10.1111/1467-9507.00108

Tomasello, M., & Josep, C. (1997). *Primate cognition.* New York, NY: Oxford University Press.

Uekermann, J., Kraemer, M., Abdel-Hamid, M., Schimmelmann, B. G., Hebebrand, J., Daum, I., … & Kis, B. (2010). Social cognition in attention-deficit hyperactivity disorder (ADHD). *Neuroscience and Biobehavioral Reviews, 34*, 734-743. doi:10.1016/j.neubiorev.2009.10.009

U.S. Census Bureau (2014). *State & County QuickFacts*, Retrieved from http://quickfacts.census.gov/qfd/states/19/19103.html

Valla, J. P., Bergeron, L., & Smolla, N. (2000). The Dominic-R: A pictorial interview for 6- to 11-year-old children. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 85–93. doi:10.1097/00004583-200001000-00020

Watson, A. C., Nixon, C. L., Wilson, A., & Capage, L. (1999). Social interaction skills and theory of mind in young children. *Developmental Psychology, 35*, 386-391. doi:10.1037/0012-1649.35.2.386

Watson, D. (2001). Dissociations of the night: Individual differences in sleep-related experiences and their relation to dissociation and schizotypy. *Journal of Abnormal Psychology,* 110, 526-535doi: http://dx.doi.org.proxy.lib.uiowa.edu/10.1037/0021-843X.110.4.526

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*, 655-684. doi:10.1111/1467-8624.00304

Wellman, H. M., Fang, F., Liu, D., Zhu, L., & Liu, G. (2006). Scaling of theory-of-mind understandings in Chinese children. *Psychological Science, 17*, 1075-1081. doi:10.1111/j.1467-9280.2006.01830.x

Wellman, H. M., & Liu, D. (2004). Scaling of theory-of-mind tasks. *Child development*, *75*, 523-541. doi: 10.1111/j.1467-8624.2004.00691.x

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition, 13*, 103-128. doi:10.1016/0010-0277(83)90004-5

Wolfe, C. D., & Bell, M. A. (2007). Sources of variability in working memory in early childhood: A consideration of age, temperament, language, and brain electrical activity. *Cognitive Development*, *22*, 431-455. doi:10.1016/j.cogdev.2007.08.007

Wright, I., Waterman, M., Prescott, H., & Murdoch-Eaton, D. (2003). A new Stroop-like measure of inhibitory function development: typical developmental trends. *Journal of Child Psychology and Psychiatry, 44*, 561–575. doi: 10.1111/1469-7610.00145

Zelazo, P. D., & Muller, U. (2002). Executive function in typical and atypical development. In U. Goswami (Ed.), *Handbook of childhood cognitive development* (pp. 445-469). Oxford:Blackwell.

Zhou, Q., Chen, S.H., & Main, A. (2012). Commonalities and differences in the research on children's effortful control and executive function: A call for an integrated model of self-regulation. *Child Development Perspectives, 6*, 112-121, doi:10.1111/j.1750-8608.2011.0017