J Autism Dev Disord (2008) 38:143–155
DOI 10.1007/s10803-007-0377-1

ORIGINAL PAPER

Beyond False Beliefs: The Development and Psychometric Evaluation of the Perceptions of Children's Theory of Mind Measure—Experimental Version (PCToMM-E)

Tiffany L. Hutchins · Laura A. Bonazinga · Patricia A. Prelock · Rebecca S. Taylor

Published online: 5 April 2007

Abstract The Perceptions of Children's Theory of Mind Measure (Experimental version; PCToMM-E) is an informant measure designed to tap children's theory of mind competence. Study one evaluated the measure when completed by primary caregivers of children with autism spectrum disorder. Scores demonstrated high test–retest reliability and correlated with verbal mental age and ToM task battery performance. No ceiling effects were observed. In addition, caregivers accurately predicted their children's ToM task battery performance. In study two the scores of primary caregivers of typically developing children demonstrated high test–retest reliability and distinguished children on the basis of age and developmental status. Ceiling effects were not evident until late childhood. The utility of the PCToMM-E and directions for future research are discussed.

Keywords Autism · Theory of mind · Assessment · Children · False belief

Introduction

Theory of Mind (ToM) broadly refers to the capacity to attribute mental states (e.g., beliefs, desires) to self and others. Its assessment has traditionally been based on performance on false belief or related tasks (discussed below). Two studies were conducted to evaluate the psychometric properties of a new method for assessing ToM knowledge in a sample of verbal and nonverbal children with ASD and typically developing children. The Perceptions of Children's Theory of Mind Measure - Experimental version (PCToMM-E) is an informant measure designed to assess the construct of ToM according to a broad definition: one that reflects the diversity in theory and assessment approaches that are part of the ToM literature. The following review describes (1) some traditional methods of assessing ToM, (2) the limitations associated with these methods, (3) parents as experts of children's knowledge and abilities and, (4) the procedures employed to evaluate of the PCToMM-E. It will ultimately be argued that a sound informant measure of children's ToM development is overdue.

Traditional Measures of ToM Development

Questions about the nature, timing, and manner in which ToM knowledge emerge have generated a variety of ToM assessment methods ranging from tasks designed to tap a child's ''developing understanding [of] conceptions of desires, emotions, beliefs, belief-desire reasoning, or psychological explanation, among others'' (Wellman, Cross, & Watson, 2001, p. 655) to assessments of the production of mental state terms (e.g., ''want'', ''think'', ''know'') (e.g., Miller, 2006; Tager-Flusberg, 1992, 1993). Among the ToM assessment procedures that have been devised, the classic false belief task (Wimmer & Perner, 1983) warrants special attention because of its prominence in research, practice, and theory for assessing typically developing children and children with ASD.

The well-known standard false belief task involves telling a story in which an object is moved from an original location to a new location without the knowledge of the

T. L. Hutchins (✉) · L. A. Bonazinga · P. A. Prelock · R. S. Taylor
Department of Communication Sciences, University of Vermont, 489 Main Street, 406 Pomeroy Hall, Burlington, VT 05445, USA
e-mail: Tiffany.Hutchins@uvm.edu

main protagonist and it is designed to tap the understanding that a person may hold a false belief where the mind's content contradicts reality. Several researchers have demonstrated that ToM understanding tapped by the false belief task is related to important aspects of social behavior (Astington & Jenkins, 1995; Lalonde & Chandler, 1995; Watson, Nixon, Wilson, & Capage, 1999). Clearly, the task taps skills that require some understanding of mental events. Operationalizing ToM based on false belief task performance alone, however, is woefully restrictive given that ToM is more accurately conceptualized as a folk psychology defined by a large set of social cognitive understandings. This problem is compounded by the fact that such tasks are scored as pass or fail. As a result

> ''we are led to believe that theory of mind is something one does or does not have-it emerges spontaneously at a single point in time. Autism research [has been] especially influenced by this narrowly defined approach to theory of mind...Thus the literature on autism often equates performance on a false-belief task to the presence or absence of a theory of mind, reducing what should be a rich, complex, unfolding mentalistic conception of people to a categorical capacity'' (Tager-Flusberg, 2001, pp. 177–178).

For these reasons, many researchers have argued for the value of aggregate measures in the form of task batteries that assess different components of ToM across levels of complexity (e.g., Hughes et al., 2000; Tager-Flusberg, 2001; Wellman et al., 2001) with the reasoning that a broader range of tasks allows for more sensitive measurement of ToM ability. A description of these tasks (hereafter referred to as standard ToM tasks) is beyond the scope of this paper (for a general review interested readers are referred to Baron-Cohen, 2000). The important point here is that performance on such tasks has come to serve as a general marker for the understanding of others' minds and thus play a pivotal role in ToM research (Begeer, Rieffe, Meerum Terwogt & Stockmann, 2003; Tager-Flusberg, 2001; Wellman et al., 2001; Yirmiya, Erel, Shaked, & Solomonica-Levi, 1998). Because of their prominence, limitations of the standard false belief task and aggregate standard ToM task batteries should be considered in light of how these limitations might be overcome by the use of a complimentary or alternative ToM assessment methodology.

Limitations of Standard ToM Tasks

An often cited limitation of standard ToM tasks is that performance is complicated by memory and linguistic factors (e.g., Happe, 1995; Kazak, Collis, & Lewis, 1997). That is, to solve the problem, both competence (the conceptual understanding) and performance factors (the ability to attend and remember the relevant information and to understand the language used) operate (Wellman et al., 2001). Several researchers have examined the influence of performance factors and have concluded that tasks like the Sally-Anne task are unnecessarily difficult thus masking ToM knowledge. This, in turn, has led to modifications in the task that are associated with enhanced performance and the result is advancement in methods to operationalize ToM. However, the influence of performance factors can not be entirely eliminated in tasks that inherently require attention, memory, and understanding of the language involved. Very young typically developing children and ''many people with autism cannot even be tested with standard theory of mind tasks, since they lack the cognitive and verbal skills necessary to answer the control questions, success on which is usually an inclusion criterion'' (Happe, 1995, pp. 845–847). In fact, many researchers in ASD employ false belief tasks only when the individual can demonstrate a minimum verbal mental age of approximately four years (e.g., Baron-Cohen, 1992) thereby excluding nonverbal children with ASD from participation. In this light, a reliable and valid informant measure of ToM knowledge for children with very limited verbal capacities would be valuable in allowing such examinations.

The standard ToM task performance of children (especially those with ASD) is prone to significant sources of error due to motivational and situational factors. Motivation requires a degree of active participation and may facilitate the performance of individuals if present while hinder performance if absent (Begeer et al., 2003). Situational factors that can impede performance include, but are not limited to, a lack of understanding of the pragmatics of the assessment situation, unfamiliarity with persons administering the test, the tendency to perseverate and repeat an action or response and frustration during difficult tasks (Tager-Flusberg, 1999). Thus, an obvious advantage of informant measures over researcher- or clinician-administered procedures involves their ability to utilize information that has been accumulated by the informant over time (McCauley, 2001) using questions that span a range of ToM knowledge areas.

Furthermore, standard ToM tasks suffer from ceiling effects when mind-reading skills are relatively good (e.g., Slaughter & Repacholi, 2003). This is true even for the most advanced ToM tasks (although there are a few notable exceptions (e.g., Happe, 1994)). The paucity of measures with relatively high ceilings has complicated efforts to investigate individual differences in ToM, the development of ToM, and mentalizing in middle and late childhood. The development of a measure that is not plagued by ceiling effects and that can be used appropriately for younger and older children would certainly contribute much to the currently available methods for assessing ToM.

It is also important to note that the explicit nature of traditional ToM tasks (e.g., ''Maxi did not see that the chocolate was moved. Where will Maxi look for the chocolate?'') does not resemble the ways that real life social dilemmas are presented. ''Not only are social demands in naturalistic settings not explicitly formulated as a problem-solving situation, they need to be created and defined as a 'social demand' by the person'' (Klin, 2000, p. 832). Thus, individuals who pass a dichotomous false belief task or who perform well on a ToM task battery may be artificially credited with ToM competence, when in fact there is a continuum of competence that is revealed in daily social dysfunction (Klin, 2000). A measure that is sensitive to fine variations in ToM and that relies on information accrued over time during real world social interaction helps to move us toward assessment of ToM that is socially valid, that adopts a dimensional approach of social cognitive abilities (Klin, 2000), and that may reveal ToM deficits that may be masked by traditional ToM tasks.

Parents as Experts of Children's Knowledge and Abilities

It makes sense that parents (or other primary caregivers) would be experts about their children's abilities *including* their ToM competence. During social interaction, parents can accumulate rich information about the child's mind and develop accurate insights into the child's perspectives. Parents also have numerous opportunities to observe child ToM knowledge as it is applied (or not applied) in a range of real-world contexts. No current measures of ToM functioning take advantage of the knowledge of those who are closest to child. The PCToMM-E is an informant measure. It relies on the important others in the child's life and the familiarity of adults who know the child best and, as such, is desirable from a family-centered perspective (McCauley, 2001).

Involving caregivers as informants and interpreters of their children's behaviors is important because it reflects the growing recognition that caregivers possess expert knowledge regarding their children's abilities, strengths, and weaknesses and, as such, are reliable and invaluable sources of information (Beatson & Prelock, 2002; Crais, 1993; Prelock, 2006; Prelock, Beatson, Contompasis, & Bishop, 1999; Prelock, Beatson, Bitner, Broder, & Ducker, 2003; Roberts-DeGennaro, 1996; Shelton & Stepanek, 1994). Indeed, high correlations have been observed between parents' and professionals' judgments of a child's developmental level (Gradel, Thompson, & Sheehan, 1981) suggesting that caregiver reports can yield valid indices of child functioning. In addition, many informant measures (which might be criticized on the basis of their potential for subjectivity and bias) have endured the scrutiny of rigorous psychometric evaluation (e.g., the *MacArthur Communication Development Inventory* (Fenson et al., 1991)) thereby demonstrating construct validity and the accuracy of the information source.

Psychometric Evaluation of the PCToMM-E

ToM is concerned with the understanding of thoughts and feelings and all mental states in one's self and others as well as the understanding of similarities and differences in mental states across different targets (Miller, 2000). The notion of ToM is broad, multifaceted (Astington, 2005) and subsumes or overlaps with constructs that include, but are not limited to, metarepresentation, pretense (Leslie, 1987), the ability to deceive (e.g., Sodian & Frith, 1992; Sodian, Taylor, Harris, & Perner, 1992), the mental–physical distinction (Baron-Cohen, 1989a), desire and intention (Astington, 1999, 2001; Wellman & Bartsch, 1988), distinctions between appearance and reality (Baron-Cohen, 1989a; Flavell, Green & Flavell, 1986), the causes of emotions in general (e.g., Baron-Cohen, 1991), the notion that seeing leads to knowing (Leslie & Frith, 1988; Perner, Frith, Leslie, & Leekam, 1989), second-order thinking (i.e., understanding embedded mental states; e.g., what Laura thinks Patty thinks; Baron-Cohen,1989b), visual perspective-taking (Leslie & Frith, 1988), affective recognition (Baron-Cohen, 2003; Prior, Dahlstrom & Squires, 1990), empathy (Baron-Cohen, 2003), and the understanding and production of mental state terms (Kazak et al., 1997; Tager-Flusberg, 1992) and speech acts (Astington, 1988; Searle, 1969).

In response to the breadth of the construct, the PCToMM-E was developed to reflect variation in the theoretical orientation and assessment procedures that exist in ToM research with typically developing children and children with ASD. The measure was designed to serve as an index of caregivers' *perceptions* of children's ToM knowledge and, by proxy, children's *actual* ToM knowledge. The developers characterize summated and averaged scores as yielding interval data that reflect a general composite of a child's ToM knowledge based on more specific component variables.

This measure was evaluated for two different populations in two separate studies. Study one (study two is described later) investigated the psychometric properties of the PCToMM-E when completed by primary caregivers of children with ASD. First, the test–retest reliability of the

PCToMM-E was examined. The measure's criterion-related construct validity was then explored by evaluating the strength of the association between scores on the PCToMM-E and verbal mental age (VMA) because VMA has been implicated as a strong predictor of ToM abilities in children with ASD (e.g., Happe, 1995; Leekam & Perner, 1991; Ozonoff & Miller, 1995). In addition, demonstrations of a measure's construct validity involve its ability to predict performance thought to be related to the construct of interest. As such, PCToMM-E scores were expected to be correlated with children's scores on ToM tasks.

Because it is an informant measure, the validity of the PCToMM-E relies on the accuracy of informants. No research to date has examined caregivers' ability to predict child ToM abilities. To explore this, parents of children with ASD were administered a measure (i.e., the Parental Predictive Measure of Child ToM Ability, described below) which asked parents to predict their child's performance on a ToM task battery. If caregivers are good at predicting child performance on such tasks, this bodes well for the PCToMM-E which relies on accuracy of knowledge. VMA and average amount of time that caregivers spent with their child per day were also examined to determine whether these factors related to accuracy of judgments.

Study 1—Method

Participants

Twenty birth parents (19 mothers and one father) and their children (three females, 17 males) who had been previously diagnosed with ASD were the participants. Inclusion criteria required parents to identify as the primary caregiver who reported spending, on average, 7.5 h per day (range = 4.2–12.4, SD = 2) with the child not counting the time the child was sleeping. On average, parents reported completion of 16.5 years of formal education (equivalent to a bachelor's degree; range = 12–20, SD = 2.7) and an annual combined household income of $65,000 (range = 12,500–92,500, SD = 18,000). No parent identified as ever having a developmental disability and all were native English speakers.

Children ranged in age from 4 to 12 years (M = 7.6, SD = 2.2). Children also represented a range of verbal abilities assessed on the basis of case history, the *Autism Observation Diagnostic Schedule* (ADOS; Lord, Rutter, DiLavore, & Risi, 1999), and a measure of receptive vocabulary (the *Peabody Picture Vocabulary Test -III*; described below) yielding age equivalent scores (M = 7.4, range = 1–19, SD = 4.4). For some analyses, it was necessary to dichotomize children on the basis of verbal abilities. Eleven children were identified as having relatively good verbal abilities. Of these 11 children, eight achieved receptive vocabulary scores indicating age appropriate levels. Two children scored one standard deviation above the mean and one child scored two standard deviations above the mean. The remaining nine children were identified as having limited verbal abilities. Of these nine children, five obtained receptive vocabulary scores falling at least one standard deviation below the mean. An additional four were identified as functionally nonverbal and obtained receptive vocabulary scores falling at least three standard deviations below the mean.

Measures

The Experimental Perceptions of Children's Theory of Mind Measure (PCToMM-E)

The PCToMM-E consists of 33 statements accompanied by a response continuum of 20 metric units (equivalent to 6.75 inches) anchored by 'definitely not' and 'definitely' with a center point of 'don't know.' The metric units corresponded to a standard engineering 30 feet per inch scale. This metric was chosen because it adequately spanned the width of a standard 8 1/2 × 11 page, provided ample room for all anchors, and yielded round values corresponding to the center (i.e., 10) and extreme points along the continuum (i.e., 0 and 20) which aids in interpretation. Respondents are instructed to carefully read each statement and place a hash mark on the continuum indicating the degree of certainty to which they feel each statement is true for the target child. In the case of uncertainly, respondents are encouraged to reflect on everything they know of the child or to think about how the child would respond in a given situation. The continuum and hash mark response arrangement were chosen to encourage respondents to conceive of certainty in a continuous rather than categorical manner and therefore reflect greater sensitivity than a categorical response arrangement. Responses for each item are scored by a ruler (possible range = 0–20) and averaged with higher values reflecting greater degrees of certainty that the target child possesses ToM knowledge across the range of content surveyed (see Appendix A for an example of the continuum).

Content of the PCToMM-E was guided by a review of the ToM literature. Items were developed to reflect the diverse theoretical perspectives noted previously and a primary goal was to represent the construct of ToM in the broadest sense. Thus, research with both typically developing children and children with ASD was considered.

Following the literature, each item was developed so that it was a face valid indicator of child knowledge of, or ability to engage in, one or some of the following: pretense, desire and intentionality, distinctions between appearance and reality, causes of emotions, mental–physical distinctions, knowledge that seeing leads to knowing, first- and second-order thinking, visual perspective-taking, affective recognition, empathy, social and logical inferencing, and the comprehension and production of mental state terms and speech acts (see Appendix B for sample items). In addition, items were generated which paralleled popular ToM tasks. That is, items were developed which essentially asked informants to rate the likelihood that a target child would pass a standard ToM task (e.g., Appendix B, item 13).

Items were developed by one expert in ASD and ToM, one expert in ToM and test development, and a student of each. The measure was subsequently revised based on - review by another expert in ToM and child development. Care was taken to ensure that all items contained only one thought, that the language used was relaxed and easily understood, and that the items provided proper coverage of the broad construct while avoiding irrelevant content. During item development, a matrix was generated to catalogue which aspects of ToM were believed to be tapped by each item. For example, the item ''My child understands that when I show fear, the situation is unsafe or dangerous'' was designed to tap first-order thinking and affective recognition whereas the item ''My child understands that when a person promises something, it means the person will do it'' was designed to tap first-order thinking, intentionality, and knowledge of speech acts. All of the aforementioned components of ToM were represented by at least (and typically more than) two items.

ToM Task Battery

To evaluate the convergent validity of the PCToMM-E, 12 of 16 items on a ToM task battery that were found to have good test–retest reliability (Hutchins, Prelock, & Chace, in review) were administered. The task battery borrowed from previously developed tasks (i.e., Hadwin, Baron-Cohen, Howlin, & Hill, 1996; Mitchell, Saltmarsh, & Russell, 1997; Silliman, Diehl, Hnath-Chisolm, Bouchard-Zenko, & Friedman, 2003), consisted of 12 test questions within seven tasks, and tapped a range of content and complexity levels.

All tasks made use of a picture storybook format and static visual stimuli. The first task targeted the ability to identify emotions associated with four different facial expressions (i.e., happy, sad, mad, scared). In the second task, children were asked to infer an emotion based on a desire. The third task assessed more advanced abilities involving the inference of belief-based emotion, reality-based emotion, and second-order belief-based emotion. The fourth task targeted the ability to infer a perception-based belief. The fifth task made use of the classic false-belief change-location task. The sixth task assessed the ability to infer a desire-based belief in the context of a change of location and the seventh task was a second-order false belief task. Thus, the ToM task battery was designed to asses a range of content and complexity levels across social and cognitive domains (for a complete description of the task battery, see Hutchins et al., in review).

In this battery, all tasks are accompanied by color illustrations and most incorporate a narrative scenario. This ToM task battery was developed specifically for children with ASD, thus, the content and response arrangements were designed to clarify the task and avoid penalizing nonverbal children with ASD by allowing them to either answer verbally or to point to a picture that showed the correct answer. In addition, illustrations with accompanying text were presented in isolation in order to avoid the potential influence of sensory distraction. Most, but not all, of the tasks included control questions. Children had to pass control questions in order to earn credit for test questions with a total of 12 points possible.

Parental Predictive Measure of Child ToM Ability

The Parental Predictive Measure of Child ToM Ability was designed to assess parental confidence in child performance on the ToM task battery. This measure consisted of 12 statements (e.g., ''When asked where Anthony will look for the book, my child will answer correctly'') based on the 12 test questions within the seven tasks directly corresponding to the ToM task battery described above. Consistent with the format of the PCToMM-E, each question was accompanied by a response continuum of 20 metric units (as described above) and the following anchors: 'definitely not,' 'probably not,' 'probably' and 'definitely' with a center point (10 cm) of 'don't know'. Parents were first familiarized with the ToM task battery and then instructed to read each statement and place a hash mark on the continuum indicating their confidence in their child's ability to correctly answer each test question.

Peabody Picture Vocabulary Test-III (PPVT-III)

The PPVT-III (Dunn & Dunn, 1997) is a popular measure of receptive vocabulary that is commonly used in autism research as an index of verbal mental age (VMA). Although the PPVT-III has demonstrated good psychometric validation (Williams & Wang, 1997), these estimates have been developed for typically developing children. Moreover, caution must be exercised when interpreting age-equivalent scores as these scores should not be taken to mean equivalent functioning. Nonetheless, previous research on ToM performance in ASD has consistently reported a strong relationship between VMA and ToM performance (e.g., Happe, 1995; Leekam & Perner, 1991; Silliman et al., 2003). Therefore, to allow examinations of criterion-related validity, VMA based on PPVT-III performance was considered.

Procedure

Participants were recruited via contact letters from local support agencies for families of children with ASD, directors of special education, and speech-language pathologists. All data were collected as part of a larger experimentally-controlled longitudinal study examining the effects of social story intervention for remediating the core deficits of ASD in nonverbal and verbal children. Only preintervention data were used in these analyses. Case histories were taken approximately one week prior to the assessments described here to diminish the potentially detrimental effects of fatigue. On the day of testing, mothers completed the Predictive Measure of ToM Ability and PCToMM-E prior to the administration of the ToM task battery and the PPVT-III to the children. These data were collected on the same day in the participants' home by graduate students in communication sciences. The ADOS was administered in the participants' homes at a later date (between 3 and 13 weeks) to ensure a proper diagnosis of ASD.

The PCToMM-E was also administered at later dates to provide estimates of test–retest reliability. For the first analysis, the measure was administered at pre-intervention (described above) and approximately one week later (M = 8 days, range = 5–13 days). For a second analysis, data from pre- and post-control phases of study were compared. Over the course of the control phase in the larger study, children were read three different stories that were selected on the basis of their language level appropriateness. In addition, all stories were carefully scrutinized for the number of references to mental states (e.g., happy, sad, remember). Only stories with very few or no mental state references were used. As a result, the content of the books were largely educational in nature (e.g., books about planets or animals). During the control phase, children were read stories three times a week for approximately eight weeks. The average lag between administrations of the PCToMM-E was 109 days or approximately 3.5 months (range = 89 – 154 days). All retesting was also conducted in participants' home.

Study 1—Results

Descriptive Statistics for the PCToMM-E

The average score on the PCToMM-E for primary caregivers of children with ASD was 11.6 (range = 6.8–16.9; SD = 2.8) out of a possible total of 20. Because some comparisons were conducted on the basis of child verbal abilities, descriptive data for this variable are also offered. The average score of respondents for children with limited verbal abilities was 10.2 (range = 6.8–13.8; SD = 2.6) whereas the average score of respondents for children with average and above average verbal abilities was 12.2 (range = 8.7–16.3; SD = 2.5).

Test–retest Reliability

Reliability for a Short Interval

Data for the one-week test–retest interval were missing for two respondents who failed to complete the PCToMM-E at time two. Thus, data for 18 parents who completed the PCToMM-E at time two provided estimates of test–retest reliability. A Pearson's product moment correlation indicated strong stability of this measure over a one week interval ($r = .94$, $p < .001$) which is a highly dependable relationship.

Reliability for a Longer Interval

Pre- and post-control data for the 10 children who participated in the control phase of the larger study previously described were compared. A Pearson's product moment correlation indicated strong stability of this measure over the 3.5 month interval ($r = .89$, $p < .01$) which is also a highly dependable relationship.

Validity

Criterion—related validity

Because of its importance as a predictor of ToM performance, we reasoned that a construct valid measure of ToM would be positively associated with children's VMA. As expected, a Pearson's product moment correlation indicated a substantial positive relation ($r = .61$, $p < .05$) with the PCToMM-E accounting for approximately 37% of the variation in children's VMA. A construct valid measure of ToM competence should also be positively correlated with children's scores on ToM tasks. Descriptive analyses of child ToM task battery performance (construed as ordinal data because a linear relationship between the score and the construct should not be assumed) indicated that the median number of tasks performed correctly was 5.5 (range 0–12). A Spearman's rho indicated a substantial positive relationship ($r = .67$, $p < .05$) with variation in scores on the PCToMM-E explaining approximately 45 percent of the variation in children's scores on the ToM task battery.

To explore parents' ability to predict (and provide accurate comment on) children's ToM abilities, all parents were asked to predict their child's performance on the ToM task battery. A Spearman's rho correlation was conducted revealing that parents' scores on the Predictive Measure of ToM Ability related strongly to children's ToM task battery scores ($r = .73$, $p < .01$) with parental prediction accounting for approximately 53% of the variation in child ToM task battery scores.

To examine factors related to accuracy of parental prediction, a coding rule was developed to operationalize accuracy. In the case that a parent's score fell at or above 10.1 cm on the continuum (i.e., were more confident of a correct response) and the child answered the test question correctly, this was deemed a correct prediction. If the child did not answer correctly, this was deemed an incorrect prediction. Likewise, when a parent's score fell at or below the 10 cm (midpoint) and the child did not answer the test question correctly, this was deemed a correct prediction and if the child did answer correctly, this was deemed an incorrect prediction. On average, parents' correctly predicted child pass and fail performance on 76% (range in a negatively skewed distribution = 38.5–92.3%) of the items (or 9 of the 12 items).

To investigate whether accuracy of parental judgments differed between parents of children with different verbal abilities, the percent of ToM task battery items correctly predicted by parents of children with limited verbal abilities were compared to parents of children with average and above average verbal abilities. An independent *t*-test revealed no difference in parental accuracy by group ($M = 76\%$ for both groups). Consistent with this, a point-biserial correlation revealed no relation between parental accuracy and child VMA, however, a correlation was found between parental accuracy and the amount of time spent daily with the child ($r = .55$, $p < .05$).

Study 1—Discussion

Study one explored the psychometric properties of the PCToMM-E when completed by parents of children with ASD. The PCToMM-E performed well under tests of criterion-related construct validity. Because it is such a good predictor, VMA estimates were expected to converge with PCToMM-E scores and this strong positive relationship was borne out in the data. The nature of the relationships between VMA and ToM knowledge is beyond the scope of this discussion. At present, child VMA appears to play a critical role in the performance on standard ToM tasks and it is related to the quality of judgments generated by primary caregivers of children with ASD who completed the PCToMM-E.

Crucially, the PCToMM-E proved to be a good predictor of child ToM task battery performance. This provides further support for the criterion-related construct validity of the PCToMM-E and rationale for its use particularly when standard ToM tasks are difficult or impossible to administer due to motivation and situational factors. In a related vein, parents were generally good at predicting children's performance on standard ToM tasks. This is not surprising from a family-centered perspective in which parents are regarded as experts on their children and are, therefore, valuable sources of information about child functioning (e.g., Prelock et al., 1999, 2003). The results also indicated, however, that some parents may be more 'expert' than others and this was contingent, to some extent, on the amount of time that parents spent with children. It is not surprising that parents who are able to spend more time with a child would provide more accurate information about that child's ability. More time spent with a child provides more opportunities to participate in dyadic interaction, to observe the child's social, language, and behavioral functioning across contexts, and to more generally develop insights into the child's ToM. Further research is needed to clarify how validity of the PCToMM-E is related to variables like time spent with child. At present, we caution that this tool is most appropriately applied to parents who spend a minimum average of 5 hours per day with the child (not counting the time the child is sleeping) as this represented the lower bound for parents in this study. We further question the use of this tool for caregivers who have not completed a high school education, who have been identified as having a developmental disability, or who are non-native speakers of English as these populations were not represented.

The findings also support the use of the PCToMM-E for parents of children with ASD with limited verbal abilities. Child VMA did not relate to parents' accuracy in predicting child performance on ToM tasks and parental accuracy of prediction did not differ as a function of child verbal ability. This suggests that ToM informant measures like the PCToMM-E may serve as valid indicators of parents' perceptions of children's ToM and children's actual ToM competence even among parents of children with the most limited verbal capacities. As such, the PCToMM-E may prove particularly useful as a means of ToM assessment for children with ASD who have traditionally been excluded from much research that makes use of standard ToM tasks.

It is noteworthy that no ceiling effects were observed in this sample of parents of children with ASD who represented a wide range in verbal abilities. Importantly, even the oldest children in our sample with the most advanced language abilities (and who incidentally approached or hit the ceiling on the ToM task battery) did not approach the ceiling on the PCToMM-E suggesting that this tool may have great utility for this population. Of course, ceiling effects are of the greatest concern when mind-reading abilities are relatively good. In addition, the PCToMM-E demonstrated good test–retest reliability at short and longer lags; however, one might expect temporal stability of the measure to be even greater among parents of typically developing children whose cognitive and social behaviors are likely to evidence less fluctuation. For these reasons, it is important to extend investigations of the PCToMM-E to typically developing samples. This would also allow examinations of criterion-related construct validity for this population on the basis of additional specific comparisons.

Study 2—Purpose

Study two extended investigations of the PCToMM-E by exploring some of its psychometric properties when completed by primary caregivers of typically developing children. An examination of test–retest reliability was followed by tests of construct validity. Although not without controversy, the age of four years is often considered a watershed for ToM development in typically developing children. In a meta-analysis of 178 studies employing a false belief task, Wellman et al. (2001) found that children under age 3.5 years typically perform below chance, children between the age of 3.5 and 4 years typically perform at chance, and children four and older typically perform above chance. Using a contrasting-groups developmental method of construct validation, PCToMM-E scores provided by parents of typically developing children ages 2.5 –3.5 years were compared to those of children ages 4.0–5.0 with the expectation that the latter would be significantly higher.

Of course, ToM continues to develop beyond the age of 5 years (although these investigations are rare). For this reason, PCToMM-E scores were expected to correlate more generally with age among children ranging in age from 2 to 12 years. Scores were also scrutinized for evidence of ceiling effects associated with age. Finally, because ToM deficits represent a core characteristic of ASD, a construct valid measure of ToM competence should distinguish between scores observed for informants of age-matched typically developing children and children with ASD such that the former should be significantly higher.

Study 2—Method

Participants

Participants were 60 mothers who were primary caregivers of a total of 72 typically developing children (36 females, 36 males) between the ages of two- and twelve-years ($M = 6.63$; SD = 2.83). The average number of hours per day that mothers spent with children (not counting the time the child was sleeping) ranged from three to 14 h ($M = 7.64$; SD = 2.7). On average, mothers reported completion of 16.7 years of formal education (range = 12–20, SD = 1.85) and an annual combined household income of \$80,000 (range = \$25,000–\$115,000, SD = \$18,000). No mother identified as ever having a developmental disability and all were native English speakers.

Procedure

Participants were recruited by circulating a recruitment letter to local preschools and elementary schools and to faculty in the psychology and communication sciences departments at a local university as well as through informal personal contacts. The PCToMM-E and a brief demographic questionnaire took approximately 10 minutes to complete. Half of all respondents (randomly selected) were provided an additional copy of the PCToMM-E and asked to complete it one week later (range 5–18 days). Respondents who complied then dated and returned this second copy in a self-addressed stamped envelope.

Study 2—Results

Descriptive Statistics for the PCToMM-E

The average score on the PCToMM-E for primary caregivers of typically developing children was 17 (range = 9.3–20; SD = 1.5). Given the wide range of child age surveyed here, a more informative description of mean scores by age is warranted. These data are presented in Fig. 1.

Although these data are based on sometimes small and uneven sample sizes, they indicate that the PCToMM-E did not evidence ceiling effects until late childhood. Although scores approach the ceiling among children ages 9 and 10 years, a true ceiling effect does not appear to be evident until age 11.

Test–retest Reliability

Thirteen mothers (37% response rate) of 17 children completed and returned the measure at time two to provide estimates of test–retest reliability. The respondents closely

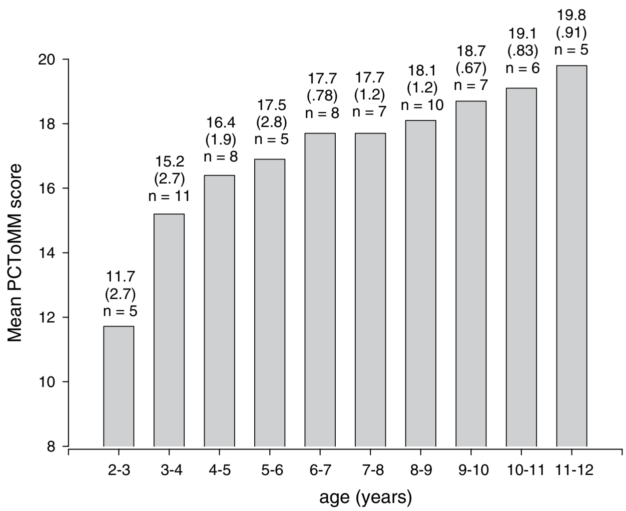


Fig. 1 Mean (standard deviation) for PCToMM-E scores by child age

represented the larger sample with regard to child age (range = 2.4–10.9 years; $M = 6.6$), number of hours spent with the child daily (not counting the time the child was sleeping (range = 3–14, $M = 7.5$), and annual combined household income (range \$35,000–\$95, 000, M = \$82,000). Data for the 17 responses indicated high test–retest reliability ($r = .98$, $p < .01$) with variation in scores at time one explaining approximately 96% of the variation in scores at time two.

Validity

Contrasting-groups Developmental Method of Construct Validation: Comparison Based on Child Age

For the total number of PCToMM-E measures completed ($n = 72$), 10 were completed by mothers of children age 2.5– 3.5 years and 9 were completed by mothers of children age 4.0–5.0. An independent t-test revealed significant differences in PCToMM-E scores, t (17) = 4.18, $p < .01$, such that mothers of younger children evidenced significantly lower scores ($M = 12.3$, SD = 2.7) than did mothers of older children ($M = 16.4$, SD = 2.1). For the total 72 measures completed, a Pearson's correlation also revealed a signification relationship between child age and PCToMM-E score ($r = .68$, $p < .01$) such that variation in child age accounted for approximately 46% of the variation in PCToMM-E scores.

Contrasting-Groups Method of Construct Validity: Comparison Based on Child Developmental Status

Data for 21 measures were omitted from the sample described in study two to create a group equivalent to the sample described in study one with respect to child age. In this way, scores for the remaining 51 PCToMM-Es completed by parents of typically developing children (mean age = 7.7; range = 4–12) were compared to those of 20 parents of children with ASD (mean age = 7.6; range = 4–12). As hypothesized, an independent t-test revealed a significant difference, t (69) = 11.6, $p < .001$, such that mothers of children identified as having ASD reported lower scores ($M = 11.77$; SD = 1.4) than did mothers of typically developing children ($M = 17.97$; SD = 3).

Study 2—Discussion

Study two explored some of the psychometric properties of the PCToMM-E when completed by mothers of typically developing children. The PCToMM-E performed well under examinations of test–retest reliability. It should be noted that the test–retest interval was relatively small (approximately one week) and we did not assess the test–retest reliability of the PCToMM-E when completed by caregivers of typically developing children for longer lags. It is worthwhile to explore the test–retest reliability of the measure over longer intervals for this population while keeping the interval short enough to ensure little or no ToM development.

The PCToMM-E performed well under examinations of construct validity. In line with decades of previous research indicating a developmental shift in early childhood (e.g., Wellman et al., 2001), the PCToMM-E discriminated between typically developing children ages 2.5–3.5 and 4.0–5.0 years of age. The PCToMM-E also discriminated between informants of age-matched typically developing children and children with ASD supporting previous research in this area (e.g., Baron-Cohen, 1989a; Flavell et al., 1986). Further, PCToMM-E scores correlated more generally with age supporting the expectation that ToM continues to develop beyond early childhood.

Moreover, the PCToMM-E did not evidence ceiling effects until late childhood (i.e., approximately age eleven). Thus, the PCToMM-E may prove to be an important contribution to existing ToM assessment methods, the majority of which suffer from low ceilings, because it is particularly well-suited to test assumptions about the development of ToM across a relatively wide range of ages. The PCToMM-E may also be preferred for use as a repeated measure. This is important when considering that scores based on standard ToM tasks typically evidence increases when administered at a second point in time (Hutchins et al., in review; Mayes, Klin, Tercyak, Cicchetti, & Cohen, 1996) and although the PCToMM-E may be vulnerable to demands (which can be addressed experimentally and procedurally), it is not vulnerable to test practice effects.

General Discussion

Two studies explored the psychometric properties of the PCToMM-E when completed by the primary caregivers of children with ASD and typically developing children. The PCToMM-E demonstrated excellent test–retest reliability for both samples. This gains importance in light of the aforementioned motivational and situational variables that are likely to generate 'noise' in standard ToM task performance. The strong test–retest estimates suggest that our sample of parents of typically developing children and children with ASD relied on relatively stable conceptions of child ToM knowledge making the PCToMM-E a tool that may be useful in research and practice.

Indeed, for all examinations of the measure's reliability and validity, the PCToMM-E performed well. Nonetheless, some limitations of the aforementioned studies deserve mention. Although we maintain that the PCToMM-E has good content validity, in retrospect, items are currently lacking that tap comprehension and production of sarcasm and irony (Baron-Cohen, 1997), the use and appreciation of humor (St. James & Tager-Flusberg, 1994), the ability to engage in creative imagination (Craig, 1997), and counterfactual reasoning (Perner, Sprung, & Steinkogler, 2004). Items sensitive to these aspects of ToM will be included in a subsequent version of the measure. Because these aspects arguably represent more advanced ToM knowledge, inclusion of such items may also be effective for raising the ceiling observed for respondents of typically developing children even higher.

In addition, too few subjects participated in either study to allow for factor analysis and this is an important next step in the evaluation of the PCToMM-E. At present, the measure is a simple summated and averaged scale and responses to each item are given equal weight. Factor analysis could be conducted to determine whether items would most appropriately be weighted differentially and it is necessary for determining whether and which subscales may comprise the measure. This is important from theoretical and clinical perspectives to identify the dimensions of ToM understanding that may be targeted for research and intervention.

The PCToMM-E appears to be a reliable and valid measure of ToM competence that can be used in concert with, or in some cases, in lieu of standard ToM tasks. Standard ToM tasks can be difficult to administer especially to children with ASD. When they are used successfully, they reduce the complex construct of ToM to a categorical response. Even when ToM task batteries are used, scores are most appropriately construed as ordinal in nature. They are also prone to ceiling effects which reduce their utility for samples with relatively good metarepresentation skills. By contrast, the PCToMM-E is quick and easy to administer, yields interval data, may be used as a repeated measure, and may provide finer levels of discrimination across a range of developmental and skill levels. It may also be particularly useful as a socially valid measure to be used in the context of intervention studies targeting ToM. Given the good psychometric properties demonstrated by the PCToMM-E thus far, this study provides strong support for further examination and development of the PCToMM-E as a research and clinical tool.

Acknowledgments Portions of this paper were presented at the annual conference of the American Speech, Language, and Hearing Association (November, 2006, Miami, FL). Data from study 2 were originally collected and analyzed by the second author as part of an undergraduate honor's thesis.

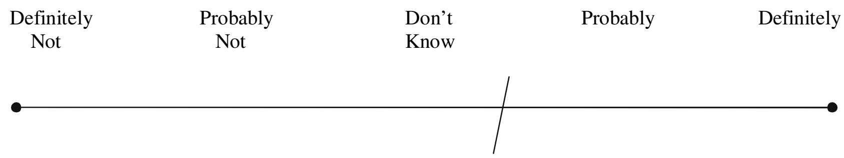
Appendix A: Instructions.

The Perception of Children's Theory of Mind Measure (PCToMM-E)

The purpose of this measure is to learn about caregivers' ideas regarding children's thoughts and feelings. Please read each statement carefully and indicate the degree to which you believe each statement is true for your child. Indicate your response by making a vertical hash mark at the appropriate point along the continuum. You may feel that you don't know for sure whether a statement is true or not. When you feel this way, reflect upon your experiences with your child and try to decide, given everything you know about this child, how certain you are that the statement is true or not true. There are no right or wrong answers and no answers are valued over any other answers. Please remember to respond as honestly and thoughtfully as possible. Your answers are completely confidential.

EXAMPLE: Read the following statement and indicate your response by making a vertical hash mark along the appropriate point on the continuum. If you don't know the answer to the question, make a slash mark somewhere underneath ''Don't Know.'' If you have more definite feelings that the statement is true or not true, make a hash mark along the point in the continuum that reflects those feelings.

Example: My child can communicate to me that s/he wants something.



Appendix B: Sample items from the PCToMM-E

1. My child understands that when someone says they are afraid of the dark, they will not go into a dark room.
2. My child understands that to know what is in an unmarked box, you have to see or hear about what is in that box.
3. My child understands that when people get what they want, they will be happy.
4. My child can pretend that one object is a different object (for example, pretending a banana is a telephone).
5. My child empathizes with others.
6. My child understands whether someone hurts someone else on purpose or by accident.
7. My child understands that when people frown, they feel differently then when they smile.
8. My child understands that, when I show fear, the situation is unsafe or dangerous.
9. My child uses the word 'know' as in ''I know it.''
10. My child understands that if two people look at the same object from a different standing point, they will see the object in different ways.
11. My child understands that when a person promises something, it means the person will do it.
12. My child understands that people often have thoughts about other peoples' thoughts.
13. If I put my keys on the table, left the room, and my child moved the keys from the table to a drawer, my child would understand that when I returned, I would first look for my keys where I left them.
14. My child understands that peoples' personalities basically don't change from day to day.
15. My child understands that people can be wrong in what they think other people want.
16. My child understands that people might not always say what they are thinking because they don't want to hurt others' feelings.

References

Astington, J. W. (1988). Promises: Words or deeds? *First Language, 8*, 259–270.

Astington, J. W. (1999). The language of intention: Three ways of doing it. In P. D. Zelazo, J. W. Astington, & D. R. Olson (Eds.), *Developing theories of intention: Social understanding and self-control* (pp. 295–315). Mahwah, NJ: Lawrence Erlbaum.

Astington, J. W. (2001). The future of theory-of-mind research: Understanding motivational states, the role of language, and real-world consequences. *Child Development, 72*, 685–687.

Astington, J. W. (2005). Introduction: Why language matters. In J. W. Astington & J. A. Baird (Eds.), *Why language matters for theory of mind* (pp. 3–25). New York: Oxford University Press.

Astington, J. W., & Jenkins, J. M. (1995). Theory of mind development and social understanding. *Cognition & Emotion, 9*, 151–165.

Baron-Cohen, S. (1989a). Are autistic children ''behaviorists''? An examination of their mental-physical and appearance–reality distinctions. *Journal of Autism & Developmental Disorders, 19*, 579–600.

Baron-Cohen, S. (1989b). The autistic child's theory of mind: A case of specific developmental delay. *Journal of Child Psychology & Psychiatry, 30*, 285–298.

Baron-Cohen, S. (1991). Do people with Autism understand what causes emotion? *Child Development, 62*, 385–395.

Baron-Cohen, S. (1992). Out of sight or out of mind? Another look at deception in autism. *Journal of Child Psychology & Psychiatry, 33*, 1141–1155.

Baron-Cohen, S. (1997). Hey! It was just a joke! Understanding propositions and propositional attitudes by normally developing children and children with autism. *Israel Journal of Psychiatry, 34*, 174–178.

Baron-Cohen, S. (2000). Theory of mind in autism: A fifteen year review. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental neuroscience* (pp. 3–20). Oxford, UK: Oxford University Press.

Baron-Cohen, S. (2003). *The essential difference: The truth about the male and female mind*. New York: Basic Books.

Beatson, J. E., & Prelock, P. A. (2002). The Vermont Rural Autism Project: Sharing experiences, shifting attitudes. *Focus on Autism & Other Developmental Disabilities, 17*(1), 48–54.

Begeer, S., Rieffe, C., Meerum Terwogt, M., & Stockmann, L. (2003). Theory of mind-based action in children from the autism spectrum. *Journal of Autism & Developmental Disorders, 33*, 479–487.

Craig, J. (1997). *An investigation of imagination and creativity in autism*. Cambridge, MA: Cambridge University Press.

Crais, E. R. (1993). Families and professionals as collaborators in assessment. *Topics in Language Disorders 14*, 29–40.

Dunn, L. M., & Dunn, L. M. (1997). *Examiner's manual for the Peabody Picture Vocabulary Test* (3rd ed.). Circle Pines, MN: American Guidance Service.

Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J., Penthick, S., & Reilly, J. (1991). *Technical manual for the MacArthur Communicative Development Inventories*. San Diego, CA: San Diego State University.

Flavell, J. H., Green, E. R., & Flavell, E. R. (1986). Development of knowledge about the appearance–reality distinction. Monographs of the Society for Research in Child Development, 5 (1).

Gradel, K., Thompson, M., & Sheehan, R. (1981). Parental and professional agreement in early childhood assessment. *Topics in Early Childhood and Special Education, 1*, 31–39.

Hadwin, J., Baron-Cohen, S., Howlin, P., & Hill. K. (1996). Can we teach children with autism to understand emotions, belief, or pretence? *Development & Psychopathology, 8*, 345–365.

Happe, F. (1994). An advanced test of theory of mind: Understanding the story characters thoughts and feelings by able autistic mentally handicapped and normal children and adults. *Journal of Autism & Developmental Disorders, 24*, 129–154.

Happe, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development, 66*, 843–855.

Hughes, C., Adlam, A., Happe, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test–retest reliability for standard and advanced false-belief tasks across a wide range of abilities. *Journal of Child Psychology & Psychiatry, 41*, 483–490.

Hutchins, T. L., Prelock, P., & Chase, W. (in review). Test–retest reliability of ToM tasks for children with ASD. *Focus on Autism and Other Developmental Disabilities.*.

Kazak, S., Collis, G. M., & Lewis, V. (1997). Can young people with Autism refer to knowledgestates? Evidence from their understanding of ''know'' and ''guess''. *Journal of Child Psychology & Psychiatry, 38*, 1001–1009.

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, 7*, 831–846.

Lalonde, C. E., & Chandler, M. J. (1995). False belief understanding goes to school: On the social-emotional consequences of coming early or late to a first theory of mind. *Cognition & Emotion, 9*, 167–186.

Leslie, A. M. (1987). Pretense and representation: The origins of ''theory of mind''. *Psychological Review, 94*, 412–426.

Leslie, A. M., & Frith, U. (1988). Autistic children's understanding of seeing, knowing and believing. *British Journal of Developmental Psychology, 6*, 315–324.

Leekam, S., & Perner, J. (1991). Does the autistic child have a metarepresentational deficit? *Cognition, 40*, 203–218.

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. (1999). *Autism Diagnostic Observation Schedule-WPS Edition*. Los Angeles: Western Psychological Services.

Mayes, L., Klin, A., Tercyak, K. P., Cicchetti, D. V., & Cohen, D. J. (1996). Test–retest reliability for false-belief tasks. *Journal of Child Psychology and Psychiatry, 37*, 313–319.

McCauley, R. (2001). *Assessment of language disorders in children*. Mahwah, NJ: Lawrence Erlbaum.

Miller, S. A. (2000). Children's understanding of preexisting differences in knowledge and belief. *Developmental Review, 20*, 227–282.

Miller, C. A. (2006). Developmental relationships between language and theory of mind. *American Journal of Speech-Language Pathology, 15*, 142–154.

Mitchell, P., Saltmarsh, R., & Russell, H. (1997). Overly literal interpretations of speech in autism: Understanding that messages arise from minds. *Journal of Child Psychology and Psychiatry & Allied Disciplines, 38*, 658–691.

Ozonoff, S., & Miller, J. N. (1995). Teaching theory of mind: A new approach to social skills training for individuals with autism. *Journal of Autism & Developmental Disorders, 25*, 415–433.

Perner, J., Frith, U., Leslie, A. M., & Leekam, S. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development, 60*, 689–700.

Perner, J., Sprung, M., & Steinkogler, B. (2004). Counterfactual conditionals and false belief: A developmental dissociation. *Cognitive Development, 19*, 179–201.

Prelock, P. A. (2006). *Autism spectrum disorders: Issues in assessment & intervention*. Austin, TX: Pro-Ed Publishers.

Prelock, P. A., Beatson, J., Bitner, B., Broder, C., & Ducker, A. (2003). Interdisciplinary assessment for young children with Autism Spectrum Disorders. *Language, Speech and Hearing Services in Schools, 34*, 194–202.

Prelock, P. A., Beatson, J., Contompasis, S., & Bishop, K. KI. (1999). A model for family-centered interdisciplinary practice in the community. *Topics in Language Disorders, 19*, 36–51.

Prior, M., Dahlstrom, B., & Squires, T. L. (1990). Autistic children's knowledge of thinking and feeling in other people. *Journal of Child Psychology & Psychiatry, 31*, 587–601.

Roberts,-DeGenarro, M. (1996). An interdisciplinary training model in the field of early intervention. *Social Work in Education, 18*, 20–29.

St. James, P. J., & Tager-Flusberg, H. (1994). An observational study of humor in autism and Down syndrome. *Journal of Autism & Developmental Disorders, 24*, 603–617.

Searle, J. R. (1969). *Speech acts: An Essay in the Philosophy of Language*. Cambridge, MA: Cambridge University Press.

Shelton, T. L., & Stepanek, J. S. (1994, September). *Family-centered care for children needing specialized health and developmental services* (3rd ed.). Bethesda, MD: Associations for the Car of Children's Health.

Silliman, E. R., Diehl, S. F., Hnath-Chisolm, T., Bouchard Zenko, C., & Friedman, S. A. (2003). A new look at performance on theory-of-mind tasks by adolescents with Autism Spectrum Disorder. *Language, Speech and Hearing Services in Schools, 34*, 236–252.

Slaughter, V., & Repacholi, B. (2003). Introduction: Individual differences in theory of mind. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind* (pp. 1–12). New York: Taylor & Francis Books.

Sodian, B., & Frith, U. (1992). Deception and sabotage in autistic, retarded, and normal children. *Journal of Child Psychology & Psychiatry, 33*, 591–606.

Sodian, B., Taylor, C., Harris, P., & Perner, J. (1992). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development, 62*, 468–483.

Tager-Flusberg, H. (1992). Autistic children's talk about psychological states: Deficits in the early acquisition of a theory of mind. *Child Development, 63*, 161–172.

Tager-Flusberg, H. (1993). What language reveals about the understanding of minds in children with autism. In S. Baron-Cohen, H. Tager-Flusberg, & D. H. Cohen (Eds.), *Understanding other minds: Perspectives from autism* (pp. 138–153). New York: Oxford University Press.

Tager-Flusberg, H. (1999). The challenge of studying language development in Children with autism. In L. Menn & N. Bernstein Ratner (Eds.), *Methods for studying language production* (pp. 313–331). Mahwah, NJ: Lawrence Erlbaum.

Tager-Flusberg, H. (2001). A reexamination of the theory of mind hypothesis of Autism. In J. A. Burack, T. Charman., N. Yirmiya., & P. R. Zelazo (Eds.), *The development of autism: Perspectives from theory and research* (pp. 173–193). Mahwah, NJ: Lawrence Erlbaum.

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.

Wellman, H. M., & Bartsch, K. (1988). Young children's reasoning about beliefs. *Cognition, 30*, 239–277.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false beliefs. *Child Development, 72*, 655–684.

Williams, K. T., & Wang, J. (1997). *Technical references to the Peabody Picture Vocabulary Test* (3rd Ed.). Circle Pines, MN: American Guidance Service.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and the constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*, 103–128.

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin, 124*, 283–307.