Technical Manual for the

Theory of Mind Inventory & Theory of Mind Task Battery

Tiffany L. Hutchins, Patricia A. Prelock, &
Laura Bonazinga Bouyea, ©2014

ToMI [revised 3/13/15]

Tiffany Hutchins, Ph.D.

Assistant Professor of Communication Sciences & Disorders
University of Vermont

Patricia A. Prelock, Ph.D., CCC-SLP

Professor and Dean of the College of Nursing and Health Sciences
University of Vermont

Laura Bonazinga Bouyea, M.S., CCC-SLP

Lecturer in the Department of Communication Sciences & Disorders
University of Vermont

Table of Contents

THE THEORY OF MIND INVENTORY (ToMI)

Rationale

'Theory of mind' is a broad and multifaceted construct (Astington & Baird, 2005) that is often used interchangeably with the terms "perspective-taking," "metacognition," "folk psychology," and "social cognition" (Hutchins, Prelock, & Bonazinga, 2012). The term 'theory of mind' has gained considerable currency among researchers and other professionals working in the areas of social cognition and developmental psychology. Although directions of influence are difficult to clarify, many researchers have concluded that theory of mind impairments underlie the social, behavioral, and communicative impairments characteristic of Autism Spectrum Disorders (ASD; e.g., Baron-Cohen, 1995; Baron-Cohen, Leslie, & Frith, 1985). This has made theory of mind relevant—often central—to the study of ASD and the development of effective interventions to support social cognition and more appropriate behaviors in this population.

Research in theory of mind (ToM) has been extremely active over the last 25 years and a wide variety of ToM tests have been developed. Across hundreds of studies, the most common ToM assessment strategy makes use of the Sally-Anne or classic false belief task (Wellman, Cross, & Watson, 2001) developed by Wimmer and Perner (1983). In this task, children are told a story in which an object is moved from an old location to a new location without the knowledge of the main protagonist. For example, Sally puts a marble in a basket and leaves the room. In her absence, Anne enters and moves the marble from the basket to a box and then she leaves. Children are asked, "When Sally returns, where will she look for the book?" Children who answer with the new (incorrect) location fail the question whereas children who answer with the old (correct) location pass the question by presumably demonstrating their knowledge that behaviors are guided by inner mental states, in this case a false belief (Hutchins, Prelock, & Chace, 2008).

The Sally-Anne task has been a valuable tool for examining the ability to attribute false beliefs—one aspect of theory of mind. On the other hand, the use of traditional measures like the Sally-Anne task and other structured elicitation procedures are associated with tremendous difficulties. Notably, traditional measures have relied almost exclusively on direct assessment of child performance. As a result, the child's cognitive and language level can influence performance so as to obscure (or artificially credit) theory of mind knowledge (Klin, 2000). Moreover, motivational factors (e.g., interest level, fatigue, attention) often operate when assessing young children and individuals with ASD (Begeer, Rieffe, Meerum Terwogt, & Stockmann, 2003). Motivational challenges may be compounded by a variety of situational factors that can also impede task performance. These include, but are by no means limited to, a lack of understanding of the pragmatics of the assessment situation, unfamiliarity with persons administering the test, and frustration during difficult tasks (Tager-Flusberg, 1999).

In addition, traditional ToM tasks are scored on a dichotomous pass/fail basis. As Tager-Flusberg (2001) argued, such procedures lead us to construe theory of mind as:

> "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'' (pp. 177–178).

For this reason, several researchers have argued for the value of aggregate measures of ToM in the form of task batteries, which tend to be quite limited in their ability to adequately tap the content domain relevant to ToM (Hutchins et al., 2012; Hutchins, Bonazinga, Prelock, & Taylor, 2008). Another approach has been to develop more content-valid standardized assessments. One such test is the *NEPSY-II* (Neuropsychological Assessment-II; Korkman, Kirk, & Kemp, 2007), which includes a 'social perception' section designed to assess affect recognition and ToM. Although the subtests tap a range of ToM dimensions and solicit either verbal or nonverbal responses, they directly test child performance and like all child performance measures, will be vulnerable to motivation, situational, attentional, linguistic, and cognitive factors.

Another drawback associated with traditional measures of ToM is that ceiling effects are common when social cognitive understanding is relatively good (Slaughter & Repacholi, 2003). This led to the development of several advanced ToM tests. For example, the *Reading the Mind in the Eyes Test* (Baron-Cohen, Wheelright, Hill, Raste, & Plumb, 2001) offers a measure that is sensitive to implicit, non-linguistic aspects of ToM (although it does require verbal skills as respondents are required to understand task instructions). The test consists of a series of photographs of the eye-region of the face and respondents are asked to choose one of four words that best describe what the person is thinking or feeling (e.g., terrified, arrogant, upset, annoyed).

Another well-known advanced measure of ToM is Happé's (1994) *Strange Stories*, which consist of a series of short vignettes accompanied by illustrations. Respondents are read a short story (e.g., about white lies, jokes, misunderstandings) and are asked questions designed to assess their comprehension of the social scenario. Although these advanced measures are innovative methods for assessing ToM in individuals with high functioning autism and Asperger syndrome,

> "interest in people with high-functioning autism can obscure the fact that most people with the disorder have moderate to severe learning difficulties. In classic autism this may be about 75%, and more than half of those affected develop no appreciable language. This means that theory of mind deficits in autism have only been examined in a fraction of sufferers; typically experiments include only children with verbal mental ages of above 4 years" (Doherty, 2009, p.179).

Finally, the explicit nature of many traditional ToM tasks are quite unlike the ways that real life social dilemmas are presented (e.g., Hutchins et al., 2012). In fact, the notion that task performance can exceed social cognitive functioning when it is applied in everyday life has been the topic of considerable concern (Astington, 2003; Davies & Stone, 2003; Klin, Schultz, & Cohen, 2000). This raises important issues about the social validity of ToM assessment.

Development

Our first effort to develop a measure of ToM that would address all of the aforementioned limitations resulted in a 33-item tool named the *Perceptions of Children's Theory of Mind Measure-Experimental version* (PCToMM-E; Hutchins, Bonazinga, et al., 2008). The 33 items were developed on the basis of their ability to be face-valid indicators of a large set of ToM competencies that varied in content and complexity. "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" (italics in original, Hutchins et al., 2012, p. 4).

> In response to the breadth of the construct and guided by the immense literature on ToM in both typically developing children and children with ASD, we sought to develop face valid indicators including (but not limited to) child knowledge of, or ability to engage in pretence, 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, and social and logical inferencing (Hutchins et al., 2012, pp. 4 – 5).

The PCToMM-E was found to have excellent psychometric properties when administered to caregivers of typically developing children (ages 2 – 12) and caregivers of children with ASD (ages 2- 12) (see Hutchins, Bonazinga, et al., 2008). The PCToMM-E has been useful as a research and clinical tool. For example, Hutchins and Prelock (2008) described how this tool could be utilized as part of a larger assessment battery to identify behavioral and social-cognitive targets of intervention that were developmentally appropriate for a young child diagnosed with Autistic disorder. Using a sample of 20 children with ASD who varied widely in their cognitive and linguistic abilities (sample described in Hutchins & Prelock, 2013a, 2013b) we also found the PCToMM-E to be a sensitive pre-post measure of ToM development in the context of an experimentally controlled study designed to assess the effects of a social cognition intervention.

Although the PCToMM-E proved to be a promising and useful measure, it was not without its limitations. Since the development and preliminary psychometric evaluation of the PCToMM-E, revisions have been made to improve the measure. These include 1) changes to the instructions to clarify the nature of the response arrangement, 2) deletion or revision of certain items based on results of statistical analyses (described in Hutchins, Bonazinga, et al., 2008; Hutchins et al., 2012), 3) the addition of items to more adequately tap the wide range of social cognitive understandings that are subsumed within the construct of ToM, 4) additional data collection from caregivers of typically developing children and children with ASD (up to age 20 years) and, 5) additional statistical analyses to examine reliability and validity and to explore the dimensionality of the measure so as to identify any ToM subscales (Hutchins et al., 2012).

Content

In its present form, the ToMI consists of 42 items designed to tap a wide range of social cognitive understandings. Each item takes the form of a statement (e.g., “My child understands whether someone hurts another on purpose or by accident”) and is accompanied by a 20-unit continuum anchored by ‘definitely not’, ‘probably not’, ‘undecided’, ‘probably’, and ‘definitely.’ The respondent is asked to read a statement and draw a hash mark at the appropriate point along the continuum. The continuum and hash mark response arrangement was favored over a more traditional Likert-type scale for its ability to be sensitive to values between anchors and therefore enhance precision. Given that it is a less common response arrangement, however, confusion about how to respond can occur. For this reason, we improved the instructions in a revision of the measure to make clear what is considered correct and incorrect forms of responding.

A formal analysis was also conducted to determine the reading level of the ToMI. According to the Flesch-Kincaid readability index (Kincaid, Braby, & Mears, 1988), the ToMI received a score of 62.2 (scores range from 0-100 with higher scores indicating easier reading) which is equivalent to a reading level of grade 8.9 which should be easily understood by 13-15 year old students.

Each item on the ToMI was developed to serve as a face valid indicator of a particular dimension of ToM. The immense theoretical and empirical research base in ToM guided the content of the ToMI. This involved consideration of the ToM literature for typically developing children (from infancy to late childhood and early adolescence) as well as individuals with ASD from across the autism spectrum (i.e., nonverbal to high functioning with precocious language). A primary goal was to develop a content valid index of ToM that reflected variation in the type and complexity of wide range of ToM understandings. Each of the 42 items comprising the ToMI belong to one of three empirically derived subscales (i.e., Early, Basic, and Advanced) that were determined through the use of principal components Analysis (PCA; described more fully below). The items, and the dimensions tapped by each item, are presented in Table 1. Additional information for using ToMI scores and the dimensions tapped is presented in the clinical applications section of this manual.

Table 1: ToMI items by factors and the dimensions intended to be tapped by each item.

Factor 1: ADVANCED THEORY OF MIND	**Dimension Intended to be tapped**
2. If it were raining and I said in a sarcastic voice “Gee, looks like a really nice day outside,” my child would understand that I didn’t actually think it was a nice day.	sarcasm
5. My child understands that people can be wrong about what other people want.	second-order false desire attribution
13. If I said “Let’s hit the road!” my child would understand that I really meant “Let’s go!”	idiomatic language
14. My child understands that people can lie to purposely mislead others.	use of language to intentionally deceive
17. My child understands that people can smile even when they are not happy.	understanding display rules
18. My child understands the difference between when a friend is teasing in a nice way and when a bully is making fun of someone in a mean way.	complex social judgment
19. My child understands that people don’t always say what they are thinking because they don’t want to hurt others’ feelings.	white lies
20. My child understands the difference between lies and jokes.	understanding lies versus jokes
21. 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.	visual perspective-taking
22. My child understands that people often have thoughts about other peoples’ *thoughts*.	second order understanding of belief
23. My child understands that people often have thoughts about other peoples’ *feelings*	second order understanding of emotion
27. My child recognizes when a listener is not interested.	complex social judgment
34. My child is able to put himself/herself in other people’s shoes and understand how they feel.	empathy
36. If I said “What is black, white and ‘*read*’ all over? It’s a newspaper!” my child would understand the humor in this play on words.	humor (play on words)
40. When we like others, we are likely to interpret their behavior in positive ways and when we don’t like others, we are likely to interpret their behavior more negatively. My child understands that previous ideas and/or opinions of others can influence how we interpret their behaviors.	biased cognition
41. My child understands that two people can see the same image and interpret it differently. For example, when looking at this image, one person might see a rabbit whereas another might see a duck.	mind as active interpreter

Factor 2: BASIC THEORY OF MIND	
1. My child understands that when someone puts on a jacket, it is probably because he/she is cold	physiologically-based behavior
4. My child understands that when someone says they are afraid of the dark, they will not want to go into a dark room.	emotion-based behavior
7. My child understands the word 'think'	mental state term comprehension
8. 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.	false beliefs in context of unexpected change of location
9. My child understands that to know what is in an unmarked box, you have to see or hear about what is in that box.	seeing leads to knowing
10. My child understands the word 'know'.	mental state term comprehension
11. Appearances can be deceiving. For example, when seeing a candle shaped like an apple, some people first assume that the object is an apple. Given the chance to examine it more closely, people typically change their mind and decide that the object is actually candle. If my child was in this situation, my child would understand that it was not the object that changed, but rather his or her ideas about the object that changed.	appearance-reality distinction
12. If I showed my child a cereal box filled with cookies and asked "What would someone who has not looked inside think is in the box?", my child would say that another person would think that there was cereal in the box.	false beliefs in context of unexpected contents
15. My child understands that when someone makes a 'guess' it means they are less certain than when they 'know' something.	certainty
16. My child understands that when someone is thinking about a cookie, they cannot actually smell, eat or share that cookie.	mental-physical distinction
26. My child can pretend that one object is a different object (for example, pretending a banana is a telephone).	pretense
29. My child understands the word 'if' when it is used hypothetically as in, "*If I had the money, I'd buy a new house*."	counterfactual reasoning
30. My child understands that when a person uses his/her hands as a bird, that the person doesn't actually think it is a real bird.	mental-physical distinction
31. My child knows how to make up stories to get what he/she wants.	ability to deceive
32. My child understands that in a game of hide and seek, you don't want the person who is 'it' to see you.	visual perspective-taking; play pragmatics

33. My child understands that when a person promises something, it means the person is supposed to do it.	speech acts: performatives
35. My child understands that when someone shares a secret, you are not supposed to tell anyone.	speech acts: basic
39. My child understands the word ‘believe’.	mental state term comprehension
42. My child understands that if Bruce is a mean boy and John is a nice boy, Bruce is more likely than John to engage in malicious or hurtful behaviors.	attribute-based behavior
Factor 3: EARLY THEORY OF MIND	
3. My child recognizes when someone needs help.	affect recognition: complex
6. My child understands that when people frown, they feel differently than when they smile.	affect recognition: expression-emotion relationship
24. My child understands whether someone hurts another on purpose or by accident.	intentionality
25. My child recognizes when others are happy.	affect recognition: basic
28. My child understands that, when I show fear, the situation is unsafe or dangerous.	social referencing
37. My child is able to show me things.	sharing attention: initiating
38. My child is able to pay attention when I show him/her something.	sharing attention: responding

As described below, the factor structure (drawn from a sample of individuals with ASD; for rationale see Hutchins et al., 2012) revealed interrelated ToM competencies reflecting a three *general* levels characteristic of the typical developmental progression. We do not suggest, however, that the developmental progression of ToM in typical children and individuals with ASD is equivalent. In fact, evidence to the contrary is accumulating (e.g., Peterson, Wellman, & Liu, 2005). We do suggest that the items comprising each factor hang together because their real world expressions share some underlying cognitive features or capacities.

Taken in reverse order, the content of factor 3 is most characteristic of the earliest ToM competencies that are known to emerge in typical development during infancy and toddlerhood. For example, the ability to engage in joint or shared attention (intended to be tapped by item 37 which taps initiation of shared attention and item 38 which taps the ability to respond to bids for shared attention) is estimated to emerge around 9 months of age (Tomasello, 1995) and has been described as a foundational skill for the advancement of social cognition (Bates, 1979; Carpendale & Lewis, 2006). Another ToM development associated with 9 months is social referencing (item 28) or the tendency of infants to look to a parent when faced with ambiguous events (Walden & Ogan, 1988). Social referencing may involve the ability to interpret others' facial expressions; an ability that we attempted to tap using at least three other items that also loaded on this factor (items 3, 6, 25). Although all of these items appear to tap affect recognition in some way, the items also differ: one appears to tap complex affect recognition (item 3) where a complex emotion is inferred from context, one appears to tap the emotion-expression relationship of basic emotions (item 6) and one appears to tap the recognition of a simple emotion (happy; item 25). Late toddlerhood has also been credited as a time when children can demonstrate some understanding of others' intention including whether acts are purposeful or accidental (item 24; Astington, 1991). In summary, the items that comprised factor 3 appear to represent some of the most significant ToM achievements that have been implicated in infancy and toddlerhood. This factor was termed “Early ToM: Reading Affect and Sharing Attention” and comprises the first subscale of the ToMI.

By contrast, factor 2 items tap ToM advancements characteristic of typically developing preschool children. Many of these are believed to require a basic metarepresentational skill with which a child makes use of mental representations and knows these are, in fact, representations. By approximately 4 years, typically developing children demonstrate an understanding of metarepresentation. This ability is associated with pretence (item 26; Leslie, 1987) and is seen as success on tasks involving false belief (items 8 and 12), the appearance-reality distinction (item 11), the mental-physical distinction (items 16 and 30), and seeing-leads-to-knowing (items 9 and 32), to name a few (e.g., Baron-Cohen, 1989; Gopnik & Astington, 1988; Wimmer & Perner, 1983). Metarepresentation is further implicated in the understanding of how psychological states guide behavior (items 1, 4; e.g., Leslie, & Frith, 1988) as well as attempts to engage in deception (item 31; e.g., Perner, 1991) although these attempts may vary considerably in their sophistication. Items intended to tap the understanding (*not* the earlier production) of cognitive terms (items 7, 10, 39) also loaded on factor 2 which has been reported to emerge around age 4 (Abbeduto & Rosenberg, 1984; Kazak, Collis, & Lewis, 1997; Moore & Furrow, 1991) and is reinforced by the finding that “proper comprehension of the lexical items ‘know’ and ‘think’ tends to go together with correct prediction of behavior” (Leslie & Frith, 1988, p. 322). The understanding of speech acts (items 33 and 35) and counterfactual reasoning (item 29) also loaded on factor 2. For each of these, the meta-level of understanding has been seen as crucial although a mature understanding of these aspects tend to emerge later than 4 years (Astington, 1988; Perner, Sprung, & Steinkogler, 2004). This factor was termed “Basic ToM: Metarepresentation and Developmentally Related Understandings” and comprises subscale two of the ToMI.

Metarepresentation is commonly viewed as a prerequisite skill that is necessary, but not sufficient, for an implicitly held theory of mind (Astington, 2003; Frith & Frith, 2000). Factor 1 items appeared to tap the most Advanced ToM knowledge, most of which are described in the literature as emerging in typical development between the ages of 6 and 8 years. These items require complex recursion, metapragmatic and metalinguistic skills, and an understanding of the mind as an active interpreter.

Recursive thinking requires the embedding of representations (e.g., Tiffany thinks about what Patty thinks) “and so first order false belief understanding is one example. However, there are more complex types of recursion than false belief understanding and these form important aspects of human thought about social matters” (Carpendale & Lewis, 2006, p. 190). These more complex forms of recursion include second-order beliefs (items 5, 22 and 23) as well the distinction between lies and jokes as this requires both the understanding that a falsehood is intended and the understanding of whether the falsehood is intended to be believed (item 20; Leekam & Prior, 1994; Sullivan, Winner, & Hopfield, 1995). Sophisticated metapragmatic competencies include the socially appropriate use of display rules (item 17; Flavell, Miller, & Miller, 2002) so that, for example, someone might smile even though she is unhappy. Advanced metalinguistic understanding involves not only the ability to identify the listener’s belief and the speaker’s intention but to distinguish various types of speech acts from each other (Keenan, 2003). This understanding includes, but is not limited to, knowledge of nonliteral idiomatic uses (item 13; Keenan, 2003), sarcasm (item 2; Keenan, 2003), and humor (e.g., play on words, item 36; McGhee, 1979).

Other items loading on factor 1 were designed to tap biased cognition (item 40), which is the understanding that one's previous experiences or expectations about others may color their interpretation of events (Pillow, 1991). Another item was intended to tap knowledge of the mind as an active interpreter (item 41; Carpendale & Chandler, 1996). Indeed, more Advanced ToM competencies "include a commonsense understanding that knowledge is interpretive and that the mind itself influences how the world is experienced" (Carpendale & Lewis, 2006, p. 193).

A mature ToM also involves the ability to make accurate social judgments (items 18 and 27), which is a particularly advanced skill. This is especially difficult for individuals with ASD considering all social encounters are embedded in context. Among other things, social judgment involves reading mental states and attitudes that may be revealed in subtle social cues and understanding their relation to the physical and social environment to extract meaningful and relevant information. In sum, the aforementioned advanced aspects of social cognition not only underscore the complexity and multifaceted nature of ToM, they also remind us that there is more to a mature understanding of ToM than mastery of false beliefs that emerge later in the typical course of development. This factor was termed "Advanced ToM: Complex Recursion, Mind as Active Interpreter, and Social Judgment" and comprises the third subscale the ToMI.

Overview and Uses

The ToMI was designed as a measure of broad ToM functioning as it includes real-world samples of behavior that caregivers can reliably and accurately identify. Because the ToMI is a caregiver-informant measure, it does not suffer from test-practice effects and so it is useful as a pre- and post-test assessment. It may be used in the context of intervention studies designed to assess the effects of a social skills curriculum or as a tool for the development of individualized instruction and progress monitoring tool in educational programs that are examining children's response to instruction. As described above, a few studies have demonstrated the sensitivity of the ToMI for this purpose; however, it must be cautioned that interventions spanning relatively short periods of time with limited intensity might not be expected to evidence change in theory of mind as measured by the ToMI. This is because ToMI scores rely on the knowledge and impressions of caregivers who accumulate insights into the child's ToM over time. Recent data (Houssa, Nader-Grosbois, & Jacobs, 2014; Hutchins & Prelock, 2008; 2013a, 2013b; Vivian, Hutchins & Prelock, 2012) suggest that fairly intensive ToM interventions of moderate time intervals (e.g., half hour sessions three times a week for 4-8 weeks) is sufficient to detect change in ToM development in clinical and experimental contexts. The sensitivity of the ToMI to detect the effects of interventions that adopt less intensive or shorter durations is unclear, thus caution is warranted in such cases. Of course, the ToMI can also be used as a matching criterion in research (e.g., Grossman, Peskin, & San Juan, in press).

Because deficits in ToM represent a universal characteristic of ASD, an important use of the ToMI is to aid in the identification of ASD and data for the ToMI's sensitivity and specificity are offered in the Reliability and Validity sections of this manual. The ToMI is particularly useful in identification of ASD among those individuals who present with the most advanced cognitive and language skills who may not be identified by existing measures. Indeed, it has long been recognized that individuals with Asperger Syndrome and high functioning autism often perform at ceiling levels on tests – even advanced tests- of direct ToM performance even though demonstrable ToM deficits are revealed in day-to-day functioning. This had lead several researchers to conclude that high functioning individuals with ASD may develop compensatory strategies to 'hack' through explicitly stated ToM problems using a nonmentalistic strategy (e.g., Happé, 1995) thus posing a serious challenge to the validity of measures of direct performance when administered to those populations.

It is important to note that the ToMI may be used as part of a larger assessment battery in the identification of psychiatric and developmental disorders, other than ASD, that can be associated with impairments in theory of mind. Although the literature is sometimes mixed, support for the presence of ToM impairment has been offered for conditions such as attention deficit hyperactivity disorder (Buhler, Bachmann, Goyert, Heinzel-Gutenbrunner, & Kamp-Becker, 2011; Buitelaar, Van der wees, Swaab-Barnveveld, & Jan Van der gaag, 1999; Perner, Kain, & Barchfeld, 2002; Uekermann, et al., 2010), fetal alcohol syndrome (GreenBaum, Stevens, Nash, Korean, & Rovet, 2009), fragile X syndrome (Cornish, Burack, Rahman, Russo, & Grant, 2005; Grant, Apperly, & Oliver, 2007), learning disability (Ashcroft, Jervis, & Roberts, 1999), intellectual disability (Abbeduto, Short-Meyerson, Benson, & Dolish, 2004), and schizophrenia (Hans, Auerbach, Styr, & Marcus, 2004). To the degree that there are broad commonalities in ToM impairments across clinical conditions, the ToMI is *not* appropriate as a tool for differential diagnosis. On the other hand, some research suggests the presence of syndrome-specific ToM deficiencies and proficiencies (e.g., Buhler et al., 2011; Cornish et al., 2005; Perner et al., 2002). Because the ToMI taps a wide range of ToM understandings and skill sets, there is a potential for the use of the ToMI for differential diagnosis in the future as more information accrues to inform the questions as to whether there are unique ToM profiles associated with specific clinical populations.

Finally, traditional ToM tasks (e.g., the Sally-Anne task) are difficult and cumbersome to adapt for use with some populations. Specifically, the ToMI may be a particularly valuable tool for assessing the ToM competencies of blind or deaf children of typically developing caregivers.

In fact, a few investigations of ToM in blind or deaf children that make use of the ToMI have been completed or are currently underway (e.g., Kossewska, in preparation).

The ToMI is designed to tap a wide range of ToM competencies; thus, it may be particularly helpful for identifying areas of strength and weakness in an individual's social-cognitive profile and developmentally appropriate targets for treatment. Hutchins and Prelock (2008) described the ways in which the earlier (and similar) version of this measure was used as part of a larger assessment battery for precisely these purposes. The chapter on how to use the ToMI in clinical decision-making elaborates on this discussion by offering additional examples of the use of ToMI scores for clinical purposes to link assessment of ToM to developmentally appropriate treatment goals and strategies for remediation.

We have previously argued (Hutchins, Bonazinga, et al., 2008, Hutchins et al., 2012) that the use of caregivers as informants, who are uniquely situated to observe their child's ToM during real world social interaction, helps move us toward an assessment of ToM that is socially valid and family-centered. Measures with social significance are important because they help to ensure that assessment is relevant and meaningful in everyday life. Moreover, given that primary caregivers are expert authorities on their children, caregivers can be recruited as valuable partners in assessment, treatment planning, and evaluation of interventions. When ToM competencies are relevant, the ToMI can act as a tool for beginning a conversation with families about the children's strengths and challenges, treatment strategies and supports that are likely to enhance outcomes, and the families' priorities for intervention.

TEST ADMINISTRATION & SCORING

Administration

Primary caregivers may be biological or adoptive parents, grandparents, or others who assume the primary responsibility for care. Previous research associated with the development of the ToMI (Hutchins, Bonazinga, et al., 2008) revealed that caregivers who spend less than five hours per day *on average* with the child (not counting the time the child is sleeping) predicted children's performance on ToM tasks less accurately than did parents who spent a minimum of five hours per day with their child. Therefore, we recommend that the respondents be primary caregivers who spend a minimum of five hours per day with the individual on which they are reporting.

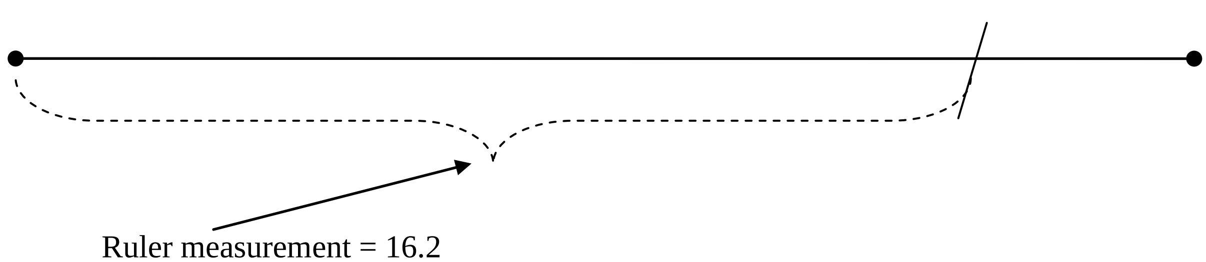
The ToMI was developed using a sample that included native English-speaking and bilingual caregivers who were fluent in English. The ToMI should not be administered to caregivers who are not fluent English speakers; however, some carefully vetted translations (e.g., French, Italian, German, Spanish, and others) of the ToMI have been validated and are available at www.theoryofmindinventory.com. Finally, the ToMI should not be administered to caregivers who report or are suspected of having any debilitating and unresolved psychiatric disorder. In such instances, assessment of ToM using direct measures of child performance is preferable.

The ToMI should be completed in a quiet, well-lit, and comfortable environment. Respondents may be informed that this test is used to assess caregivers' ideas about their children's thoughts and reasoning. Respondents should then be asked to read the instructions in their entirety and to pay close attention to the examples of correct and incorrect response strategies. Our experience suggests that respondents sometimes circle an anchor or place an 'X' at some point along the continuum; however, this will introduce error and complicate scoring and interpretation. For this reason, it is important for respondents to carefully read the instructions and to understand the importance of making a single hash mark that intersects the continuum at the point that best reflects their attitude. Test administrators should avoid answering questions about the content of the test. In our experience, questions rarely arise but when they do, variability in administrators' responses has the potential to introduce error. For this reason, we recommend that respondents be directed to "interpret the question the best you can and try to give your most general answer using everything you know about this child". The ToMI takes approximately 10 minutes to complete.

Scoring by Ruler

Each of the 42 items comprising the ToMI is scored using the ruler that accompanies the Examiner's Manual. The ruler spans 20 metric units and yields possible scores for each item that range from 0 – 20 with higher values reflecting greater degrees of confidence that a target child possesses a particular ToM understanding. An example of correct responding (i.e., drawing a vertical hashmark that intersects the continuum) is given below. Whether the hashmark is slanted or perpendicular to the continuum is irrelevant and in both cases, the item is scored by measuring the precise point of intersection as shown below.

Definitely Not	Probably Not	Undecided	Probably	Definitely



To enhance precision, rounding to full integers is discouraged and each item should be scored to one decimal place (e.g., a score of 16.2 is preferable to a rounded score of 16). In fact, frequent users of the ToMI will soon discover that small changes in subscale and composite scores often translate into large differences in percentile ranks thus underscoring the importance of precision in measurement. Hand scoring using the ruler takes approximately five minutes.

Calculating Composite and Subscale Scores by Computer

After scoring by ruler, ToMI users have the option to enter values into a computer program, which will automatically derive composite and subscale scores and their associated percentile ranks. To access this feature, visit www.theoryofmindinventory.com and link to the "For Professionals" site. Here you can enter relevant demographic information and individual ToMI item scores. All demographic information except age (in years) is optional, is not recorded to the server in any way, and the information provided can only be saved by you as downloads or printouts from your computer. This feature is available to aid professionals who wish to generate a personalized report—another feature available on this site.

For computer scoring, enter the value obtained from ruler scoring in the box on the right hand side of each item. When all values are entered, click 'Submit' to view the report. As stated above, the report will provide composite and subscale reports along with their associated percentile rank. The report will also provide raw scores for individual items within each subscale. This provides a quick way to scrutinize the data and may be helpful in determining whether particular ToM dimensions may be appropriate targets of treatment. Strategies for this use are discussed more fully in the Clinical Decision-Making section of this manual.

Please note that no normative tables are available in this manual and standard scores based on preliminary norms can only be obtained using the computer scoring procedure described above. This is because the present version of the ToMI is offered as a research version of the tool with preliminary norms and with data still in collection. Norms for the ToMI are updated periodically and linked to the report generator to give users the most recent norms based on the largest normative sample to date. Data for sample size by age and gender will be made available in a subsequent release of the tool.

STANDARDIZATION SAMPLE

Typically Developing Sample. A small national sample of 37 ToMI forms were completed by primary caregivers (all mothers) of children identified as typically developing. This sample was from only five states (i.e., Vermont, Texas, Illinois, California, and Massachusetts). A local (Vermont only) sample of 60 mothers of 87 typically developing children also participated in this study. Like the ASD sample described above, this sample over-represented the New England region of the United States.

For the combined samples (n = 124), caregiver age ranged from 23 to 52 (*M* = 38; *SD* = 5), education ranged from 12 (completion of high school) to 20 (doctoral degree) years (*M* = 16.67; S*D* = 1.98), and gross annual combined income ranged from less than $20,000–$275,000 (*M* = $102,355; *SD* = $44,210). All caregivers identified as the child's primary caregiver and reported spending an average of 7.4 hours with the child per day (*SD* = 2.46; not counting when the child was sleeping). A total of 120 (97%) caregivers were native English speakers, all were fluent in English, and 16 (12.9%) were fluent in more than one language.

Children were 62 females (50%) and 62 males (50%) who ranged in age from 2 years, 0 months to 12 years, 8 months (*M* = 7, *SD* = 2.22). Typically developing children were identified on the basis of parent report and parents' responses to a questionnaire that was designed to screen for a variety of clinical conditions (e.g., uncorrected visual or hearing impairment, language or psychiatric disorders). Only parents who reported the absence of any condition and the absence of any parental concern for any condition were included in analyses. On the basis of caregiver report, no child had ever been diagnosed with a disability with the exception that four had received services in the past to remediate a speech (not language) impairment which was not an inclusion criterion.

ASD Sample. A small national sample of 104 ToMI forms were completed by primary caregivers (99 mothers and 5 fathers) of children diagnosed with ASD. The national sample was drawn from 14 states and each major geographic region in the United States (i.e., Northeast, Southeast, Northwest, Southwest, Midwest). A local (Vermont) sample of 31 mothers and their children diagnosed with ASD also added to this sample, thus, the New England region of the Unites States if over-represented in the present sample.

For the combined sample (n = 135), caregiver age ranged from 24 to 59 (M = 42; SD = 6.8), education ranged from 12 (completion of high school) to 20 (doctoral degree) years (M = 15.40; SD = 2.21), and gross annual combined income ranged from less than $1,000 – $400,000 (M = $67,350; SD = $48,283). All caregivers identified as the child's primary caregiver and reported spending an average of 7.8 hours with the child per day (*SD* = 3.38; not counting when the child was sleeping). A total of 127 (94.1%) caregivers were native English speakers, all were fluent in English, and 19 (14.1%) were fluent in more than one language. Children were 27 females (20%) and 108 males (80%) who ranged in age from 3 years, 4 months to 17 years, 8 months (M = 10.12, SD = 4.19). On the basis of caregiver report, 74 (55%) children were currently diagnosed with autistic disorder, 30 (22.2%) were diagnosed with Pervasive Developmental Disability—Not Otherwise Specified (PDD-NOS), 29 (21.5%) were diagnosed with Asperger Syndrome, and 2 reported ''other'' (and indicated ''autism spectrum disorder''). Also according to parent report, 20 (14.8%) children were characterized as functionally nonverbal, 24 (17.7%) were characterized as having limited language (i.e., uses 2 – 3 word utterances), and 91 (67.5%) were characterized as verbal (uses a variety of words to communicate flexibly and for a range of purposes).

Descriptive Statistics for ToMI

Typically developing sample. For the combined typically developing sample, ToMI scores ranged from 4 to 20 (*M* = 15.62; *SD* = 3.34). Descriptive data for the overall ToMI scores were examined by age. These data are presented in Figure 1. Inspection of the descriptive data by subscale scores is informative. These data are presented in Figure 2. A line corresponding to a score of 15 has been added to the graph to indicate the point at which parents, on average, endorsed ToM competencies as "probably" present.

Figure 1: Descriptive data for composite ToMI scores by age for the typically developing sample.

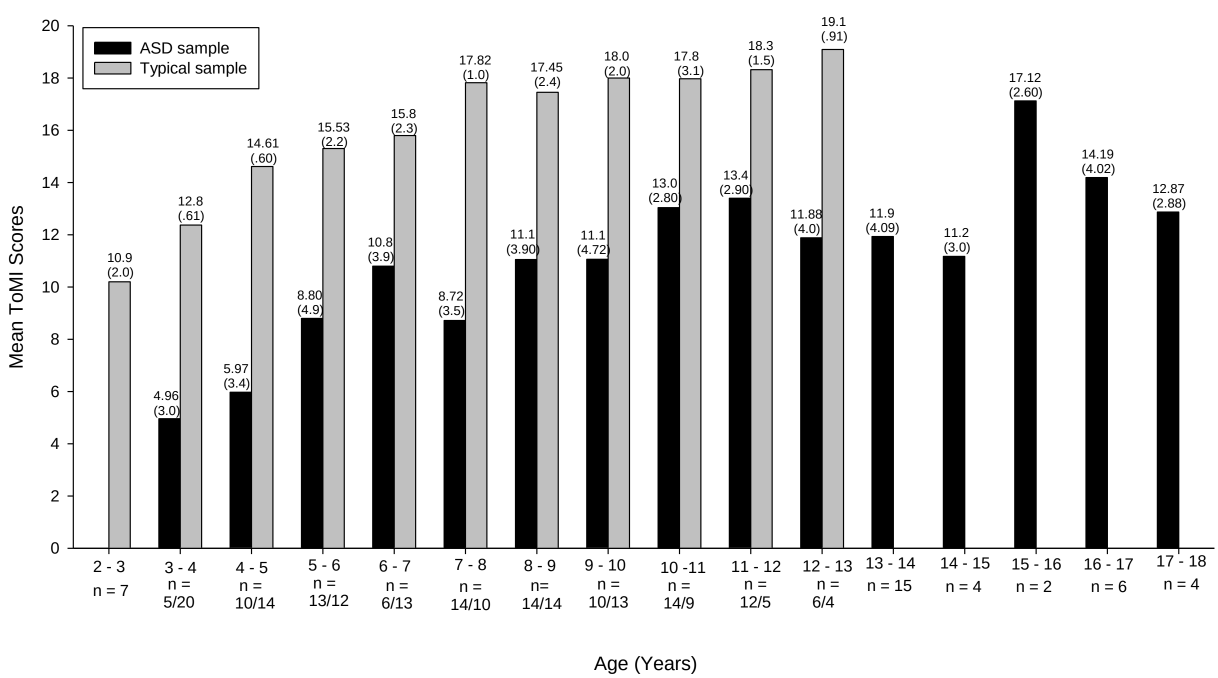
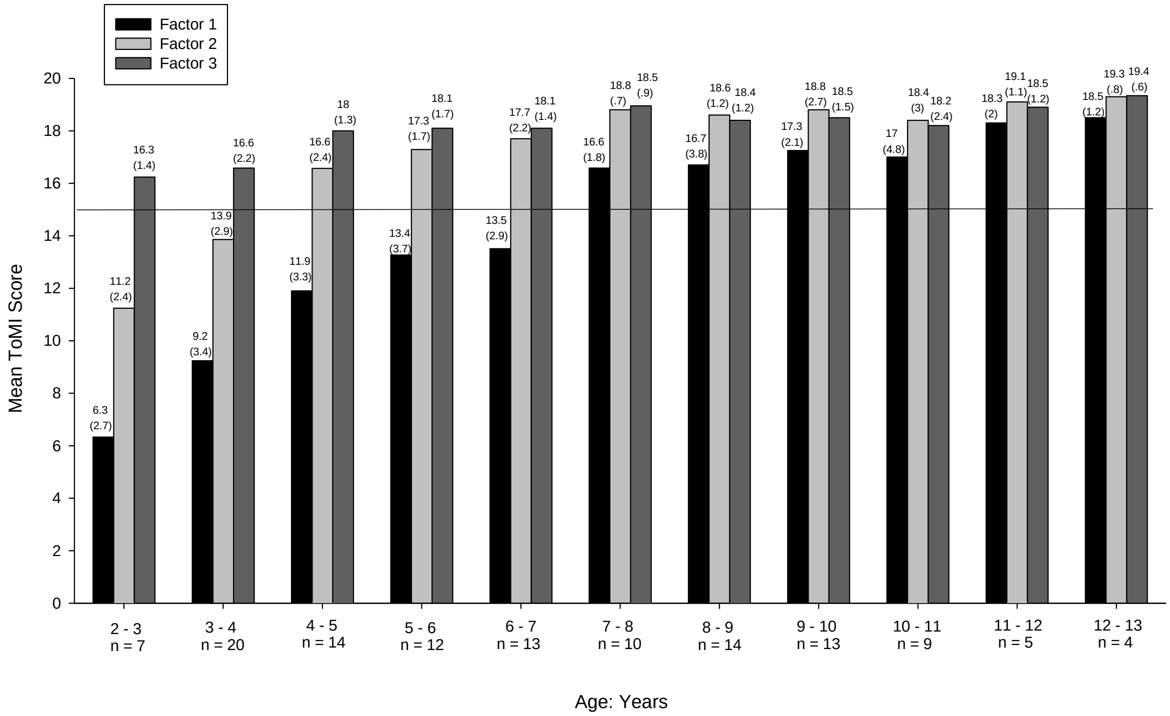
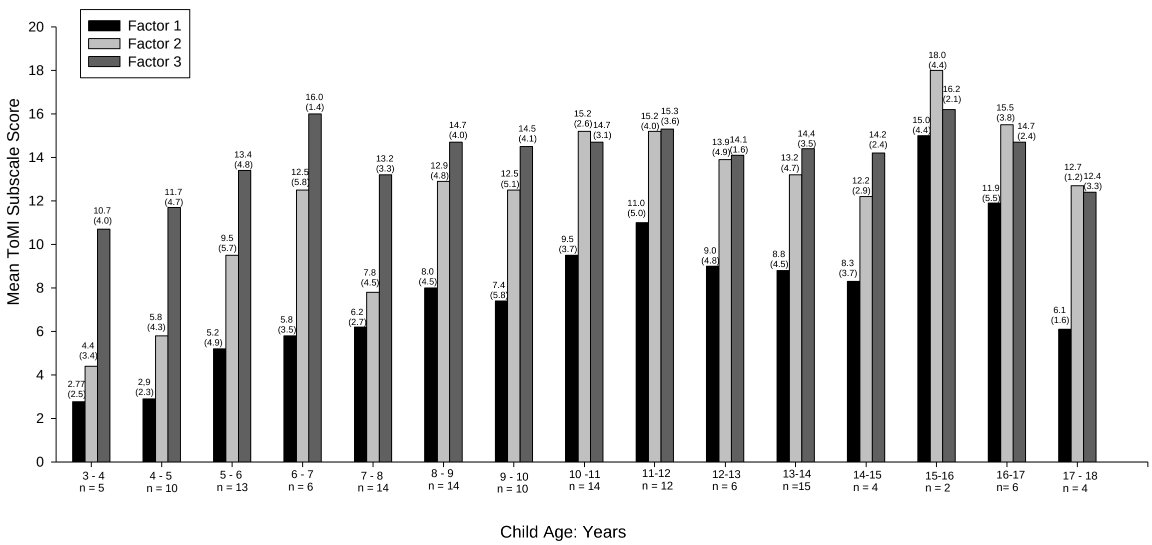


Figure 2: Descriptive data for subscale ToMI scores for the typically developing sample.



ASD sample. For the combined ASD sample, ToMI scores ranged from 1.1 to 19.95 (*M* = 10.8; *SD* = 4.5). Descriptive data by age (year) are provided in Figure 1 for the entire sample. Because some comparisons were conducted on the basis of the child's verbal abilities, descriptive data for this variable are also offered. The average score for the combined sample for children characterized as nonverbal on the basis of parent report was 6.54 (*SD* = .71), which was similar to the average score for children with limited language (*M* = 6.52; *SD*= .70). The average score for children characterized as verbal was 12.68 (*SD* = .35). With regard to receptive language (local sample only; as measured by the PPVT-4), 11 children scored three standard deviations below the mean, five scored two standard deviations below the mean, and two scored one standard deviation below the mean. An additional 10 children scored in the normal range, two scored one standard deviation above the mean, and one scored three standard deviations above the mean. These data indicate that children with ASD with a wide range of verbal abilities were represented in this study. Descriptive data for ToMI subscales for the ASD sample by age are presented in Figure 3.

Figure 3: Descriptive data for ToMI subscales for the ASD sample by age



Normative Data: Raw Scores and Preliminary Norms

At present, the ToMI yields raw scores and percentile ranks. Raw scores are typically regarded as useful insofar as they can be converted to derived scores (McCauley, 2001); however, raw scores for subscales that reflect a developmental progression (as the early, basic, and advanced subscales of the ToMI do) may be useful for developing qualitative impressions of intra-cognitive differences related to ToM for any given individual. That is, examination of the raw data can be useful when the user is interested in analyses at the individual item level and/or subscale level. Suggestions for how to use raw scores to develop impressions about the relative strength and challenge areas in theory of mind development and the identification of targeted treatment goals are addressed more fully in subsequent sections.

Percentile ranks are a kind of ordinal-level standard score. Percentile ranks reflect the percent of individuals in the normative sample whose scores fall at or below a given score. Percentile ranks are a popular derived score. Not only are they relatively easy to interpret (McCauley, 2001) but they may be preferable when distributions are skewed (Fenson, Marchman, Thal, Dale, Reznick, & Bates, 2007). As can be seen in Figures 1 and 2, ToMI scores for the typically developing sample are negatively skewed with ceiling effects evident for most subscale and composite scores occurring around age eight years and ToMI scores for the ASD are only slightly negatively skewed with no ceiling evident. Given skewness in the data, we opted against the use of other types of standard scores believing that percentiles best capture the theory of mind abilities of children relative to other children of the same age. At present, percentiles are presently only available for ToMI subscale and composite scores (i.e., they are not available for individual items), but as noted above, inspection of raw scores by item can be helpful in interpretation of the ToMI.

It is critically important to emphasize that the percentile ranks that are currently available are based on a very small sample. This means that the percentiles may be unstable. Research is currently underway to address this limitation and the percentiles that are offered in the computer scoring and generation of reports must be regarded as preliminary and tentative estimates. At present, percentiles are also calculated solely on the basis of age (not gender). This represents another important goal of future versions of the ToMI as slight theory of mind advantages have been observed for females (Charman, Ruffman, & Clements, 2002).

Clearly, caution must be exercised in the interpretation of ToMI percentiles and there are a few points in particular that warrant special attention. First, given that no ceiling effects were observed in the ASD sample (even among the oldest and highly verbal individuals), we suspect that the ToMI can be appropriately used for older adolescents and adults. In this situation, scores for older individuals who are at risk for poor theory of mind development, would be compared to the oldest sample of typically developing sample—12 years of age. This is possible because the typically developing sample did, in fact, evidence ceiling effects. Of course, interpretation of scores for this use must be approached with caution. First, literal interpretation is only appropriate in situations where percentiles indicate scores in the low average or below average range. That is, if an 18 year old with ASD achieves ToMI scores around the 50^{th} percentile when compared to typically developing 12-year-olds, it would *not* be appropriate to conclude 1) that the 18-year-old has theory of mind functioning that is equivalent to a 12-year-old, or 2) the 18 year old is doing fine with regard to theory of mind development. On the other hand, scores are more meaningful when the 18-year -old obtains a low percentile compared to the 12-year-olds in the normative group, in which case one would conclude deficits in ToM. For persons older than 12 years who have scores in the normal range, literal interpretation of scores is complicated by the fact that, despite ceiling effects in the typically developing sample, ToM development is expected to continue in later adolescence and adulthood and so this can lead to overestimation of ToM knowledge and skills.

Second, we have adopted a criterion to identify a clinical range cut-off based on our data for sensitivity and specificity (described below). This cut-off is the 10^{th} percentile for the composite ToMI score and it is consonant with many other parent-informant measures (e.g., Fenson et al., 2007). Scores that fall at or below the 10^{th} percentile are denoted by an asterisk on the computer-generated report (theoryofmindinventory.com). These scores are at the subscale and composite level (and not available for individual items); however it should be noted that it is the composite score data that were used in estimates of sensitivity and specificity and they are considered the most robust indicator as opposed to the use of a single subscale or combination of subscales.

In the development of percentiles, it is important to note that the nature of the percentile ranks is ultimately determined by the nature of the underlying distribution of scores. For ages at which skills in question are beginning to emerge, scores tend to be compressed. This means that "very small shifts in raw scores can sometimes produce large shifts in percentile scores" (Fenson et al., 2007, p. 33). As such, a 2-point difference for very young children can result in a 40-point change in percentile ranks while a 2-point different for an older child can result in a change of only 1 or 2 percentiles. In a related vein, it is also true (and largely explained by a small sample size) that raw data often corresponded to a range of percentile rankings (e.g., a score of 9.4 may correspond to a range between the 1st and the 7th percentiles). When these situations occur, a conservative approach may be to use the larger value and assume higher ToM competence. Another approach is to interpolate a single value by simply determining the midpoint of the range (e.g., 4th percentile as it is half way between the 1st and the 7th percentile). For the sake of accuracy, we suggest that the range and the midpoint of percentile ranks be reported when they occur; however, there may be clinical or research purposes that preclude this approach and this must be left to ToMI users to determine.

The ToMI as a Criterion-Referenced Measure

Of course, norm-referenced measures may be expressed and employed as criterion-referenced measures. Conversely, all criterion-referenced measures are ultimately rooted in some index of normative performance on the test (McCauley, 2001). That is, cut-scores (often initially determined arbitrarily) are typically adjusted to maximize their ability to make correct decisions. For example, a cut-score of 90% on a driver's license exam is likely to be adjusted downward when it results in too few good drivers passing the test. Similarly, a cut-score of 50% will be adjusted upwards when it results in oodles of reckless drivers lawfully obtaining their driving privileges. Unlike many traditional criterion-referenced measures that adopt an adjustment approach like that just described, ToMI cut-scores were developed using our preliminary normative data. Subscale and composite cut-scores were identified by age and cut-scores represent extreme and low scores. When any one of the four obtained scores (early, basic, advanced, composite score) falls below the cut-score, this should be taken as an indication that the individual is at risk for poor ToM outcomes. The cut-scores by age are presented in Table 2.

Table 2: Cut-scores indicating risk for poor ToM outcomes (early, basic, advanced, composite scores) by age.

Age (years)	Early subscale	Basic subscale	Advanced subscale	Composite score
2-3	13	7	--	7
3-4	13	8	--	10
4-5	15	10	6	12
5-6	15	11	7	13
6-7	15	11	10	14
7-10	15	13	11	14
10-11	15	13	12	14
11-12	15	14	13	14
12-13	16	14	13	15

RELIABILITY AND VALIDITY

Reliability refers to the characteristics of dependability of measurement and is seen as a necessary, but not sufficient criterion for validity (McCauley, 2001). Reliability of the ToMI was examined by first looking at the measure's test-retest reliability—an index of temporal stability. This was followed by an examination of the ToMI's internal consistency—a measure of homogeneity of content. When internal consistency is high, this is generally taken as evidence that the items on a measure tap a unitary construct (in this case, ToM).

Typically, the first step in establishing validity is logical and theoretical as opposed to statistical (McCauley, 2001). Content validity refers to the degree to which items on a measure adequately tap the construct of interest while avoiding irrelevant content (McCauley, 2001). The content validity of the ToMI was considered during the item development phase and in the item revision of the original version (described earlier in this manual) and a primary goal was to develop a content valid measure of ToM using a panel of experts that included one expert in ASD and ToM, one expert in ToM and test development, and a student well-versed in all of these topics.

Formal validity of the ToMI was explored in several different analyses for typically developing and ASD samples. As described below, the psychometric properties of the ToMI were examined in several tests of construct validity. These include 1) tests of criterion-related validity (where test scores are judged against a criterion to evaluate the test's performance), 2) factor analysis (a statistical technique to explore the dimensionality of a measure), and 3) a contrasting groups method of construct validation (where groups who should differ on the construct are compared). In addition, we also include data for sensitivity, specificity, and overall accuracy of classification.

Reliability: Typically Developing Sample

Test-retest reliability. Data for 29 parents from the local sample who completed the ToMI at time two provided estimates of test–retest reliability. A Pearson's product moment correlation indicated strong stability of this measure using an interval of 14 – 78 days ($M = 27.5$; $SD = 19.4$; $r = .89$, $p < .001$), which is a highly dependable relationship with variation in scores at time 1 accounting for approximately 80% of the variation in scores at time 2. The Standard

Error of Measurement (SEM; an index of the average amount of error in a test) was 1.50, which is acceptable (i.e., it is less than one-third of the standard deviation; McCauley, 2001).

Internal consistency. Internal consistency was assessed using Cronbach's alpha, which resulted in a very high estimate of content consistency ($\alpha = .98$). Examination of the alpha based on item deletion indicated that removal of two items resulted in slight improvement in Cronbach's alpha. These two items ("My child understands the word desire" and "My child understands the word need" which appeared to assess knowledge of low frequency mental state terms) were subsequently dropped from further analysis.

Validity: Typically Developing Sample

Criterion-related validity. As expected, a Pearson's product moment correlation indicated a substantial positive relationship between ToMI scores and PPVT-4 standard scores ($r = .73$, $p < .05$) with variation in ToMI scores accounting for approximately 53% of the variation in children's receptive vocabulary.

A construct valid measure of ToM competence should also be positively correlated with children's scores on ToM tasks. To examine this, we compared ToMI scores to scores obtained on the ToM Task Battery (described in the final section of this manual). A Spearman's rho (because the ToM Task Battery data are best construed as ordinal in nature) indicated a substantial positive relationship ($r = .66$, $p < .05$) with variation in scores on the ToMI explaining approximately 44 percent of the variation in children's scores on the ToM task battery.

Contrasting-groups developmental method of construct validity. The age of four years is often considered a time when significant ToM developments can be observed. 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. Data from the sample (i.e., national and local) sample of typically developing 2.5 to 3.5 years olds (*n* = 16) were compared to data from mothers of typically developing 4 to 5 year olds (*n* = 13). An independent t-test indicated that there was a significant difference between mothers of 2.5 to 3.5 year olds (*M* = 11.27; *SD* = 2.40) and mothers of 4 to 5 years olds (*M* = 14.79; *SD* = 2.08), $t\ (27) = 4.12$, $p < .01$.

Reliability: ASD Sample

Test-retest reliability. Twenty-eight mothers (46.6% response rate) of 37 children completed and returned the measure at time two to provide estimates of test–retest reliability. The test-retest interval was between 12 and 44 days (*M* = 25.46; *SD* = 10.43). Data for the 37 responses indicated high test–retest reliability ($r = .89$, p < .01). This is a highly dependent relationship with variation in scores at time one explaining approximately 80% of the variation in scores at time two. The SEM was 1.11, which is acceptable (again, it was less than one-third of the standard deviation).

Internal consistency. Internal consistency was assessed using Cronbach's alpha, which resulted in a very high estimate of content consistency ($\alpha = .98$). As with the TYD sample, examination of alpha based on item deletion indicated that removal of two items resulted in slight improvement in Cronbach's alpha. These two items (the same as indicated earlier) were dropped from further analysis.

Validity: ASD Sample

Criterion – related validity. ToM not only develops in early childhood but it continues to develop into late childhood and beyond (although these studies are relatively rare). Thus, it was expected that a construct valid measure of ToM would show increases in scores that correspond to increases in child age among children who are typically developing (ages 2 – 12). A significant correlation was found between child age and ToMI scores (r = .72, p < .05). Variation in child age accounted for approximately 52% of the variation in ToMI scores.

Criterion-related validity was further examined by correlating ToMI and ToM Task Battery scores. A subset of 24 children (between 2-9 years) from the local sample completed the ToM Task Battery. Child participants between the ages of 2 and 9 years were chosen to avoid ceiling effects in ToM task battery performance. A Spearman's rho revealed a significant relationship (r = .82, p < .05) with variation in ToMI scores accounting for 67% of the variance in scores on the ToM Task Battery.

Contrasting Groups Method of Construct Validation. We also expected a construct valid measure of ToM to distinguish typically developing children and individuals with ASD. An independent t-test revealed a significant difference, t (257) = 10.04, p < .001, such that mothers of children identified as having ASD (ages 3 – 17) reported lower scores (M = 10.8; SD = 4.42) than did mothers of younger (ages 2 – 12) typically developing children (M = 15.6; SD = 3.37).

Factor Analysis. Data for all ASD participants (national and local sample, n = 135) were submitted to exploratory principal components analysis (PCA). Three rotation methods were explored: one was orthogonal (varimax) and two were oblique (direct oblimin and promax). Although oblique rotation is recommended for correlated factors (which applied to some of our factors), all rotation methods resulted in nearly identical factor solutions and varimax rotation was most effective in achieving a simple structure. Of course, many measures remained complex in that they continued to load on more than a single factor (> .30).

Varimax rotation with Kaiser normalization yielded a six-factor solution. Upon inspection of the rotated matrix, factors 4 and 5 were comprised of two items and were dropped from further analysis as is convention (Costello & Osborne, 2005). Factors 1 – 3 accounted for 62.7% of the cumulative variance (i.e., factor 1 = 52% with Eigenvalue of 24.9; factor 2 = 6.4% with Eigenvalue of 3; factor 3 = 4.4% with Eigenvalue of 2.1). The rotated component matrix for the three retained factors are presented in Table 3

Table 3: Factor Structure 3/13/15]

Factor 1 (*n* = 16)	Factor 1	Factor 2	Factor 3
2. If it were raining and I said in a sarcastic voice "Gee, looks like a really nice day outside," my child would understand that I didn't actually think it was a nice day.	**.39**		
5. My child understands that people can be wrong about what other people want.	**.59**		
13. If I said "Let's hit the road!" my child would understand that I really meant "Let's go!"	**.53**	.35	
14. My child understands that people can lie to purposely mislead others.	**.63**	.51	
17. My child understands that people can smile even when they are not happy.	**.74**		
18. My child understands the difference between when a friend is teasing in a nice way and when a bully is making fun of someone in a mean way.	**.73**		
19. My child understands that people don't always say what they are thinking because they don't want to hurt others' feelings.	**.84**		
20. My child understands the difference between lies and jokes.	**.69**		
21. 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.	**.74**		
22. My child understands that people often have thoughts about other peoples' *thoughts*.	**.75**	.35	
23. My child understands that people often have thoughts about other peoples' *feelings*	**.66**	.42	
27. My child recognizes when a listener is not interested.	**.56**		.51
34. My child is able to put himself/herself in other people's shoes and understand how they feel.	**.68**		
36. If I said "What is black, white and '*read*' all over? It's a newspaper!" my child would understand the humor in this play on words.	**.66**	.41	
40. When we like others, we are likely to interpret their behavior in positive ways and when we don't like others, we are likely to interpret their behavior more negatively. My child understands that previous ideas and/or opinions of others can influence how we interpret their behaviors.	**.60**		
41. My child understands that two people can see the same image and interpret it differently. For example, when looking at this image, one person might see a rabbit whereas another might see a duck.	**.52**	.46	
Factor 2 (*n* = 19)			
1. My child understands that when someone puts on a jacket, it is probably because he/she is cold		**.56**	
4. My child understands that when someone says they are afraid of the dark, they will not want to go into a dark room.		**.65**	.35
7. My child understands the word 'think'	.44	**.62**	
8. 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.	.39	**.44**	

9. My child understands that to know what is in an unmarked box, you have to see or hear about what is in that box.		**.47**	
10. My child understands the word ‘know’.		**.71**	
11. Appearances can be deceiving. For example, when seeing a candle shaped like an apple, some people first assume that the object is an apple. Given the chance to examine it more closely, people typically change their mind and decide that the object is actually a candle. If my child was in this situation, my child would understand that it was not the object that changed, but rather his or her ideas about the object that changed.	.35	**.54**	
12. If I showed my child a cereal box filled with cookies and asked “What would someone who has not looked inside think is in the box?”, my child would say that another person would think that there was cereal in the box.		**.72**	
15. My child understands that when someone makes a ‘guess’ it means they are less certain than when they ‘know’ something.	.56	**.62**	
16. My child understands that when someone is thinking about a cookie, they cannot actually smell, eat or share that cookie.	.49	**.64**	
26. My child can pretend that one object is a different object (for example, pretending a banana is a telephone).		**.55**	.52
29. My child understands the word ‘if’ when it is used hypothetically as in, “*If I had the money, I’d buy a new house*.”	.41	**.56**	
30. My child understands that when a person uses his/her hands as a bird, that the person doesn’t actually think it is a real bird.		**.71**	
31. My child knows how to make up stories to get what he/she wants.	.43	**.56**	
32. My child understands that in a game of hide and seek, you don’t want the person who is ‘it’ to see you.		**.79**	
33. My child understands that when a person promises something, it means the person is supposed to do it.		**.71**	
35. My child understands that when someone shares a secret, you are not supposed to tell anyone.	.43	**.67**	
39. My child understands the word ‘believe’.	.53	**.65**	
42. My child understands that if Bruce is a mean boy and John is a nice boy, Bruce is more likely than John to engage in malicious or hurtful behaviors.		**.68**	
Factor 3 (*n* = 7)			
3. My child recognizes when someone needs help.			**.54**
6. My child understands that when people frown, they feel differently than when they smile.			**.56**
24. My child understands whether someone hurts another on purpose or by accident.	.41		**.58**
25. My child recognizes when others are happy.			**.62**
28. My child understands that, when I show fear, the situation is unsafe or dangerous.		.32	**.67**
37. My child is able to show me things.		.43	**.64**
38. My child is able to pay attention when I show him/her something.			**.72**

Accuracy of Classification

The data sets described above were used to calculate sensitivity, specificity, positive predictive validity, negative predictive validity, and overall accuracy. The data used in these analyses were drawn from the samples described above.

Sensitivity. Sensitivity refers to the correct detection rate or the ability of a test to give a positive result when the person being assessed truly does have the disorder (McCauley, 2001). Our data reveal that 88.3% of individuals diagnosed with ASD (who varied widely in verbal and cognitive skills) were captured by the ToMI when composite scores fell at or below the 10^{th} percentile when compared to a typically developing normative sample. This figure rises to 91.2% identification using a criterion of less than the 20^{th} percentile and 94.3% identification using less than the 30^{th} percentile. This finding was observed among the youngest children sampled (currently age three) although -as would be expected- it is the Early subscale score that accounts for the majority of the variance between groups in most cases. This was also observed for the oldest and highest functioning individuals in our sample lending credibility to the notion that the ToMI may be particularly useful in the identification of ASD for those with the most advanced language and cognitive skills.

Specificity. Specificity refers to correct rejection rate or the ability of a test to give a negative result when the person being assessed truly does not have the disorder (McCauley, 2001). Analyses for specificity revealed that 93.3% of typically developing individuals were correctly rejected by the ToMI when composite scores exceeded the 10^{th} percentile. It is important to keep in mind, however, that this analysis was performed to explore the ToMI's accuracy for classifying persons already identified as typically developing or having a diagnosis of ASD. As stated previously, the ToMI is not intended (at present) as a tool for differential diagnosis and persons who are at risk for poor ToM development may obtain scores that fall below the 10th percentile because they have conditions other than ASD.

Predictive validity. Positive predictive validity is the percent of positive tests where the individual actually has the disorder. Analyses revealed positive predictive validity at the level of 91.2 %. Negative predictive validity is the percent of negative tests where the individual does not have the disorder. Negative predictive validity was found at the level of 90.9%.

Overall accuracy. The overall accuracy of the ToMI for identification of ASD using the criterion of the 10th percentile for the composite score was 91.1% meaning that in 91% of cases, the ToMI made a correct decision. The Rows by Column contingency table for accuracy data are represented in Table 4.

Table 4: Rows by Column contingency table for accuracy of classification.

		Diagnosis of ASD	
		Y	N
ToMI composite score < 10th percentile	Y	83	8
	N	11	111

USING TOMI SCORES IN CLINICAL DECISION-MAKING

Clinicians across work settings are seeing greater numbers of children with ASD who exhibit a range of social communication challenges with specific deficits in ToM. Since varying abilities in ToM often characterize the child with ASD, speech-language pathologists (SLPs) need to understand the available interventions, consider the profiles of children with Early, Basic, and/or Advanced ToM in determining the most appropriate interventions, and recognize those social cognitive skills (e.g., emotion recognition, metarepresentation, empathy) most likely to be positively impacted by intervention (Prelock, 2011). This chapter highlights interventions that may be appropriate to support the social cognition of children with ASD who demonstrate both strengths and challenges in Early, Basic, and Advanced aspects of ToM. Many of the suggested interventions for Early and Basic ToM are drawn from those treatments identified by the National Standards Project (National Autism Center, 2009) with established or emerging evidence. Notably, most of those interventions described for supporting Advanced ToM do not have an evidence-base. Although not an exhaustive review, the interventions represent those frequently used to support children’s shared attention, emotion recognition, and perspective-taking. Notably, the interventions can be used to facilitate more than one component of ToM.

For those students whose scores on the ToMI indicate a need for support in the Early ToM category, interventions focusing on joint attention and social communication in the context of caregiver-child interaction and naturalistic settings are described. These include joint attention training, Floor time/DIR, Relationship Development Intervention (RDI), and More Than Words. For those students whose scores indicate a need for support in the Basic ToM category, interventions focus on increasing perspective-taking, comprehending a variety of mental states, and understanding the feelings and beliefs of self and others using Social Stories™, comic strip conversations, and Talkability™. For those students whose scores on the ToMI suggest a need for support in the Advanced ToM category, interventions emphasize understanding and use of idioms and sarcasm and detecting and accurately interpreting subtle cues in a social context through the use of Social Stories™, comic strip conversations, Talkability™, Social Thinking® and other social cognition curriculums.

Following the description of potential interventions that can support various aspects of ToM, a number of case studies are presented to demonstrate applications of ToMI scores to intervention planning including the articulation of specific treatment goals and the identification of programs that might achieve those goals. This manual concludes (see Appendix) with an example of report writing, which includes interpretation of standard and raw TOMI scores and how these can be used to describe an individual's competency across the three ToM subtests thereby informing intervention targets.

Interventions to Support Early ToM

Skills that are captured by the Early ToM subtest include sharing attention, basic emotion recognition, intentionality, and social referencing. An individual's capacity for ToM at the Early level might include the ability to engage in response to and/or initiation of joint attention towards an object, person or an event of interest. Additionally, an individual begins to develop understanding of communicative partners as agents of intentionality who communicate verbally and/or nonverbally with purpose. There is also the emergence of affect recognition allowing communicative partners to engage in interactions of shared enjoyment or other shared mental states. These instances of shared enjoyment or mental states often include the presence of social referencing, which is at the heart of the shared engagement often seen in early social routines, such as peek-a-boo.

Joint attention training. Children with ASD communicate primarily to regulate the behavior of others so they are less likely to point or show objects and make gaze shifts back and forth between a person and an interesting object or event (Prelock, 2006). There are two types of joint attention (i.e., response to and initiation of) that are important to establish in young children with ASD. The first, response to joint attention, involves the ability to 'read' the direction of another's eye gaze, head turn or point to infer the object of another's intention. It is seen in infants as young as 6 months and is established by 12-15 months in typical development (Sullivan et al., 2007). The second, initiation of joint attention, involves intentionally directing another person's attention to share an experience with that person. It usually develops by 12 months and is well established by 18 months (Schietecatte, Roeyers & Warreyn, 2012). Failure to develop joint attention has been linked to limited symbolic play, slower and less well developed language, and challenges in peer relationships (Mundy & Burnette, 2005; Murray et al., 2008; Wetherby, 1986) and deficits in social and cultural learning more generally. Thus, joint attention is often considered a priority treatment goal or a pivotal skill because of its critical role in fostering early social and symbolic connections (Koegel & Koegel, 2006; Prelock, 2006).

Several studies examining the effectiveness of joint attention training have been conducted over the last decade. Some studies have relied on parents as intervention agents for their children with ASD, following the logic that parents spend the most time with the child and thus will have the greatest number of opportunities to provide instruction. Other studies have targeted direct work by clinicians and special educators with children with ASD. Notably, the research in joint attention training has utilized strong experimental designs, including both randomized control trials and multiple baseline single subject designs. The effects of joint attention training have been powerful for increasing verbalization (Drew et al., 2002), facilitating reciprocal social interaction (Aldred, Green, & Adams, 2004), and increasing response to and initiation of joint attention that generalized to other contexts (Jones, Carr, & Feeley, 2006; Kasari, Freeman, & Paparella, 2006; Kasari, Paparella, Freeman, & Jahromi, 2008; Kasari, Gulsrud, Wong, Kwon, & Locke, 2010; Schertz & Odom, 2007; Whalen & Schreibman, 2003). Play is often used as the treatment context with both behavioral and developmental instructional strategies employed (White et al., 2011). Overall, joint attention is an established intervention important to establishing shared meaning making for children with ASD who typically have difficulty with this early developing skill.

Floortime™/Developmental Individual-Difference Relationship-Based (DIR). Floortime™ is a relationship-based intervention designed to facilitate a sense of relatedness between a child with ASD and a caregiver or interaction partner (Greenspan & Wieder, 1997a, 1997b, 1998, 2001). This intervention focuses on supporting children's social-emotional development as they master six developmental milestones including the ability to: 1) regulate and share attention; 2) engage in trusting and intimate relationships; 3) establish two-way communication; 4) problem solve in social-emotional situations; 5) use ideas functionally; and, 6) build connections between ideas (Greenspan & Wieder, 1998; 2001).

Greenspan and Wieder (1998) established four major goals for using Floortime™ based on the expected milestones for social emotional development. These include encouraging attention and intimacy, establishing two-way communication, cultivating the expression of ideas and feelings, and connecting logical thought—all goals requiring the development of early and foundational aspects of ToM including the sharing of attention and reading of affect. Floortime™ involves several 20 to 30 minute interaction periods throughout the day in which a parent or interaction partner is engaging a child through rough and tumble play to acting out adventures with toys (Greenspan & Wieder, 1998). Floortime™ is a child-centered intervention in which the interaction partner follows the child's lead while facilitating initiations. At first this interaction occurs nonverbally and then moves to more verbal encounters as the child increases their success linking emotions and behaviors. This is an important initial Floortime™ goal as it fosters a child's ability to attend to and establish an intimate connection with the adult. The adult expands on the child's actions and attempts to turn every behavior into a circle of communication (i.e., opening an interaction through verbal and/or nonverbal initiations and closing the interaction when there is a response to the initiation) (Greenspan & Wieder, 1998; Prelock, 2006). Initially, circles of communication are simple reactions but grow to communicate emotions through enticement (e.g., interventionist builds a bridge for the child's favorite cars to go under) and playful obstructions (e.g., favorite cars are out of reach and require the adult to access). The exploration of emotions is a critical component of Floortime™ (Gerber, 2012) as it builds on a child's emotion recognition—an early developing ToM construct.

Establishing two-way communication is a second Floortime™ goal that occurs by engaging in dialogue without words and capitalizing on facial expressions to communicate intentions. Goal three in the Floortime™ intervention approach encourages the expression and use of feelings and ideas—building on a child's developing emotion recognition and perspective taking. Typically, this is done through drama and make-believe as a clinician plays with a child with ASD using characters from a favorite video, asks questions about why a character is doing something or describes the feelings the character has when the child says or does something. The final goal for Floortime™ is to establish logical thought in the child with ASD. It requires the interventionist to support the ability of the child with ASD to link their ideas to their feelings and the feelings of others so the child establishes some logical and meaningful connection with the world. Intervention would support the child with ASD to explain why someone is hungry, how hungry they are, what someone likes to eat, how someone might feel if they don't get enough to eat and what could be done if someone could not get the food they want. At this point in the intervention the child is opening and closing multiple circles of communication, elaborating on the adult's content, and responding to the adults emotion.

Although not an established intervention, the available research indicates that Floortime™ is an emerging intervention approach that supports the social emotional development of children with ASD. Greenspan and Wieder (1997b) completed a retrospective chart review for 200 children and found that outcomes following intervention were good for 58% of the children and moderate for 25% of the children, while 17% of the children continued to have difficulty. They also reported that those children with the strongest outcomes continued to improve their socialization the longer they received intervention. A variation of the Floortime™ approach was compared to an adult-directed approach and results indicated significant positive results for Floortime™ (DeGangi & Greenspan, 1997). Ingersoll and colleagues (2005) also investigated the effectiveness of Floortime™ using a single subject multiple baseline design and found that children with ASD increased their spontaneous speech with the interventionist as well as their parents who had not been trained in the intervention. Wieder and Greenspan (2005) also reported on a follow up study 10-15 years after 16 boys with ASD received comprehensive intervention using Floortime™ and DIR consultation. The boys received intervention for two to five years between two and eight years of age. They exhibited significant improvements in their ability to be empathetic, creative, and reflective. They were also described as being successful in school and having positive peer relationships. Solomon and colleagues (2007) reported on a parent- training program they used for families of 60 children with ASD that incorporated

Floortime™.DIR. They designed an 8-12 month intervention where parents were asked to engage in 15 hours a week of one-to-one interaction with their children. Pre- to post-intervention results revealed 45.5% of the children achieved good to very good progress with parent satisfaction at 90% (Gerber, 2012). Most recently, a pilot study conducted by Pajareya and Nopmaneejumruslers (2011) revealed that parents who received home-based training using the principles of DIR and incorporated Floortime™ strategies approximately 15 hours a week saw positive outcomes in their children's ability to relate, engage, and communicate. A single subject AB design with a three and a half year old child with ASD found that following parent training in Floortime™ strategies, the child increased the number of circles of communication and the parent reported greater satisfaction in interacting with her child (Dionne & Martini, 2011). Like joint attention training, Floortime™ has real value for establishing affective connections in children with ASD with their communicative partners. Further, although this intervention primarily builds on Early ToM development, it can and does address some of the more Basic and Advanced aspects of ToM discussed later in this chapter.

Relationship Development Intervention (RDI). RDI (Gutstein & Sheely, 2002a, 2002b) is similar to Floortime™ as it capitalizes on shared and joint attention that occurs naturally between children and their parents. It follows three principles of intervention (i.e., functions precede means, social referencing, and co-regulation) with targeted behaviors to be developed (Gutstein, 2000). The first principle, function precedes means, defines the reason for a particular behavior or skill. One of most basic functions is emotion sharing—a key element to Early ToM development. Competence is another critical function in that children with ASD must be proficient in sharing their emotion, take responsibility for their role in an interaction, and manage when the unexpected occurs. Two strategies that capitalize on the principle of function precedes means include guiding and pacing. Guiding facilitates the selection of experiences to share, while setting clear limits for expected behavior. Pacing requires modifying one's communication style so more complex skills are broken down into simpler ones that can be practiced to ensure mastery. For example, while a mom and her son are playing with a tambourine, instead of guiding her son's participation with her voice telling him what to do with the instrument, she would instead nod her head and show her son how to hold the tambourine. The child would then imitate his mom's tapping of the tambourine using one hand and gradually adding shaking the tambourine with his other hand.

The second RDI principle involves the child with ASD in social referencing or the ability to perceive and process relevant information in social relationships. Social referencing is a foundational skill for connecting with an interactive partner. Indirect prompts are used to give the child social clues related to the next steps in a social encounter. Co-regulation is the third principle which requires the child with ASD and their parent or another adult to work together in an interaction so that they can be successful sharing an experience.

There are several levels with multiple stages to the RDI intervention model. Goals are established within each stage with a process for determining skills acquisition. As an example, there are four stages in level one, *Laying the Foundation for Relational Development*. These stages (emotional attunement and attending, social referencing, excitement sharing or regulating, and coordinating actions as in simple games) are particularly relevant to a child's developing ToM. The interventionist's goal in emotional atunement is to facilitate a child's face-to-face gazing through laughter, sharing joy or soothing distress. In social referencing a child is expected seek out the adult's facial expressions to reference what actions might be expected, particularly in unfamiliar situations. For excitement sharing or regulating, novel elements of an experience are shared to foster a child's increased challenge and excitement. Finally, in coordinated actions the child learns how to connect his actions with the adult's when they are engaged in shared activities.

Currently, research is limited on the effectiveness of RDI although some preliminary research reported positive results in communication and social interaction as well as autism symptoms for 10 children, ages 5 to 11 years (Gutstein, 2007). After participating in RDI, Gutstein noted that three children were no longer classified on the autism spectrum and three other children appeared to have reduced symptoms. Parental perceptions of change were also reported with some of the greatest improvement in excitement, nonverbal language, initiation, eye contact, and sharing emotions (Gutstein, 2007; 2009). Similar to the strategies in Floortime™, RDI supports experience sharing and reciprocal communication—valuable components for fostering early perspective taking skills in children with ASD.

More than Words—The Hanen Program for Parents of Children with ASD. More Than Words (MTW) is a parent training intervention designed to support the communication and social skills of children with ASD (Sussman, 1999). The intervention focuses on teaching parents that communication depends on a child's ability to pay attention, find enjoyment in reciprocal communication, imitate and understand others, have fun interacting with others, and practice what is learned. Because the program focuses on parents' understanding of what communication is, their child's learning style, and the function of their child's communication, parents learn ways to affect their child's communication.

There are four major goals in the MTW program: 1) help children interact with others; 2) support communication in new ways; 3) teach communication for new reasons; and, 4) connect what is being said with what is actually happening (Sussman, 1999). Through the first goal, parents learn that their children can and do enjoy doing things with them and children learn interaction with their parents creates enjoyment. During intervention focused on the second goal, parents help their child communicate using a number of methods, including the use of visual supports and gestures. Intervention focused on goal three facilitates intentional communication as parents set up situations that require their child to make requests to get what they want. To achieve the final intervention goal, parents make sure what they say is meaningful so their children can make connections between what is said and what is happening. Parents develop their interactions around their children's motivators—preferences or interests that are likely to engage them at their communication level.

MTW has some emerging evidence to support its use with young children with ASD and their families. The first controlled trial using the MTW program principles to enhance parent understanding of ASD and support social communication in their children with ASD had a measurable impact on both the parents' and children's communication skills (McConachie, et al., 2005). A second study examining the social interaction of three children with ASD between the ages of two and three, following their mothers' participation in MTW yielded positive results for the mothers' use of spontaneous interaction strategies and the children's increased vocabulary (Girolametto, et al., 2007). A recent multisite study revealed that children who played with a limited number of toys showed more improvement in their communication skills following their parents' participation in MTW compared to those children in a community-based intervention only. Children showed gains in initiating joint attention, requesting and communicating intentionally which were maintained four months post intervention (Carter et al., 2011). Prelock and colleagues (2011) also examined the impact of MTW on four children with ASD and their families and found increased social and symbolic communicative acts from pre-to post-intervention and increased vocabulary. As an intervention for supporting a child's emerging theory of mind, MTW fosters opportunities for joint attention and reciprocal communication as well as emotion sharing—key elements to a developing ToM.

Interventions to Support Basic ToM

Skills that reflect Basic ToM competencies involve metarepresentation and developmentally-related ToM understandings. For this level of ToM, individuals progress into more complex emotion recognition and affective reasoning, develop mental state term comprehension, understand physiological- and affective-based behaviors, and learn the distinction between appearance and reality as well as the ability to distinguish between mental and physical entities. Additionally, the capacity for explicit first-order thinking and visual perspective-taking emerges, further informing one's own and others' beliefs about the world. This capacity for perspective-taking is associated with the metarepresentation of others' thoughts, feelings, and beliefs. It is also associated with the ability to make inferences about shared knowledge or experiences. The individual begins to understand that "seeing leads to knowing" and that multiple individuals can have varying perspectives of the same object or event. The social reasoning that occurs with the consideration of others' mental states and perspectives is, in turn, associated with the demonstration of empathetic and deceptive behaviors.

Social Stories™. Social Stories™ (Gray, 2010) provide a way for children with ASD to navigate those daily social interactions they find particularly confusing or troublesome. Social Stories™ are short stories that use printed words or words paired with pictures to highlight relevant social cues often missed by children with social pragmatic challenges, particularly those with ASD. Social Stories™ consider several sentence types, but most often they include: 1) *descriptive* sentences-providing information about the people, setting and activities involved; 2) *directive* sentences-defining the child's expected behavior; and, 3) *perspective* sentences-describing the beliefs, feelings or emotional reactions of others (Gray, 1995; Gray & Garand, 1993). Typically, they are developed in collaboration with the child and family or educational team to target problematic behaviors. During intervention, a priming strategy may be used where the story is read prior to an anticipated challenging situation and there is some support for the notion that this is facilitative of optimal outcomes (Scattone, 2007). It should also be noted that Social Stories™ do not have to be used in the context of challenging situations and Gray (2010) recommends that they also be used to celebrate success and can be effective in giving meaning to praise (Hutchins, 2012).

Social Stories™ are theoretically potent for addressing Basic ToM skills (as well as Advanced skills discussed below) as they focus on the understanding of internal states and their relations to behavior and social context. In fact, Gray (1998) suggested that impairments in ToM and the tendency of individuals with ASD to acquire the cognitive style of weak central coherence limit their access to social knowledge. Thus, the theory behind Social Stories™ is that they "translate these 'secrets' surrounding social interaction into practical, tangible social information" (Gray, 1998, p. 169). Because they draw on what is personally relevant and motivating for the child with ASD and focus on visual versus verbal information, Social Stories™ are seen as particularly effective for children with ASD.

Social Stories™ have been identified as one of 11 established treatments in ASD (National Autism Center). Social Stories™ have been found to advance the social communication of children with ASD by increasing greetings, requests, comments, and appropriate word use (Adams, Gouvousis, VanLue, & Waldron, 2004; Delano & Snell, 2006; Scattone, Tingstrom, & Wilczynski, 2006; Swaggart et al., 1995; Thiemann & Goldstein, 2001).

They can also facilitate the child's the use of compliments (Dodd et al., 2008), labeling and explaining emotions (Bernad-Ripoll, 2007), and two-way conversations (Crozier & Tincani, 2007; Sansosti & Powell-Smith, 2006; 2008). Further, they have been successful in decreasing undesired behaviors during social interactions such as echolalia, excessive voice intensity, and tantrums (Brownell, 2002; Lorimer et al., 2002; Norris & Datillo, 1999).

Social Stories™ have traditionally been defined as most appropriate for school-aged children who have some interest in print and/or have an ability to read. They have been successfully used, however, with younger children and those with limited verbal abilities (Hutchins and Prelock, 2013a, 2013b). If using Social Stories™ with children who display lower cognitive abilities, some modifications may be required such as using simple and fewer sentences that are paired with familiar pictures or visual symbols. It is also important to identify and define the social behavior requiring change, and to collect baseline data. The child's ability level is a consideration for the number and linguistic complexity of the sentences used to ensure the child is processing the relevant information. Notably, Gray (1994,1998) suggests comic strip conversations are prerequisite strategies for the development of Social Stories™ (for a recent review of Social Stories™, see Hutchins, 2012).

Comic Strip Conversations. Comic strip conversations are simple drawings that illustrate an ongoing conversation for children with ASD who struggle to understand the rapid exchange of information in reciprocal communication. Also developed by Gray (1994, 1998), comic strip conversations are theoretically similar to Social Stories™ except they are used while talking and drawing about a social situation. In collaboration with the interventionist, the child identifies the who, what, when, where, and why (Gray, 1994) surrounding a challenging situation as well as possible solutions to the problem. This is done using drawing (e.g., stick figures, talking bubbles, thinking bubbles) and writing conventions so as to talk about an event using a structured visual format. With the interventionist, the child with ASD creates a potential plan for each solution, talking through the advantages and disadvantages of each, and prioritizing the most appropriate solutions while discarding those that are not feasible (Prelock, 2006). Although there is limited literature to support the use of comic strip conversations, Glaeser, Pierson, and Fritschmann (2003) report on success with one student with ASD who had fewer incidents on the playground and in the classroom following intervention using comic strip conversations. More recently, Foran, Hutchins, Prelock, & Murray-Close (in review), Hutchins and Prelock (2013a), and Vivian, Hutchins, and Prelock (2012) found positive results in parents' perception of behavior change in their children with ASD following intervention using comic strip conversations.

Gray (1998) suggests that comic strip conversations might be most effectively used with children with Asperger syndrome or children with ASD who have higher cognitive abilities, as they are required to identify what another person is thinking. Again, however, recent research suggests that comic strip conversations may be used effectively for children with more severe limitations (Hutchins & Prelock, 2013a) although research in this area is generally lacking.

Talkability: People skills for verbal children on the autism spectrum. *Talkability™* is a program for parent-led intervention in ToM skills for verbal children who have been identified with ASD, Nonverbal Learning Disorder (NLD) or impairments in social communication (Sussman, 2006). This intervention approach educates parents in strategies for interacting with their children and supports the social interactions between their child and his/her peers. This program draws from the underlying premise that children navigate interpersonal relationships and particularly, conversation and social interactions, by understanding that words and behaviors have meaning and that they learn to reflect on the thoughts and feelings of others.

Talkability™ begins by determining a child's learning style, sensory preferences and aversions as well as the caregiver's interactive style. Next, caregivers are taught to instruct their children on how to understand nonverbal communication and ways to initiate, maintain, and end conversations. Then constructs underlying the child's ability to "tune-in to others" and use perspective-taking are introduced. Talkability™ addresses the Basic skills (e.g., metapreresentation, pretend play) intended to be tapped by the ToMI but it also progresses into the Advanced skills as it progresses into more complex aspects of metarepresentation. These involve, for example, the nature of the child's friendship skills and the intervention provides strategies for effective "coaching" during naturalistic social interactions.

The development of ToM as well as the benefit of "parent talk" in a child's developing ToM provides the theoretical framework for this program; however, no research studies are available examining the effectiveness of implementing the components of the Talkability®. Sussman (2006) recommends that this program is most useful for children who are verbal and between the ages of 3 and 7 years. She indicates that children may have a diagnosis of Asperger syndrome, High functioning Autism or NLD although the program could be applied to a student with no diagnosis who demonstrates a social communication disorder with specific challenges in conversational skills and social relatedness.

Interventions to Support Advanced ToM

The dimensions that encompass Advanced ToM competencies as assessed by the ToMI include: second-order thinking, interpretation of complex and second-order emotions and thinking, complex visual perspective-taking, complex social judgments, the capacity for empathy and understanding the nuances of language such as seen in humor, idioms, and sarcasm. An individual's social understanding at the Advanced level might include: making abstract aspects of language more concrete; drawing attention to the nonverbal and verbal features of a communicative partner's language that signal changes in intentionality; and, instructing students on the "hidden rules" of changing social contexts. Additionally, the focus is on increasing gestalt processing and problem solving skills that inform a greater understanding of social themes. Advanced ToM skills also consider increasing a child's ability to attend to and interpret others' thoughts and beliefs in order to navigate dynamic and novel social situations. Often times, instructing children on how and why their behaviors affects others' thoughts and feelings leads to greater understanding, and therefore, an increase in competency within social exchanges.

Although there are many skill-specific resources and a few comprehensive intervention programs available for supporting more advanced aspects of ToM, we have chosen only a handful of curricula and resources to present here. As opposed to the majority of strategies described above which are considered established or emerging interventions, interventions to support advanced aspects of ToM have (generally speaking) not yet been the focus of rigorous empirical scrutiny. As such, the interventions are chosen for inclusion here, not on the quality of the evidence-base, but on the basis of their popularity in clinical practice with evidence of efficacy coming almost entirely from clinician testimonials. It is also important to note that those interventions previously described for supporting Basic ToM skills (i.e., Social Stories™, Talkability™ and comic strip conversations) can also be appropriately applied to facilitating aspects of Advanced ToM.

Comprehensive Interventions to Increase Advanced Theory of Mind

Building Social Relationships. Building Social Relationships is a curricular guide that presents a five-step model for developing social skills programming for children and adolescents with ASD. It involves explicit social skills instruction emphasizing frequent practice within the context of meaningful social exchanges occurring in the natural setting (Bellini, Peters, Benner, & Hopf, 2007; Gresham, Sugai, & Horner, 2001). Bellini (2006) suggests that for a program to be successful, it must be designed using several guiding principles. First, it must consider that children with ASD often want to establish meaningful relationships. Second, social skills should be taught that increase the success of a child in the social setting. Third, instead of teaching "appropriate" or polite behaviors, targeted skills should consider what is expected for the student's age and the social context. This is followed by a focus on functional skills that elicit a positive response from a student's peers. Next, a student should be taught how to flexibly recognize and adapt to both the environment and to the verbal and nonverbal cues of a communicative partner.

Bellini (2006) highlights the importance of identifying the obstacles that preclude a child from developing and maintaining successful social relationships. He recommends that a team of trained professionals conduct interviews and observations across both structured and unstructured settings. This information is then combined with the results of several standardized assessment tools as well as checklists and interviews provided in the guide. As result of this comprehensive assessment, guidance is provided on developing appropriate goals and objectives for the student.

The next step in Building Social Relationships is to distinguish between competence and performance. That is, it is important to differentiate the lack of skill acquisition from a performance-deficit as these are likely to have different implications for intervention planning.

Following the identification of social skills to be targeted, the next step is Selection of Intervention Strategies. This involves considering several intervention approaches ranging from computer-based mind reading programs to less structured role playing that facilitates cognitive flexibility and more structured approaches to teaching theory of mind, such as the social story™ and video-modeling. Additionally, emotion and behavior regulation strategies, and social problem solving approaches are presented. Following a discussion of intervention strategies to promote skill acquisition Bellini makes several recommendations for strategies that enhance performance. Reinforcement and contingency strategies, gaming skills, environmental modifications, peer-mediated interventions, and priming are included.

After interventions have been selected to target skill acquisition and/or performance it is recommended that the clinician consider the implementation of the intervention through thoughtful planning guided by several considerations. These considerations include the format of the program (e.g., group or individual), selection of peer models, assembling the team, gathering materials and resources, and determining the length and frequency of sessions. Additionally, connecting the treatment objectives with the activities embedded within each session is stressed, as well as defining prompt hierarchies, behavioral supports, reinforcement, and data collection procedures.

The final step is evaluating and monitoring progress, which purports to guide the clinician through taking baseline measurements at the start of intervention as well as follow-up assessments along the way (e.g., every three months, quarterly, etc.). Recommendations for recording data include frequency recording (the number of times a behavior occurs), duration recording (the length of time a behavior occurs), time sampling (recording behaviors at intervals), latency recording (amount of time between a stimulus and a response), and response ratios (frequency of response in relation to the number of opportunities.). The manual describes these procedures and their ideal uses as well as discusses factors affecting intervention implementation and progress monitoring.

Bellini (2006) does not explicitly state any cognitive and linguistic requisites for application of the principles, but indicates ideal use is to develop comprehensive social interactions skills programming for children and adolescents with ASD and/or those students with related social difficulties. Building Social Relationships presents a broad range of assessment tools and intervention techniques that can be chosen or modified based on verbal abilities and/or developmental level.

Social Thinking®. Social Thinking® (Winner, 2000, 2007) is a cognitive behavioral approach to intervention that aims to foster students' social communication skills and increase their social understanding through the instruction of a set of core principles and concepts. The curriculum is embodied in the ILAUGH model (Winner, 2000, 2007) with a focus on: *Initiation*—the ability to use verbal and nonverbal communication skills to effectively begin an interaction, seek assistance or gather information within the social setting; *Listening with Eyes and Brain*—the ability to interpret social cues by integrating auditory and visual information (e.g., nonverbal behaviors) to formulate a deeper understanding of the social message; *Abstract and Inferential Language/Communication*—the ability to recognize that most language is not intended for literal interpretation and requires flexibility in thinking about the intended message in a given context; *Understanding Perspective*—the ability to understand the emotions, thoughts, beliefs, experiences, motives and intentions of one's self and others which is crucial to participation in groups and in interpreting information; *Gestalt Processing*—the ability to convey information through concepts and not just facts such as making an inference or a prediction about an underlying concept being discussed by synthesizing multiple details and perspectives across communicative partners; and, *Humor and Human Relatedness*—the ability to appreciate humor, relate to others, engage in shared enjoyment as well as in use self-regulatory patterns that allow individuals to effectively share space with others. Understanding and use of these six basic concepts and/or skills are considered integral to the development of effective communication skills as they allow one to initiate, negotiate, and terminate social interactions in an efficient manner and are necessary for the development of empathy and emotional regulation.

The Social Thinking® approach is intended to provide students with instruction, through coaching, in "how" and "why" their behaviors affect others' thoughts and feelings, thereby creating a positive or negative impact on others' impressions. This intervention was developed in response to some of the limitations of traditional social skills training programs where discrete social skills, or individual dimensions of ToM are taught with little generalization to social competency in "real life" interactions (Krasney, Williams, Provencal, & Ozonoff, 2003). Teaching students how to relate and respond to others' emotions as well as one's own, while also facilitating shared enjoyment is critical to the development of a sophisticated social cognition and Advanced ToM, which is a primary goal of this model. The Social Thinking® model relies heavily on teaching the meta-cognitive and metalinguistic underpinnings of pragmatic language and social communication; therefore, specific requisite cognitive and linguistic skills are recommended. It is suggested that appropriate candidates possess at least average to above average intellectual capacity and strong to excellent language abilities. This intervention is most appropriate for individuals functioning on the higher end of the autism spectrum.

Social Thinking® is both popular and aggressively marketed, however, very little data have been offered in support of its efficacy with anecdotal reports of effectiveness coming primarily from practitioners in the field and the developers of the intervention themselves. In a single study, Crooke, Hendrix, and Rachman (2007) assessed the effectiveness of Social Thinking® for six boys with Asperger's Syndrome (AS) and High Functioning Autism (HFA) using a pre-post no control design. The authors categorized verbal and nonverbal behaviors as "unexpected" or "expected," which were measured during unstructured and semi-structured social exchanges, at baseline and then during the successive eight weeks of 60-minute treatment. Results indicated a robust and significant increase in "expected" behaviors from baseline to generalization although limitations in treatment time, sample size, familiarity with the interventionists and lack of a control group were noted.

Skill-Specific Resources for Supporting Advanced Theory of Mind

***Navigating the social world: A curriculum for individuals with Asperger's syndrome, high functioning autism, and related disorders*.** Navigating the Social World is a resource guide that aims to support parents, clinicians, and paraprofessionals in the development of social-emotional skills programming for students with Asperger's Syndrome, high functioning autism and related socio-cognitive disorders (McFee, 2004). The manual includes several components. The first section focuses on basic nonverbal communication and emotion recognition. Additionally, the authors present a few techniques for recognizing stress, self-monitoring, and facilitating stress regulation. Section two focuses on more general social and communication skills by presenting discrete skill training in areas such as conversations and social scripts to support asking for help, and then moves more broadly to conflict resolution. In section three, abstract thinking skills and decoding figurative language are emphasized. The final section focuses on behavioral issues often associated with deficits in social skills.

In terms of the cognitive and linguistic prerequisites that are needed to successfully use the curriculum, McAfee indicates that most of the activities included in this resource are targeted towards students needing "basic instruction". However, it is also noted that there are some activities geared towards more high-functioning individuals with autism, suggesting the need for verbal and cognitive abilities that would support candidacy for participation in those programs. Notably, program activities can be adapted and individualized to meet the skill levels of most students.

***Teaching Perspective Taking to Children with Autism Spectrum Disorders*.** This is a teaching program and resource guide for parents, teachers and clinicians working with young children with autism (Brennan, 2011). The guide purports to synthesize research on the core deficits often associated with ASD and was developed based on several theoretical hypotheses, including: *affective theory*, which suggests that the social difficulties often seen in ASD can be attributed to an underlying disturbance in the capacity for emotional interaction (Kanner, 1943); *cognitive theory,* which suggests an underlying disturbance in the ability to infer mental states in the self and others (Baron-Cohen, 1988, 1995; Baron-Cohen, Leslie & Frith, 1985); and *intersubjectivity theory,* which hypothesizes an underlying deficit in formulating and coordinating representations of the self and others (Rogers & Pennington, 1991). Brennan (2011) notes the difference across these theories, however also highlights a set of common and core deficits, such as pretend and symbolic play, pragmatic language skills, and theory of mind, which serve as the main components targeted within this resource.

Each chapter in the guide contains a teaching program that is broken down into individual steps with suggestions for generalization. The curriculum begins with early developing non-verbal skills, such as eye gaze and requesting with eye contact, progresses through more distal eye-gaze tracking, pointing and then more complicated attention switching. Ideally, mastery of these skills serves as a pre-requisite prior to moving to more sophisticated symbolic language and play development. Subsequent strategies are presented to support the understanding of thoughts, emotions, and mental states more generally, to enhance the metacognitive skills required for intentionality and conversational skills, and advance perspective-taking by focusing on interventions that teach false belief, nested beliefs, and the subtle nuances of social language.

As described by Brennan (2011), the early chapters (1-3) are recommended for teaching children with little verbal language (and which can be modified for nonverbal children with autism). The later chapters (4-6) target skills for children who demonstrate mastery of the pre-requisite skills outlined in the earlier chapters and with receptive and expressive language abilities typically seen in children 4 to 7 years of age.

***Teaching Children with Autism to Mind Read: A Practical Guide*.** This resource is offered as a practical guide based on intervention techniques developed in an experimental study focusing on key aspects of social understanding, such as mental-state teaching or "mind-reading" (Howlin, Baron-Cohen, & Hadwin, 1999). The guide was developed on the premise that naturalistic teaching should occur by breaking down the student's needs into smaller steps, following a typical developmental progression and with systematic reinforcement within the interactions. Further, the authors suggest that by using errorless learning procedures the student does not have a chance to practice mistakes, which facilitating rapid acquisition. Lastly, Howlin and colleagues suggest that one way to enhance generalization is to teach the underlying concepts, thereby avoiding solely training to the task.

The guide is divided into three main parts categorized by different classes of mental states. Within each section the reader is provided with the general principles underlying that stage, assessment procedures, ways to establish a baseline, materials needed. and teaching procedures to be used. Teaching Children with Autism to Mind Read was developed for children with ASD between the ages of 4 and 13 years. The authors indicate that appropriate candidates should demonstrate verbal abilities typical for a five year old, the age at which children typically begin to "mind read". Although the materials have been developed for instruction through age 13, the authors suggest the potential for adapting the materials to meet the individual needs of older students.

***The Hidden Curriculum*.** The Hidden Curriculum (Myles, Trautman, & Schelvan, 2004) is a resource that provides guidelines or a "set of rules" that are not explicitly taught in school or academic settings but play an important role in social competence and navigating social interactions. The curriculum focuses on observing and interpreting nonverbal social cues, understanding the implicit rules embedded in social interaction, and mastery of idiomatic and metaphorical expressions with the aim of providing instruction that will be relevant and useful for a variety of social situations. In addition to outlining settings or situation-specific rules, this resource also presents several strategies for observing behavior, problem-solving, and modifying behaviors.

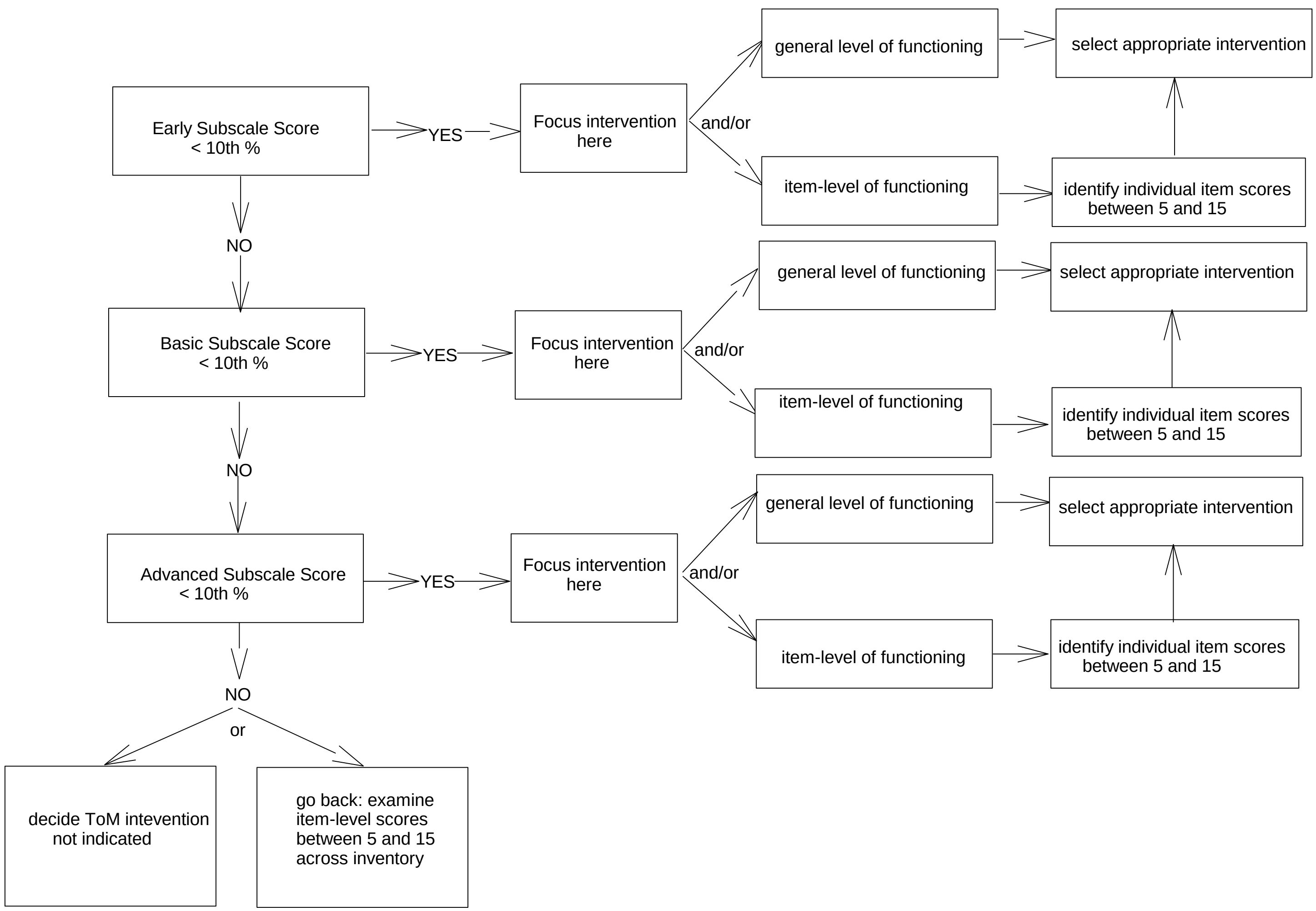
Myles and colleagues (2004) do not explicitly state any pre-requisite cognitive and linguistic skills that affect candidacy for instructional use of this resource. There are many components that refer to the school setting, suggesting it's appropriateness for students who are school aged, however, there is also a section on hidden rules that are common in the workplace.

Clinical Application of the ToMI: Interpreting Scores

We encourage clinicians to apply what they have learned about a child's ToM following administration of the ToMI in different ways depending on the purpose of assessment. For the purposes of identification, recall that our sensitivity and specificity data suggest that a criterion of less than the 10^{th} percentile for the ***composite*** score is useful in the identification of ASD although the ToMI is not designed as a tool that will necessarily aid in differential diagnosis at this time. For the purposes of treatment planning, results can be used to create profiles of social cognitive strengths and challenges using two strategies. One strategy is to consider the child's subscale (standard) scores to establish areas of *general* intervention focus at the Early, Basic, or Advanced levels of ToM. A second approach is to look within (and sometimes across) the subscales at the item-level (raw scores) to identify *specific* areas of strength or weakness to guide treatment planning. We imagine that clinicians will find it useful to consider scores at both levels to inform whether treatment planning is most appropriately conducted to target general or specific ToM areas (or both).

With regard to general ToM levels, attention should focus on subscale percentiles. In alignment with the sensitivity and specificity data described for composite scores, we recommend that less than the 10^{th} percentile be used as a criterion for the identification of subscale scores that can be considered in the clinical range (again, this is denoted by an asterisk on the online generated report). Examination of these norm-referenced subscale scores may be conducted to zero-in on *general levels* of intervention focus and various intervention procedures and programs were described earlier in this chapter to inform that purpose. When more than one subscale indicates functioning in the clinical range, we recommend that intervention begin at, or at least incorporate, the lowest level of functioning which corresponds to the earliest stages of typical ToM development. This is consistent with a descriptive-developmental approach to intervention and is a common standard in clinical sciences including communication sciences and disorders (Paul, 2007). In line with this descriptive-developmental approach, an ***idealized*** decision-tree for guiding the treatment planning process using ToMI subscale scores in presented in Figure 4.

Figure 4: Idealized decision tree for treatment planning using the ToMI.



Scrutiny of subscale percentiles can be followed by an examination of specific areas of ToM functioning at the item-level in which case the following guidelines for decision-making are recommended. First, raw item-level scores falling below 5 are taken as indicators that the child has yet to develop this ToM competency. Intervention for this target may not be appropriate, however, and this should be informed by whether the subscale percentile for this item falls in the clinical range (discussed immediate above). For example, a child could obtain a score of 4/20 on a particular item and although this indicates a low score on the item, it may be an age-appropriate score; thus, intervention is not indicated. But even when a subscale percentile indicates functioning in the clinical range *and* a raw score on a particular item within that subscale is less than 5, intervention for that target may still not be appropriate. To explain why this is so, it is first instructive to note that significant heterogeneity has been documented in the social cognitive functioning within individuals (e.g., Gonzalez-Gadea et al., 2013) and heterogeneous scores on items within subscales has been observed on the ToMI as well. So, although Early, Basic, and Advanced levels represent statistically and theoretically coherent subscales, variation in item scores is not only assumed across but also *within* subscales. Consequently, it is possible that the level of competency for a particular item may be so low that that area represents a target that is not developmentally appropriate and where the child is unlikely to benefit from training. Vygotsky's (1934) concept of the Zone of Proximal Development (ZPD; the discrepancy between a child's present competency and the level s/he reaches in solving problems with assistance) is instructive here and we have developed some tentative opinions for which levels of competency (as measured reliably on the ToMI by assessing primary caregiver's confidence in a child's acquisition of a particular ToM skill) may be most responsive to treatment. These rough and rudimentary guide posts are as follows:

*** NOT DEVELOPED: Item-level scores falling below 7.0** (i.e., scores falling in the range of "probably not" and including "definitely not"): caregiver confidence in the presence of this ToM knowledge area or skill set is very low. Competence is not readily apparent, is not actualized, or exists primarily as a potential. Clinicians should consider whether these aspects of ToM are developmentally appropriate targets for intervention. Advanced skills may be unlikely to benefit from intervention directed at this competency at this time. Clinicians are encouraged to focus on early and foundational ToM competencies that are pivotal to other areas of functioning. The clinician might also consider basic level skills as appropriate.

***UNCERTAIN: Item-level scores between 7.1 and 12.9** (i.e., scores falling between the range of "probably not" and "probably"): caregiver is unsure whether competency is present and uncertainty is either informed or uninformed. When caregiver uncertainty is rooted in lack of item clarity or lack of relevant information (e.g., the caregiver hasn't had the opportunity to observe the child in a situation that would demonstrate this ToM aspect), the items should be omitted from interpretation. On the other hand, when caregiver uncertainty is informed, it may reflect inconsistent child performance or partial acquisition of a ToM skill (e.g., the caregiver might think "sometimes he seems to understand this and other times he doesn't"). This is generally considered a good starting point for intervention as it suggests some degree of understanding of this ToM aspect. There is potential that training in this area will provide opportunities for meaningful growth.

***DEVELOPED: Item-level scores above 13** (i.e., scores falling in the range of "probably" and including ratings of "definitely"): caregiver reports relative certainty that this competency is present. Score suggests that this ToM knowledge area or skill set is established.

Thus, this area of ToM development is not a priority for intervention unless intervention is deemed necessary for enhancing the generalization or maintenance of this skill.

Although our recommendations are based on theory and our clinical impressions of the use of the ToMI thus far, they are provided with the hope that these benchmarks will provide strategies and some level of confidence for clinicians seeking to make informed decisions about treatment planning. We also imagine scenarios in which these guidelines can be appropriately jettisoned as illustrated below. In the section that follows, we present five case studies that demonstrate the use of the ToMI to guide clinical decision-making. In these case studies, we consider both subscale and individual item scores to guide intervention targets and offer suggestions for appropriate treatment strategies. In many cases the recommendations for treatment are in alignment with suggested benchmarks provided above and other times they are not. In each case, a rationale is provided.

Clinical Application of the ToMI: Case Studies

Case Study 1: ST. ST is a 4 year, 9 month old male diagnosed with ASD. He uses some instances of intentional communication, some two word combinations (e.g., want more), some maladaptive behaviors to communicate discontent, and demonstrates difficulty with motor speech coordination. He is currently in the early stages of using the Picture Exchange Communication System (PECS) and a few signs for communicating his basic needs (e.g., more, all done). ST enjoys music, cause-and-effect toys, as well as games that involve water, tickling, and those that have a sensorimotor component (i.e., swinging, chase). His mother completed the ToMI and the results are presented below.

COMPOSITE SCORE

Composite Mean = 5.081 Percentile = 1st-6th*

EARLY ToM SUBSCALE

Subscale Mean = 11 (of 20) Percentile = 1st-6th*

Subscale Item Scores

Item	designed to assess	score
3.	affect recognition (complex)	10.0
6.	affect recognition (emotion-expression relationship)	13.5
24.	Intentionality	0.0
25.	affect recognition (basic)	16.0
29.	social referencing	15.5
37.	sharing attention – initiating	6.0
38.	sharing attention – responding	16.0

BASIC ToM SUBSCALE

Subscale Mean = 4.3526 (of 20) Percentile = 1st-6th*

Subscale Item Scores

Item	designed to assess	score
1.	physiologically-based behavior	12.4
4.	emotion-based behavior	15.2
7.	mental state term comprehension	0
8.	false beliefs in context of unexpected change of location	0
9.	seeing leads to knowing	5.0
10.	mental state term comprehension	.2
11.	appearance-reality distinction	10.0
12.	false beliefs in context of unexpected contents	0
15.	Certainty	0
16.	mental-physical distinction	14.0
26.	Pretense	12.5
29.	counterfactual reasoning	0
30.	mental-physical distinction	11.4

31.	ability to deceive	0
32.	level 1 visual perspective-taking	2.0
33.	speech acts	0
35.	speech acts	0
39.	mental state term comprehension	0
42.	attribute-based behavior	0

ADVANCED ToM SUBSCALE

Subscale Mean = 3.3563 (of 20) Percentile = 1st-6th*

Subscale Item Scores

Item	designed to assess	score
2.	Sarcasm	5.0
5.	second-order false desire attribution	0
13.	idiomatic language	5.6
14.	use of language to intentionally deceive	5.1
17.	understanding display rules	14.5
18.	complex social judgment	3.0
19.	white lies	0
20.	understanding lies versus jokes	2.0
21.	level 2 visual perspective-taking	0
22.	second order understanding of belief	0
23.	second order understanding of emotion	3.0
27.	complex social judgment	8.0
34.	Empathy	7.5
36.	humor (play on words)	0
40.	biased cognition	0
41.	mind as active interpreter	0

According to the ToMI results, ST demonstrates the greatest competency in Early ToM (subscale mean = 11.00), less competency in Basic ToM (subscale mean = 4.35), and the least competency in Advanced ToM (subscale mean = 3.35). This makes sense in light of the typical developmental progression of ToM where Early ToM is mastered in the first two years of life, Basic skills are mastered in the preschool years, and Advanced ToM is typically acquired in middle childhood (approximately between ages 7 – 9; Hutchins, et al., 2012). In light of ST's age then, advanced ToM skills should not be considered appropriate targets at this time. It is also noteworthy that scores for the composite and all subscale scores fell below the 10^{th} percentile and this is not surprising given ST's diagnosis of ASD (and a very common result as well).

With regard to scrutiny of item-level Early subscale scores, scores for intentionality fell in the 'latent' range, scores for two dimensions of affect recognition (complex and emotion-expression relationship) and for initiating shared attention fell in the 'emerging' range, and scores for basic affect recognition and responding to bids for shared attention fell in the 'developed' range. Although ST has yet to master most of the Basic ToM skills, the understanding of physiologically-based behavior, seeing-leads-to-knowing, and several aspects involving metarepresentation (e.g., appearance-reality, mental physical distinctions) appear to be emerging. Additionally, while he would not yet be expected to demonstrate an Advanced ToM, he may have a relative strength in understanding display rules.

From a descriptive-developmental approach, effective intervention should focus on Early ToM skills but may also include some emerging skills captured in the Basic subscale of the ToMI. More specifically, intervention should target increasing shared attention and engagement particularly in the area of initiation. Not only does there appear to be readiness for this intervention target, but it is a foundational and pivotal ToM skill. Intervention should also focus on complex affect recognition where ST demonstrates some competency but where there is clear room for advancement. If intervention were to include Basic ToM targets as well, we recommend focusing on increasing ST's understanding of physiologically-based behaviors, the appearance reality distinction, the mental-physical distinction, and strategies to support engagement in pretend play. As described above, Advanced skills are not expected given ST's age, are deemed unlikely to be responsive to treatment based on item-level scores, and are not recommended targets at this time. The following goals focused on a strategically selected set of Early and Basic ToM skills for supporting ST and were as follows:

Goal Area 1: initiation of shared attention
ST will initiate shared attention (e.g., eye gaze, verbalization or gesture) towards a highly preferred object that is placed out of reach for __ out of ____ opportunities across three consecutive sessions.

Goal Area 2: physiologically-based action
ST will match a character's mental state, specifically a basic physiologic state with a behavior in a picture-based story when presented with at least one situation nal cue with ___ accuracy.

Goal Area 3: pretense and shared attention
ST will engage in 3-5 circles of communication surrounding a high preference toy, which include 1-3 play expansions during one Floortime™ session with ____ accuracy.

Goal Area 4: affect recognition (complex)
ST will recognize complex emotions in context (e.g., distress, embarrassment, surprise) of a family member following an event at home with ____ accuracy.

Goal Area 5: mental-physical distinction
ST will identify mental and physical entities with ____ accuracy when presented with a forced-choice closed-ended questions (e.g., "Joan has a real puppy. Mel is thinking of an imaginary puppy. Who can feed the puppy? Who can change the size of the puppy?).

Recommended Intervention Approaches:

- Floortime/DIR
- Joint Attention Training
- Social Stories™
- Teaching Children with Autism to Mind Read

Case Study 2: KF. KF is a 15-year old male diagnosed with autistic disorder. He is verbal and uses language in a variety of ways to communicate with others. He enjoys video games, American history, science and sports. KF's performance on a test of nonverbal intelligence yielded a standard score of 80 with a percentile rank of 9. His standard score on a test of receptive vocabulary was 54 placing him at the first percentile. His caregiver completed the ToMI generating the profile provided below.

COMPOSITE SCORE

Composite Mean = *9.04 Percentile = 1st

EARLY ToM SUBSCALE

Subscale Mean = *12.86 (of 20) Percentile =1st

Subscale Item Scores

Item	designed to assess	score
3.	affect recognition (complex)	4.6
6.	affect recognition (emotion-expression relationship)	16.3
24.	Intentionality	9.4
25.	affect recognition (basic)	16.0
29.	social referencing	14.8
37.	sharing attention – initiating	13.4
38.	sharing attention – responding	15.5

BASIC ToM SUBSCALE

Subscale Mean = *10.64 (of 20) Percentile = 1st

Subscale Item Scores

Item	designed to assess	score
1.	physiologically-based behavior	13.1
4.	emotion-based behavior	11.4
7.	mental state term comprehension	20.0
8.	false beliefs in context of unexpected change of location	3.7
9.	seeing leads to knowing	14.7

10.	mental state term comprehension	18.8
11.	appearance-reality distinction	16.8
12.	false beliefs in context of unexpected contents	16.5
15.	Certainty	3.0
16.	mental-physical distinction	1.9
26.	Pretense	14.7
29.	counterfactual reasoning	13.7
30.	mental-physical distinction	13.5
31.	ability to deceive	1.7
32.	level 1 visual perspective-taking	12.5
33.	speech acts	9.3
35.	speech acts	2.9
39.	mental state term comprehension	12.0
42.	attribute-based behavior	2.0

ADVANCED ToM SUBSCALE

Subscale Mean = *5.15 (of 20) Percentile = 1st

Subscale Item Scores

Item	designed to assess	score
2.	Sarcasm	.1
5.	second-order false desire attribution	6.0
13.	idiomatic language	17.2
14.	use of language to intentionally deceive	2.0
17.	understanding display rules	13.4
18.	complex social judgment	2.2
19.	white lies	2.3
20.	understanding lies versus jokes	10.8
21.	level 2 visual perspective-taking	2.0
22.	second order understanding of belief	2.9
23.	second order understanding of emotion	4.6
27.	complex social judgment	7.0
34.	Empathy	1.5
36.	humor (play on words)	1.2
40.	biased cognition	4.4
41.	mind as active interpreter	10.0

KF shows a pattern of results for subscale scores similar to that of ST (case study 1).

ToMI results for KF suggest greatest competency in Early ToM (subscale mean = 12.86), followed by Basic ToM (subscale mean = 10.64), which was followed by Advanced ToM (subscale mean = 5.15) with all scores being below the 10^{th} percentile in relation to typical age norms. With regard to Early ToM, complex affect recognition was (barely) in the 'latent' range, scores for intentionality, social referencing, and initiating shared attention were in the 'emerging' range, and scores for two aspects of affect recognition and responding to shared attention were in the 'developed' range.

For Basic ToM, mental state comprehension and the appearance-reality distinction appear to be 'developed', false belief understanding and mental physical distinction are 'latent', and the understanding of physiologically-based behavior, seeing leads to knowing, pretense, and counterfactual reasoning are in the 'emerging' range. Several Advanced skills are yet to develop including certainty, speech acts, ability to deceive, emotion-based behavior, and attribute-based behavior. Although Advanced ToM skills are an age-appropriate goal for KF (and some dimensions appear to be 'emerging'), they do not appear to be developmentally appropriate on the basis of his language, cognitive, and ToM scores on the ToMI and, thus, are not targeted at this time.

As with the previous example, intervention could focus on either Early or Early *and* Basic ToM skills. For KF, we believe that it may be useful to ensure consistency in affect recognition and developing intentionality prior to moving to the Basic ToM skills. Despite the relatively low score on complex affect recognition (i.e., 4.6 which is barely below the 'emerging' range), this can be viewed as a foundational skill so there is some rationale for its selection as a target; especially given the fact that we are already operating within the floor subscale and in light of the high scores observed on the other dimensions of affect recognition. Our initial treatment plan focused on these Early skills and were as follows:

Goal Area 1: intentionality
KF will identify his own motivational/intentional states (emotions, cognitions, desires) in real-life social situations with ___ accuracy out of ___ trials.

Goal Area 2: intentionality
KF will identify others' motivation/intentional states of others in real-life social situations with ___ accuracy out of ___ trials.

Goal Area 3: affect recognition (complex)
KF will identify his complex emotions (e.g., upset, guilty, proud, worried) for self and family members following a conflict with ___ accuracy out of ____ trials.

Goal Area 4: affect recognition (complex)
KF will recognize the complex emotions (e.g., surprised, upset, worried, etc.) of family members following a conflict with ___ accuracy out of ____ trials.

Recommended Intervention Approaches:
- Social Stories™ (with a focus on the perspective of self and others)
- Comic strip conversations (considering his stronger verbal ability and his ability to draw)
- Teaching Perspective Taking to Children with Autism (Chapters 3- 5)
- Teaching Children to Mind Read

Case Study 4: JM. JM is an eight-year old male diagnosed with autistic disorder. He is functionally nonverbally and has some non-verbal communicative behaviors (i.e., pointing and a few signs) and can communicate using an alternative augmentative device (i.e., Dynavox). He enjoys trains, snuggling, and Disney movies. JM's performance on a test of nonverbal intelligence yielded a standard score of 101 with a percentile rank of 52. He also obtained a standard score on a test of receptive vocabulary of 54 placing him at the first percentile. His caregiver completed the ToMI generating the profile provided below.

COMPOSITE SCORE

Composite Mean = *10.13 Percentile = 1st

EARLY ToM SUBSCALE

Subscale Mean = *13.30 (of 20) Percentile =1st

Subscale Item Scores

Item	designed to assess	score
3.	affect recognition (complex)	14.6
6.	affect recognition (emotion-expression relationship)	20.0
24.	Intentionality	9.9
25.	affect recognition (basic)	16.4
29.	social referencing	16.4
37.	sharing attention – initiating	15.6
38.	sharing attention – responding	16.6

BASIC ToM SUBSCALE

Subscale Mean = *10.55 (of 20) Percentile = 1st

Subscale Item Scores

Item	designed to assess	score
1.	physiologically-based behavior	16.2
4.	emotion-based behavior	5.3
7.	mental state term comprehension (think)	9.8
8.	false beliefs in context of unexpected change of location	15.9
9.	seeing leads to knowing	16.5
10.	mental state term comprehension (know)	10.0
11.	appearance-reality distinction	13.4
12.	false beliefs in context of unexpected contents	3.3
15.	Certainty (knowing/guessing)	5.8
16.	mental-physical distinction	9.8
26.	Pretense	14.9
29.	counterfactual reasoning	10.7
30.	mental-physical distinction	15.3
31.	ability to deceive	3.6
32.	level 1 visual perspective-taking	14.3
33.	speech acts (promises)	14.0
35.	speech acts (secrets)	5.0
39.	mental state term comprehension (believe)	8.3
42.	attribute-based behavior	8.3

ADVANCED ToM SUBSCALE

Subscale Mean = *7.76 (of 20) Percentile = 1st

Subscale Item Scores

Item	designed to assess	score
2.	Sarcasm	4.5
5.	second-order false desire attribution	8.8
13.	idiomatic language	14.8
14.	use of language to intentionally deceive	2.3
17.	Understanding display rules	13.6
18.	complex social judgment	6.1
19.	white lies	5.2
20.	Understanding lies versus jokes	5.2
21.	level 2 visual perspective-taking	14.2
22.	second order understanding of belief	6.3
23.	second order understanding of emotion	6.9
27.	complex social judgment	15.7
34.	Empathy	5.0
36.	humor (play on words)	6.3
40.	biased cognition	3.0
41.	mind as active interpreter	14.0

According to the ToMI results, JM demonstrates the greatest competency in Early ToM (subscale mean = 13.3), less competency in Basic ToM (subscale mean = 10.55), and the least competency in Advanced ToM (subscale mean = 7.76). Although scores for composite and subscale scores fall below the 10th percentile, examination of item-level scores reveals marked strength in Early ToM with scores for all dimensions in the 'developed' range with the exception of intentionality and complex affect recognition which are in the 'emerging' range. While these targets are appropriate for intervention, several Basic ToM skills appear to be 'emerging' as well (i.e., emotion-based behavior, mental state term comprehension, appearance-reality, certainty, mental physical distinction, counterfactual reasoning, level 1 visual perspective taking, speech acts [promises]). JM has yet to develop most Advanced ToM skills. Therefore, intervention was focused on JM's emerging Early and Basic ToM skills. The following goals were developed and provide an example of Early and Basic ToM goals for a nonverbal child with severe limitations:

Goal Area 1: intentionality
JM will identify the motivational states of family members during daily routines by pointing to pictures with ___ accuracy out of ___ trials.

Goal Area 2: emotion-based behavior
JM will match a character's mental state with a behavior in a literature-based story when presented with at least one situational cue with ___ accuracy.

Goal Area 3: appearance-reality
JM will identify appearances (i.e., what something looks like, e.g., an apple) and realities (i.e., what something is, e.g., a candle) by pointing to pictures when presented with deceptive objects (e.g., a candle that looks like an apple) and simple forced choice questions with ____ accuracy.

Goal Area 4: certainty
JM will identify knowing and guessing cognitive states when presented with short stories by pointing to pictures with ____ accuracy.

Goal Area 5: counterfactual reasoning
JM will identify situational contingencies through pointing to pictures when verbally presented with a set of hypothetical events (i.e., 'if, then' statements) with ____ accuracy.

Recommended Intervention Approaches:
- Social Stories™ (with a focus on the perspective of others)
- Teaching Perspective Taking to Children with Autism (Chapters 3- 4)

Case Study 4: CM. CM is a 16 year-old girl, diagnosed with ASD. She is a verbal communicator and demonstrates the ability to communicate using a variety of non-verbal communicative behaviors, such as gestures, facial expressions, and intentional body language. CM enjoys drawing, bicycling, animals, and vacationing with her family. CM has a concomitant anxiety disorder and struggles with establishing and maintaining friendships. Her caregiver completed the ToMI generating the profile provided below:

COMPOSITE SCORE

Composite Mean = 17.52 Percentile = 17th

EARLY ToM SUBSCALE

Subscale Mean = 19.28 (of 20) Percentile =60th

Subscale Item Scores

Item	designed to assess	score
3.	affect recognition (complex)	18.0
6.	affect recognition (emotion-expression relationship)	20.0
24.	Intentionality	17.1
25.	affect recognition (basic)	20.0
29.	social referencing	20.0
37.	sharing attention – initiating	20.0
38.	sharing attention – responding	20.0

BASIC ToM SUBSCALE

Subscale Mean = 19.32 (of 20) Percentile = 53rd

Subscale Item Scores

Item	designed to assess	score
1.	physiologically-based behavior	20.0
4.	emotion-based behavior	19.2
7.	mental state term comprehension (think)	20.0
8.	false beliefs in context of unexpected change of location	19.0
9.	seeing leads to knowing	19.8
10.	mental state term comprehension (know)	20.0
11.	appearance-reality distinction	20.0
12.	false beliefs in context of unexpected contents	20.0
15.	Certainty (knowing/guessing)	15.5
16.	mental-physical distinction	20.0
26.	Pretense	20.0
29.	counterfactual reasoning	17.3
30.	mental-physical distinction	20.0
31.	ability to deceive	20.0
32.	level 1 visual perspective-taking	20.0
33.	speech acts (promises)	20.0
35.	speech acts (secrets)	20.0
39.	mental state term comprehension (believe)	20.0
42.	attribute-based behavior	18.3

ADVANCED ToM SUBSCALE

Subscale Mean = *14.62 (of 20); Percentile = 8th %

Subscale Item Scores

Item	designed to assess	score
2.	Sarcasm	20.0
5.	second-order false desire attribution	15.4
13.	idiomatic language	12.2
14.	use of language to intentionally deceive	12.0
17.	Understanding display rules	19.0
18.	complex social judgment	10.2
19.	white lies	10.0
20.	Understanding lies versus jokes	15.4
21.	level 2 visual perspective-taking	18.1
22.	second order understanding of belief	16.0
23.	second order understanding of emotion	16.0
27.	complex social judgment	15.0
34.	Empathy	14.6
36.	humor (play on words)	12.0
40.	biased cognition	12.3
41.	mind as active interpreter	18.0

According the subscale percentiles, CM demonstrates age appropriate skills on both the Early and Basic scales. On the Advanced subscale, however, she was below the 10^{th} percentile. Items where she scored between 5 and 15, indicating 'emerging' skill areas included: second-order false desire attribution, idiomatic language, use of intentionality to deceive, white lies, complex social judgment, understanding lies verses jokes, biased cognition. and interpreting humor. In line with the descriptive-developmental approach to intervention planning (see Figure 4), the following Advanced ToM goals were developed to guide intervention planning with the hope of addressing ToM challenges in several ways that would be relevant and meaningful to her day-to-day social functioning:

Goal Area 1: idiomatic language
CM will increase her capacity to interpret the nuances of language by identifying the thoughts and intentions underlying a given idiomatic phrase and/or figure of speech, when provided with a pictorial representation of a character, with ___accuracy out of ___ trials.

Goal Area 2: complex social judgment
CM will identify the thoughts, beliefs, and mental states of persons and whether or not they match or are different from details included in naturalistic and second order and nested false belief scenarios as indicated by ___accuracy across ___trials, with moderate support.

Goal Area 3: complex social judgment
CM will identify behaviors that are expected to change due to variations in relationship and/or context and on the thoughts, feelings, and belief's of others' when provided with a pre-trained processing strategy (e.g., social behavior map, social response pyramid) for __ out of ___ opportunities across one academic day.

Goal Area 4: complex social judgment
CM will identify 3-5 observations or details from a social situation, the social theme or conflict, and identify two "hidden rules" that apply to variations in the context, time, and audience involved in the conflict with ___ accuracy and minimal support.

Goal Area 5: white lies
CM will identify different kinds of lies (i.e., white lies, bold-faced lies, fabrication, and broken promises) on the basis of short stories and hypothetical scenarios with ____ accuracy and minimal support.

Goal Area 6: biased cognition
CM will predict the social attributions (positive, negative, neutral) of characters who have information (positive, negative, neutral) about other characters in a story with ____ accuracy and minimal support.

Recommended Intervention Approaches:

- Comic strip conversations
- Teaching Perspective Taking to Children with Autism (Chapter 6)
- Social Thinking® : Think Social & Social Behavior Mapping
- Navigating the Social World
- The Hidden Curriculum

Case Study 5: JD. JD is a 17 year-old adolescent male, diagnosed with pervasive developmental disorder-not otherwise specified. Although he is a verbal communicator, JD's spontaneous speech is generally characterized by short phrases that are functionally communicative. He also demonstrates the ability to communicate using some basic non-verbal communicative behaviors, such as gestures, some facial expressions and some intentional body language. He enjoys music, the computer, and creating works of art using wind chimes. JD's performance on a test of nonverbal intelligence yielded a standard score of 90 with a percentile rank of 25. His performance on a test of receptive vocabulary was 68 placing him at the 8th percentile. His caregiver completed the ToMI generating the profile provided below.

COMPOSITE SCORE

Composite Mean = *13.46 Percentile = 1st

EARLY ToM SUBSCALE

Subscale Mean = *15.31 (of 20) Percentile =1st

Subscale Item Scores

Item	designed to assess	score
3.	affect recognition (complex)	11.8
6.	affect recognition (emotion-expression relationship)	12.7
24.	Intentionality	13.8
25.	affect recognition (basic)	14.2
29.	social referencing	14.7
37.	sharing attention – initiating	20.0
38.	sharing attention - responding	20.0

BASIC ToM SUBSCALE

Subscale Mean = *13.31 (of 20) Percentile = 1st

Subscale Item Scores

Item	designed to assess	score
1.	physiologically-based behavior	14.9
4.	emotion-based behavior	12.2
7.	mental state term comprehension (think)	12.8
8.	false beliefs in context of unexpected change of location	12.6
9.	seeing leads to knowing	9.4
10.	mental state term comprehension (know)	13.1
11.	appearance-reality distinction	16.3
12.	false beliefs in context of unexpected contents	13.8

15.	Certainty (knowing/guessing)	14.0
16.	mental-physical distinction	15.6
26.	Pretense	14.8
29.	counterfactual reasoning	14.0
30.	mental-physical distinction	15.3
31.	ability to deceive	7.2
32.	level 1 visual perspective-taking	12.7
33.	speech acts (promises)	13.9
35.	speech acts (secrets)	9.3
39.	mental state term comprehension (believe)	14.1
42.	attribute-based behavior	16.9

ADVANCED ToM SUBSCALE

Subscale Mean = *12.07 (of 20)   Percentile = 1st

Subscale Item Scores

Item	designed to assess	score
2.	Sarcasm	15.0
5.	second-order false desire attribution	13.8
13.	idiomatic language	15.0
14.	use of language to intentionally deceive	12.2
17.	understanding display rules	13.9
18.	complex social judgment	14.7
19.	white lies	9.6
20.	understanding lies versus jokes	14.0
21.	level 2 visual perspective-taking	13.7
22.	second order understanding of belief	9.9
23.	second order understanding of emotion	9.5
27.	complex social judgment	6.0
34.	Empathy	11.3
36.	humor (play on words)	15.5
40.	biased cognition	12.3
41.	mind as active interpreter	18.8

Like many children on the spectrum, JD demonstrates skills in the lowest 10th percentile across the Early, Basic, and Advanced ToM subscales. He also shows a common profile in that subscale mean scores are highest for Early ToM (mean = 15.31), followed by Basic ToM (mean

= 13.31), and Advanced ToM (mean = 12.07), although the mean scores across subscales are somewhat atypical in that they are relatively flat (compare with cases 1- 4). Moreover, inspection of item-level raw scores on the Early subscale suggests that all Early skills are 'emerging' except shared attention which is 'developed'. Inspection of Basic ToM items reveal that all skills are 'emerging' with the exception of appearance-reality, mental-physical distinction, and attribute-based behavior which are 'developed'. Similarly, almost all items on the Advanced subscale fall in the 'emerging range'. The exceptions here are the understanding of humor (play on words) and mind-as-active-interpreter which fall in the 'developed' range.

Based on JD's flat (but relatively strong) pattern of performance across subscales, we shifted our focus away from individual subscale performance to item-level performance to identify viable targets of intervention from across Early, Basic, and Advanced skill levels. Several ToM skills that fell in the 'emerging' range and which constitute good candidates were included in the initial intervention plan. Goals 1 – 3 focused on Early skills, Goals 4-5 focused on Basic skills, and Goals 6-7 focused on Advanced skills and were are follows:

Goal Area 1: affect recognition (basic)
JD will identify the basic emotional states (i.e., happy, sad, mad, scared) of characters presented in literature read and real-life scenarios with ___ accuracy out of ___ trials.

Goal Area 2: intentionality
JD will identify behaviors that match a characters motivational/intentional state that are embedded within a social scenario when presented with 1-3 situational cues with ___ accuracy.

Goal Area 3: affect recognition (complex)
JD will identify the complex emotions (e.g., guilty, proud, worried etc.) of a character following an event, conflict or interaction based on his/her desires, beliefs and thoughts that were identified prior to the event with ____ accuracy.

Goal Area 4: emotion-based behavior
JD will identify the feelings of story characters (e.g., scared, excited, worried, surprised) who behave in ways consistent with those feelings with ___ accuracy across ___ trials, with moderate support via prompting.

Goal Area 5: mental state term comprehension
JD will identify a character's thoughts, beliefs and knowledge about a social scenario when presented with a cue that the character has had a sensory experience (e.g., seeing, hearing, feeling etc.) with an object or event portrayed in the scenario, with ___ accuracy.

Goal Area 6: level II visual perspective-taking
JD will identify photographs that correspond to the visual fields of others when seated at various positions in a room filled with objects with ____ accuracy.

Goal Area 7: empathy
JD will identify the emotional states of characters in short narratives that rely heavily on inference of mental states (i.e., use little explicit mental state language) in the context of explicit situational contexts with ____ accuracy.

Recommended Intervention Approaches:
- Comic strip conversations
- Teaching Children with Autism to Mind Read
- Teaching Perspective Taking to Children with Autism, Chapters 3-6
- Building Social Relationships
- Social Thinking® : Social Behavior Mapping

THE THEORY OF MIND TASK BATTERY

Description

There are several research and clinical purposes for which direct assessment of an individual's theory of mind competencies is desirable. For such instances, we have developed The Theory of Mind Task Battery. Since the development of the original measure (described in Hutchins, Prelock, & Chace, 2008), the Theory of Mind Task Battery has undergone revision to enhance the content of the test as well as the stimulus materials used during administration. In its current form, the Theory of Mind Task Battery consists of 15 test questions within nine tasks (tasks A – I). Tasks are presented as short vignettes, which are arranged in ascending difficulty and represent a variety in terms of content and complexity ranging from the ability to identify facial expressions to the ability to infer second-order false beliefs. The tasks are presented in a story-book format. Each page has color illustrations and accompanying text. Memory control questions are included which must be passed in order for credit to be given on the test questions. The Theory of Mind Task Battery was designed to assess the ToM understanding of younger and older children who vary widely in their cognitive and linguistic profiles. The test is appropriate for nonverbal individuals with ASD as respondents can indicate responses either verbally or through pointing. As described more fully below, preliminary norms for each task have been established using a relatively small sample of typically developing children ages 2- 12 and the task battery can be used appropriately for clinical populations for whom chronological age exceeds age 12 years when ceiling effects are not observed.

Administration

The ToM Task Battery should be administered in a comfortable and quiet environment with the administrator seated next to the child. The test booklet is placed flat on a table in between the administrator and the child. The administrator should introduce the test as an activity by stating "I am going to read you some short stories and ask you questions about the story. You can answer with words or you can point to the answer." In our experience, there are some children who will want to read the story along with the administrator. We actually discourage this practice because it has the potential to interfere with performance (e.g., by changing the pragmatics of the interaction or through a shift to a focus on reading). If this occurs, the administrator should pause and inform the child that "for these stories, I will read and I want you to listen carefully and answer a few questions."

There are no opposing pages in the ToM Task Battery booklet. This is necessary so that the child does not look back at the previous pages to figure out the answer to a question. The one exception to this is Task D where there are opposing pages and the child is allowed to look at the previous (opposing) page to answer the test questions. For each page, the administrator reads the text verbatim while pointing to the images that correspond to what is said. The administrator should read and point at a relaxed and relatively slow pace. For example, for the first two pages of Task B, the administrator should do the following:

Administrator:	"This is Brynn" (points to Brynn). (wait 1 second) "Brynn wants a cookie to eat" (points to cookie). (wait 3 seconds, turn page) "What does Brynn want?" If no response from child then use prompt 1 → "Does Brynn want a cake (points to cake), a cookie (points to cookie), a lollipop (points to lollipop), or a candy bar (points to candy bar).

Pointing should also be carried out to make salient the relevant aspects of the stories that involve perception (i.e., what people see or don't see) and knowledge (i.e., what people think). For example, several tasks require that the respondent consider a character's perception or belief. When these aspects are present, they should be emphasized through pointing while reading. To illustrate, consider Task D:

Administrator: "Jasmin is at the park" (points to Jasmin)
"Vince is also at the park" (points to Vince)
"Jasmin and Vince are looking at a statue" (points to statue)
(wait 1 second)
"When Jasmin (points to Jasmin's face) looks at the statue (moves pointing finger from Jasmin's face to statue), what does she see?"
(point to each of the four response choices)

The ToM Task Battery Score Form should be completed during the administration of the tasks; however, scoring should be done as covertly as possible so as not to distract the respondent. Two score forms are available for use: a long score form and a short score form. Both forms note when to skip items and both provide spaces to score responses and write answers to justification questions. However, the long form expands on this by providing brief information about the construct intended to be tapped by each item. It also includes the actual text of the battery as well as the prompts that accompany each item in the event of a non-response. The long form may be desirable when the battery is administered by those who have limited experience with the battery or who are using the battery with individuals who are expected to need a high degree of prompting (i.e., those with the most limited language or cognitive capacities). In these situations, use of the long form should facilitate administration as it avoids the need to reference the manual to find the required prompts.

The ToM Task Battery makes use of a simple stopping rule. The entire task is completed when all *control* questions are answered correctly and there are no stopping rules associated with test questions. Because the order of tasks in ToM Task Battery is arranged in ascending difficulty, testing should be terminated when the child responds incorrectly *to any five control questions* (they need not be in succession). When this number of control questions is failed, performance factors such as attention and motivation and child factors such as limitations in general cognitive functioning, language comprehension, and/or working memory are probably responsible. Thus, performance on further tasks cannot be considered a valid indicator of ToM functioning. Of course, there are situations when the entire ToM Task Battery should be given. Specifically, if there is a need for repeated measurement then termination of a task or the entire battery can result in test practice effects. We also find that many children enjoy the activities that make up the ToM Task Battery and prefer (or at least expect) to read and respond to all the pages in the booklet. In these cases, the battery is completed but the items that follow the stopping criterion are not scored (i.e., have a score of zero).

In situations where the child fails to respond, specific prompts to elicit a response to all test questions are given below. All test questions have 2 levels of prompting except the first four emotion recognition tasks for which the question is simply repeated a maximum of two times. No prompts are given for control questions and a lack of response to a control question constitutes failure for that control item and, by extension, the associated test question. For some purposes, professionals may have a need to probe the quality and nature of the child's reasoning behind a given response. This approach may be desirable when used with older or highly verbal individuals who may perform well or at ceiling levels on tests of ToM but who, nevertheless, evidence profound difficulties in social cognition. For these purposes, we have included justification questions that may only follow a *correct* response to a test question (and they are not used for control questions). When justification questions are used, the child's response should be written verbatim on the ToM Task Battery Score form.

Prompts and Justification Questions:

Test Question 5: Desire-based emotion question: *<u>How will Brynn feel if she gets a cookie</u>?*
Prompt 1: *Point to the face that shows how Brynn will feel if she gets a cookie.* Prompt 2: *If Brynn gets a cookie, will she feel happy, sad, mad or scared*? Justification (verbal children with correct answer only): *Why will Brynn be happy?*

Test Question 6: Perception-based belief question: *__Where does Patty think her glasses are__?* Prompt 1: *Point to where Patty thinks her glasses are.* Prompt 2: *Does she think they are in the drawer, on the desk, on the table, or on the chair?* Justification (verbal children with correct answer only): *Why will Patty think they are on the table?*

Test Question 7: Other-perception question: *__When Jasmine Looks at the statue, what does she see?__* Prompt 1: *Point to the picture that shows what she sees.* Prompt 2: *When Jasmine looks at the statue, does she see this, this, this, or this (pointing to each image)*

Test Question 8: Other-perception question: *__When Vince Looks at the statue, what does she see?__* Prompt 1: *Point to the picture that shows what she sees.* Prompt 2: *When Vince looks at the statue, does he see this, this, this, or this (pointing to each image)*

Test Question 9: Perception-based action question: *__Where will Franklin go to get the keys?__* Prompt 1: *Point to where Franklin will go to get the keys*. Prompt 2: *Will Franklin go to the couch, the desk, the drawer, or the bed?* Justification (verbal children with correct answer only): *Why will Franklin to go the couch**?***

Test Question 10: Standard false belief question: *__Where will Anthony look for the book first__*? Prompt 1: *Point to where Anthony will look for the book first.* Prompt 2: *Will he look in the drawer, on the desk, on the table, or on the chair?* Justification (verbal children with correct answer only): *Why will Anthony look for the book on the table?*

Test Question 11: Belief-based emotion question: *__If Lee thinks his Dad got him an airplane, how will Lee feel__*? Prompt 1: *Point to how Lee will feel if he thinks that his Dad got him an airplane.* Prompt 2: *If Lee thinks his Dad got him an airplane, will Lee feel happy, sad, mad or scared?* Justification (verbal children with correct answer only): *Why will Lee feel happy?*

Test Question 12: Reality-based emotion question: *__How will Lee feel when his Dad gives him the train__*? Prompt 1: *Point to the picture that shows how Lee will feel when his Dad gives him the train.* Prompt 2: *When Dad gives Lee the train, will Lee feel happy, sad, mad or scared?* Justification (verbal children with correct answer only): *Why will Lee feel sad?*

Test Question 13: Second-order emotion question: *__When Dad gives Lee the train, how does Dad think Lee will feel__*? Prompt 1: *Point to the picture that shows how Dad thinks Lee will feel.* Prompt 2: *When Dad gives Lee the train, does Dad think Lee will feel happy, sad, mad or scared?* Justification (verbal children with correct answer only): *Why does dad think Lee will be happy?*

Test Question 14: Message-desire discrepant question: *__Which bowl does Russ really want__*? Prompt 1: *Point to the bowl that Russ really wants.* Prompt 2: *Does he want a bowl of salad, a bowl of spaghetti, a bowl of bread, or a bowl of soup?* Justification (verbal children with correct answer only): If Russ wanted the bowl on the table, why did he ask for the bowl on the counter?

Test Question 15: Second-order false belief question: *<u>What does Mom tell Grandfather? Does she tell him that Enrique thinks he is getting roller blades, a bike, a basketball, or a baseball glove</u>?* Prompt 1: *Point to the picture that shows what Mom will say Enrique thinks he is getting.* Prompt 2: *Will Mom say that Enrique thinks he is getting roller blades, a bike, a basketball, or a baseball glove?* Justification (verbal children with correct answer only): *Why will Mom say Enrique thinks he is getting roller blades?*

Development and Content

Most of the content of the ToM Task Battery was gathered from the vast body of literature examining ToM impairments in ASD as well as typical development of ToM. This body of literature included a diverse set of previously developed measures from which ideas were gathered and content was adapted to accommodate a story-book format that could make use of static visual supports. Content was selected to create a battery that spanned a range of ToM content and that had items that varied in complexity and difficulty. For all tasks in the ToM Task Battery, children are presented with one correct response option and three plausible distracters, making the chance of correct responding in the absence of ToM knowledge equal to 25%. This is true for both test and control questions. During development, care was taken to ensure that the language used in the tasks is informal and easily understood. We also developed content to suit a diverse American population. Thus, several different races and ethnicities are represented in the story content.

The domain that is intended to be tapped by each task is briefly described below:

TASK A: The *<u>Emotion Recognition Task</u>* is intended to assess children's recognition of emotional states. Specifically, children are asked to identify a happy, sad, mad, and scared face. Two panels consisting of four illustrations each (two correct and two distracters) were presented in order to reduce response bias due to a process of elimination. Four points (one for each emotion) are possible for this task.

TASK B: The *<u>Desire-Based Emotion Task</u>* was developed from several research paradigms (e.g., Wellman, 1988; Wellman & Banerjee, 1991; Wellman & Bartsch, 1994) and it is intended to assess children's understanding of desires. More specifically, this task is designed to tap the understanding that people are happy when desires are satisfied. One point is possible for this task.

TASK C: The *Seeing Leads to Knowing Task* was developed from several research and experimental paradigms (e.g., Baron-Cohen & Goodhart, 1994; Friedman, Griffin, Brownell, & Winner, 2003; Leslie & Frith, 1988) and it is intended to assess children's knowledge that perceptions influence beliefs. The specific content of this understanding is the notion that seeing something (and more generally hearing about something) provides access to knowledge. Children who acquire this understanding should be able to attribute knowledge or ignorance to an observer on the basis of whether the observer was able to access information via seeing (or hearing). One point is possible for this task.

TASK D*:* The *Line of Sight Task* (Flavell, 1992) is intended to assess the understanding that people may not see the same thing depending on positioning. A total of two points (one for each characters' perspective) is possible for this task.

TASK E: The *Perception-Based Action Task* (Hadwin, Baron-Cohen, Howlin, & Hill,1996) was adapted to assess the understanding that perceptions influence behavior. Thus, this task has one additional layer of understanding compared to the Seeing Leads to Knowing Task. For Perception-Based Action, the child must understand that 1) knowledge can be gained through visual perception (e.g., seeing keys on a couch leads to knowledge that keys are on the couch) and, 2) that knowledge drives behavior (e.g., knowing the keys are on the couch will result in seeking behavior such that the person will now look for the keys on the couch). One point is possible for this task.

TASK F: A *Standard False Belief Task* (Wimmer & Perner, 1983) is intended to assess children's ability to infer belief in the context of an unexpected location change. Following the recommendation of Siegal and Beattie (1991), the test question for the item modeled after the classic false belief task was modified to include the word *first* (i.e., "Where will Anthony look for the book first?") to limit the potential that this question would be misinterpreted. That is, this question should not be interpreted as "Where will someone need to look in order to be successful in finding the object?" Like the Perception-Based Action Task, this task also includes an understanding of the knowing-looking connection; however, the Standard False Belief Task adds yet another layer of complexity because it must also include the understanding that people can have a belief that contradicts reality. One point is possible for this task.

TASK G: The *Belief- and Reality-Based Emotion and Second Order Emotion Task* was adapted (Hadwin et al., 1996) to assess the understanding that beliefs, as well as events contrary to beliefs, can cause emotion. This task also incorporated a second-order emotion task to assess children's understanding that an observer will incorrectly infer a protagonist's emotion based on a false belief about the protagonist's desire. This adds another degree of complexity and requires recursive thinking (i.e., thinking about what someone thinks about someone else's emotions/desires). A total of three points is possible for this task.

TASK H: The *Message-Desire Discrepant Task* was adapted (Mitchell et al., 1997) to assess the ability to infer the belief of another when interpreting a statement of desire in the context of a change location (i.e., false belief). This task was chosen because it represents a distinct facet of ToM while conferring advantages over other tasks (e.g., the more traditional Smarties, false-contents task) by avoiding response errors due to an overly literal interpretation of the test question (see Fodor,1992; Mitchell et al.,1997). A total of one point is possible for this task.

TASK I: A *Second-Order False Belief Task* was adapted (Silliman et al., 2003; originally adapted from Sullivan, Zaitchik, & Tager-Flusberg, 1994) to tap knowledge of second-order false beliefs. This task is believed to be the most challenging test of ToM in the battery. Not only is complex recursion involved (thinking about what someone else thinks about what someone else thinks) but it also includes the element of a false belief. As described more fully below, it is not uncommon for some older typically developing children to fail Tasks G and H but pass Task I. We suggest that, with regard to ToM knowledge, Task I may be the most challenging but that patterns of performance like that just noted are possible due to the item construction of G and H and their larger number of associated control questions for which a pass is required for credit on the test questions. One point is possible for this task.

Scoring

Test Questions. In line with the scoring procedures of previously developed ToM tasks, children do not receive credit for items when the associated control question is not passed. The total score is simply the number of correct responses to test questions. This number is determined by looking at the score form and it is noted at the bottom of the form. Possible scores range from 0 – 15 with higher values indicated greater ToM knowledge. It is important to note, however, that this is construed as an ordinal scale. Because the test advances in difficulty and latter items arguably reflect greater or more developed ToM knowledge, inspection of raw ordinal scores in isolation may obscure ToM knowledge when early items are failed and latter items are passed. Fortunately though this is rarely the case and to reiterate, the battery is organized to proceed from relatively simple and early-emerging aspects of ToM to more complex, sophisticated, and latter emerging aspects of ToM.

Optional Justification Questions. Optional justification questions are available for nine of the 15 test questions (see below). These nine questions were chosen because, from a pragmatic standpoint, they lend themselves to justification. In addition, justification answers to these nine questions can be succinctly articulated when mastery of the ToM area being tested is present. When employing justification questions, a variety of scoring options are available. One option is to conduct purely qualitative analyses and to look for patterns in cognition or communication that might reflect gaps in ToM knowledge and suggest areas of clinical focus. Another option is to score according to an ordinal scale where values reflect variation in the completeness or correctness of a response. The coding scheme chosen will be guided by the specific needs of professionals working with children with ASD and other developmental disabilities. Nonetheless, we offer some suggestions for such an ordinal scale as follows:

1 = no response; "I don't know"; or otherwise clearly incorrect response

2 = incomplete response where child does not reference inner mental states (e.g., wanting, thinking, knowing) OR partially incorrect response where child references inner mental states but the wrong mental states are invoked or important details are missing and the links from mental states to real-world consequences are vague or absent.

3 = a complete correct response

Test Question 5 Justification question: *Why will Brynn be happy?*

1 = I don't know
2 = because she likes cookies (incorrect mental state)
3 = because she got the cookie [as to mean, she got what she wanted], because she got what she wanted, because she wanted the cookie

Test Question 6 Justification question: *Why will Patty think they are on the table?*

1 = because they are on the table (absence of mental state reference)
2 = because she thinks about those kinds of things (reference to mental state but not related to perception-based knowledge and the idea that seeing leads to knowing)
3 = because that is where she saw them, she saw them there last

Test Question 9 Justification question: *Why will Franklin go to the couch***?**

1 = because they are on the couch (absence of mental state)
2 = because he didn't see the ones on the bed (reference to mental state but unresponsive to question; it is unclear whether the child understands the links between seeing, knowing, and behaving)

3 = because that is where he saw them, because he knows they are there, because he knows they are there because that is where he saw them

Test Question 10 Justification question: *Why will Anthony look for the book on the table?*

1 = because that is where people put books (absence of mental state reference; incorrect response)
2 = He is going to look for it there but it really isn't there (unresponsive to question; also reference to mental state but does not reflect clear knowledge of false beliefs)
3 = because that is where he left it, he saw the book on the table last, because that is where he thinks it is

Test Question 11 Justification question: *Why will Lee feel happy?*

1 = because it is his birthday (incorrect response)
2 = because he thinks it will be fun to play with the airplane (reference to mental state but no stated connection to wants and desires)
3 = because he thinks he is getting what he wants, because that is what he wanted for his birthday

Test Question 12 Justification question: *Why will Lee feel sad?*

1 = because he wanted the airplane now and not later (reference to mental state but incorrect response)
2 = because his birthday is not fun now (reference to mental state but no stated connection to wants and desires)
3 = because he didn't get what he wanted, because he thought he was getting a plane but he didn't, he is disappointed because he did get what he wanted the most

Test Question 13 Justification question: *Why does dad think Lee will be happy?*

1 = because dad likes trains (reference to mental state but incorrect response)
2 = because kids like presents (reference to mental state but incomplete response), because dad gave him the train (no reference to mental state but partially correct response)
3 = because dad thinks he wants a train, because dad got a train thinking that's what Lee wanted

Test Question 14 Justification question: *If Russ wanted the bowl on the table, why did he ask for the bowl on the counter?*

1 = because Russ is hungry (reference to physiological state but incorrect response)
2 = because Russ wants to eat spaghetti (reference to mental state but incomplete response),
3 = because Russ didn't know that she switched the bowls, because the bowls got switched and he thinks the one he wants is on the counter

Test Question 15 Justification question: *Why will Mom say Enrique thinks he is getting roller blades?*

1 = because Enrique wants roller blades for his birthday (reference to mental state but incorrect response)
2 = because that is what she told him (no reference to mental states but partially correct response)
3 = because that is what she thinks he thinks, because she tried to trick him into thinking he was getting roller blades, because she told him he was getting roller blades and she doesn't know he found the bike

Reliability and Validity

The following analyses were conducted using two different samples. These samples and their data sources are described below.

ASD sample. Subjects were 41 children (8 girls, 33 boys) ranging in age from 4 to 12 years (*M* =7.3) and diagnosed with Autism (n = 18), Pervasive Developmental Disorder–Not Otherwise Specified (PDD-NOS, n = 17) or Asperger's Disorder (n = 6) using the criteria in the *Diagnostic and Statistical Manual of Mental Disorders–Fourth Edition* (DSM-IV; American Psychiatric Association, 1994). Subjects were participants in a two studies designed to assess the effects of social-pragmatic interventions for addressing the core deficits of ASD. These studies were longitudinal, however, all data reported here were taken at pre-intervention and reflect baseline functioning. Diagnoses were made between 22 months and 8 years of age by a developmental pediatrician, pediatric psychiatrist, or psychologist with experience in the diagnosis of children with autism. The *Autism Diagnostic Observation Schedule* (ADOS; Lord, Rutter, DiLavore, & Risi, 1999) was used to confirm the participants' diagnoses of ASD. Children represented a range of verbal abilities assessed on the basis of case history and the *Peabody Picture Vocabulary Test–Third Edition* (PPVT-III; Dunn & Dunn, 1997).

Typically development sample. Subjects were 55 typically developing children (25 girls, 30 boys) ranging in age from 2 – 12 years (M = 5.6). Subjects were participants in two separate cross-sectional study. One study was a validation study of the ToMI and the other was an eye-tracking study where typically developing children served as a comparison group to a group of children diagnosed with ASD. Typically developing children were identified on the basis of parent report and parents' responses to a questionnaire that was designed to screen for a variety of clinical conditions (e.g., uncorrected visual or hearing impairment, language or psychiatric disorders). Only parents who reported the absence of any condition and the absence of any parental concern for any condition were included.

Reliability

Internal consistency. Internal consistency of the original version was examined using Cronbach's alpha, which is a measure of homogeneity of content (McCauley, 2001). When alpha estimates are high, it is generally assumed that items on a measure tap a unitary construct. According to conventional guidelines, an alpha of .70 was considered "adequate," an alpha of .80 was considered "good," and an alpha of .90 was considered "excellent." Analysis revealed that internal consistency of the battery achieved $\alpha = .91$ at T1and $\alpha = .94$ at T2, representing excellent intertask agreement.

Test-retest reliability by item. In the original version of the measure, all but three items demonstrated acceptable levels of test-retest reliability (reported in Hutchins, Prelock, & Chace, 2008). These items were subsequently omitted or revised. For the one item that was significantly revised (i.e., Task D), it has yet to be formally evaluated for test-retest reliability and so its test-retest performance is currently unknown.

Test-retest reliability and interval length. To evaluate the effect of length of test-retest interval, comparisons were conducted between shorter (2–7 weeks) and longer intervals (8–16 weeks) in terms of change in score. An independent samples *t* test revealed no effect. Consistent with this result, a Pearson's *r* revealed no relation between change in score from T1 to T2 and the length of the interval between administrations. Thus, analyses of both differences and correlations converged to find that variation in interval did not significantly affect test-retest reliability.

Test-retest reliability and Verbal Mental Age (VMA). To explore whether reliability varied as a function of VMA, (assessed by the PPVT-III) participants were divided into two groups according to their change in score between T1 and T2. "Consistent performers" (change of 0–1 points) and "inconsistent performers" (change of 2 or more points) were compared in terms of VMA. An independent samples *t* test revealed no effect.

To explore whether consistency of performance varied as a function of diagnosis, the change in score from T1 to T2 was compared among children diagnosed with autism ($n = 8$) and PDD-NOS ($n = 7$). Because only two children in the current sample were diagnosed with Asperger's Disorder, data for these cases were dropped from this analysis. An independent samples *t* test revealed no effect of diagnosis on reliability. Incidentally, the mean for the two children with Asperger's Disorder was nearly identical to the means of the other two groups. Reliability and VMA were also examined by the use of a Pearson's *r*, which included the full sample of participants. No relationship was found between VMA and change in score between T1 and T2. In summary, analyses of differences and correlations converged to find that variation in VMA was not associated with variation in reliability.

Validity

Criterion-related validity. A construct valid measure of ToM measured via child performance should be positively correlated with scores on the ToMI informant measure. A Spearman's rho (because ToM task battery data are best construed as ordinal in nature) for the

ASD sample described in the ToMI validation section of this manual indicated a substantial positive relationship (*r* = .66, *p* < .05) with variation in scores on the ToMI explaining approximately 44 percent of the variation in children's scores on the ToM Task Battery. Even stronger correlations are evidenced for the aforementioned typically developing sample (*r* = .82, *p* < .05) indicating approximately 67% shared variance.

A construct valid measure of ToM should also correlate with age. Total ToM Task Battery scores were correlated with child age using the typical sample descrbied above. Results indicated a strong positive relationship between the two measures (r = .66, p < .01) with age explaining approximately 44% of the variance in task battery scores.

Descriptive and Normative Data

Item difficulty. The difficulty of each item was estimated for the ASD sample using the original version of the ToM Task Battery. The original version was evaluated for test-retest reliability and internal consistency (Hutchins, Prelock, & Chace, 2008) and then revised based on improvement in item reliability. The text of the original version and the present version are identical except for three items, which were revised or omitted. Thus, difficulty estimates for the task that was substantively reworked (i.e., Task D) are not available for the ASD sample.

The difficulty of each item was estimated for the typically developing sample using original and revised versions of the battery. For this analysis it was necessary to create an age-matched typically developing sample. Thus, a subsample of 39 typically developing children ages 4 – 12 (M = 7.8) was used for the following analyses.

The difficulty index of each item is presented in Table 5. These data confirm that the tasks are ordered in a *general* increasing order of difficulty.

Table 5: Difficulty index of all tasks for ASD and typically developing samples.

Task and Test Question	Item Difficulty ASD Sample (% who answered item correctly)	Item Difficulty: Typically Developing age-matched Sample (% who answered item correctly)
Task A		
Test Question 1	**86%**	**100%**
Test Question 2	**86%**	**97%**
Test Question 3	**89%**	**97%**
Test Question 4	**92%**	**100%**
Task B		
Test Question 5	**81%**	**97%**
Task C		
Test Question 6	**74%**	**93%**
Task D		
Test Question 7	**--**	**89%**
Test Question 8	**--**	**82%**
Task E		
Test Question 9	**59%**	**74%**
Task F		
Test Question 10	**57%**	**82%**

Task G		
Test Question 11	**56%**	**75%**
Test Question 12	**39%**	**54%**
Test Question 13	**34%**	**52%**
Task H		
Test Question 14	**44%**	**54%**
Task I		
Test Question 15	**17%**	**43%**

Descriptive Statistics: Mean Total Score by Age.

The mean total score by age was calculated using data from the typically developing sample. These data can be used as a general benchmark of expected performance for total score, however, it is again noted that these data are based on a small sample and this is particularly true for certain age categories (e.g., age 2 – 3 years) which much be interpreted with great caution. These data are presented in Table 6.

Table 6: The mean total score by age was calculated using data from the typically developing sample.

Age (n)	**Mean Total Score (Standard Deviation)**
2-3 (n = 4)	**3.5 (.71)**
3-4 (n = 6)	**8.5 (3.5)**
4-5 (n =8)	**11.8 (3.3)**
5-6 (n = 9)	**11.5 (2.1)**
6-7 (n= 7)	**12.4 (1.1)**
7-8 (n= 5)	**12.8 (3.0)**
8-9 (n = 5)	**12.4 (2.1)**
9-10 (n = 7)	**13.4 (1.5)**
11-12 (n = 4)	**15 (0.0)**

Preliminary Norms

Given the usefulness of normative data for research purposes and clinical decision-making, we estimated the age at which most typically developing children passed each test question in the ToM Task Battery. These data are presented in Table 6. These norms were derived using data from Sample B (*n* = 55), which constitutes a small sample and, as such, the norms reported here must be approached with caution. Larger samples are needed to ensure the stability of the norms and this is a focus of our future research. Nevertheless, the norms reported here are generally consistent with the findings in the literature for typically developing children. The norms were determined through examination of pass and fail rates by age. Specifically, the age above which >80% (i.e., 80 – 100%) of the sample performed correctly was identified for each item.

Table 7: Age at or above which 80% (i.e., 80 – 100%) of the sample performed correctly was identified for each item

Task and Test Question	**Preliminary Age Norms**
Task A	
Test Question 1	**2 years; 0 months**
Test Question 2	**2 years; 3 months**
Test Question 3	**2 years; 4 months**
Test Question 4	**2 years; 6 months**
Task B	
Test Question 5	**2 years; 6 months**
Task C	
Test Question 6	**3 years; 5 months**
Task D	
Test Question 7	**4 years; 6 months**
Test Question 8	**4 years; 8 months**
Task E	
Test Question 9	**4 years; 6 months**
Task F	
Test Question 10	**4 years; 6 months**
Task G	
Test Question 11	**5 years; 5 months**
Test Question 12	**--**

Test Question 13	**--**
Task H	
Test Question 14	**--**
Task I	
Test Question 15	**8 years; 2 months**

As shown in Table 7, this criterion was not reached for some of the most difficult items in the task battery (i.e., 12, 13, 14). Upon inspection of the items, this was not surprising. These items appear to be the most difficult items in terms of working memory demands and they employ relatively complex language. On the other hand, test question 15 is similar in this respect and is actually considered the most difficult item with regard to ToM processing; however, preliminary norms could be established for this item. What might account for this pattern in the data? Crucially, items 12, 13, and 14 use a greater number of memory control questions. Consequently, there are simply more opportunities to fail them. It is probably for this reason that we observed failure on these items even among some older typically developing children (e.g., 8, 9, 10 years), which made the establishment of norms tenuous. Closer scrutiny of the data revealed that when these items were failed, it was due to failure on a memory control item as opposed to perfect performance on control items and failure on the test questions. Interpretation of performance patterns on all items, including these for which norms could not be established, is taken up later in this chapter.

Interpretation and Clinical Decision-Making

As clinicians interpret the results of the ToM Task Battery, it is important to recognize that other factors beyond ToM understanding may be influencing a child's performance. That is, a child may not fail an item because of a ToM deficit but instead might be challenged by the linguistic and cognitive complexities of the tool. It is important therefore, that clinicians interpret the results with caution, particularly for items where the child does not pass the comprehension probes or control items. In this situation, it is common to attribute failure on the task to failures of motivation, attention, executive dysfunction, or linguistic complexity but may be attributed to a combination of these or other factors such as a misunderstanding of the pragmatics of the test-taking situation. In any event, no information regarding ToM functioning can be gleaned when control items are failed. When control items are passed and associated test items are failed, this is typically taken as evidence for ToM impairment and it is recommended that interventions begin at the earliest level from a descriptive-developmental perspective (i.e., look to the first item failed and focus intervention there).

When using scores on the ToM Task Battery, the clinician may consider which tasks reflect varying levels of ToM competency as they correspond to the three Early, Basic, and Advanced subtests on the ToMI. As previously described, items on the ToM Task Battery are presented in ascending order of difficulty and tasks correspond to the Early, Basic, and Advanced levels measured by the ToMI as follows: Tasks A and B represent Early ToM skills; tasks C – F represent Basic ToM skills, and tasks G – I represent Advanced ToM Skills.

Interpretation of performance can be strengthened if the clinician uses the results of the ToM Task Battery in conjunction with the ToMI (for a sample write up of ToM Task Battery and ToMI results, see Appendix). In this case, the clinician can look for patterns in performance (at the general or item-level of analysis when there are indeed corresponding items) on both tools, particularly for Task Battery items where the control item is failed and thusly no conclusions can be drawn as to the cause of item failure on the basis of the Task Battery alone. Our clinical impressions involving the joint use of the measures is that scores for general and item-level areas sometimes converge (and indeed the two measures correlate at around .6 - .7 depending on the specific samples used) to provide a coherent portrait of ToM functioning. Other times, scores diverge which complicates interpretation of the tests. For example, if a score demonstrates understanding of a ToM skill (e.g., seeing-leads-to knowing) on the ToM Task Battery (Task C) but not on the ToMI (item 9), it may suggest that the child is using a compensatory strategy to hack a battery item and has not acquired the competence to apply a mentalistic understanding to solve ToM problems in real-life. It could also mean, however, that that ToM knowledge is present when applied in the Task Battery context but that understanding is not applied or transferred to real-world behavior. In this case, the child may not understand the full range of situations for which ToM understanding is relevant. It may also signal failure to detect or define ToM problems in real-world situations. Conversely, an item on the ToMI could yield high scores when the corresponding item is failed on the ToM Task Battery. This may indicate that an ecologically valid competency is present but cannot be expressed on a test of performance where linguistic and other task demands are too high. Of course, divergent results can also be the result of error in measurement. Whatever the case, clinicians are encouraged to carefully consider potential sources of disagreement in scores and it is possible (for the reasons stated above) that disagreement in scores highlights an area of ToM that may actually represent a developmentally appropriate intervention target and a potential area for growth as there is indication that some facet of ToM understanding is in place. In conclusion, when either testing option is available and/or the source of inconsistent scores is unclear, our preference and recommendation (and the rationale underpinning the ToMI to begin with) is to rely on ToMI scores: it is a relatively comprehensive and detailed ToM assessment, it is specifically designed to avoid the problems associated with direct tests of performance, and it is very well-validated.

REFERENCES

Abbeduto, L., & Rosenbberg, S. (1984). Children's knowledge of the presuppositions of know and other cognitive terms. *Journal of Child Language, 12*(3), 621-641.

Abbeduto, L., Short-Meyerson, K., Benson, G., & Dolish, J. (2004). Relationship between theory of mind and language ability in children and adolescents with intellectual disability. *Journal of Intellectual Disability Research, 48*(2), 150-159.

Adams, L., Gouvousis, A., VanLue, M., & Waldron, C. (2004). Social story intervention: Improving communication skills in a child with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 19* (2), 87-94.

Aldred, C., Green, J., & Adams, C. (2004). A new social communication intervention for children with autism: Pilot randomized controlled treatment study suggesting effectiveness. *Journal of Child Psychology and Psychiatry, 40,* 1-11.

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

Apple, A. L., Billingsley, E., & Schwartz, I. S. (2005). Effects of video modeling alone and with self-management on compliment-giving behaviors of children with high functioning ASD. *Journal of Positive Behavior Interventions, 7(*1), 33-46.

Ashcroft, A., Jervis, N., & Roberts, C. (1999). A theory of mind (TOM) and people with learning disabilities: The effects of a training package. *Journal of Applied Research in Intellectual Disabilities*, *12*(1), 58-68.

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

Astington, J. W. (1991). Intention in the child's theory of mind. In D. Frye & C. Moore (Eds.), *Children's theories of mind: Mental states and social understanding* (pp. 157-172). Hillsdale, NJ: Erlbaum.

Astington, J. W. (2003). Sometimes necessary, never sufficient: False belief understanding and social competence. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 13-38). New York: Psychology Press.

Astington, J. W., & Baird, J. A. (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.

Baron-Cohen, S. (1989). Are autistic children 'behaviorists'? An examination of their mental-physical and appearance-reality distinctions. *Journal of Autism and Developmental Disorders, 19*(4), 579-600.

Baron-Cohen, S. (1995). *Mindblindness*. London: MIT Press.

Baron-Cohen, S., & Goodhart, F. (1994). The 'seeing leads to knowing' deficit in autism: The Pratt and Bryant probe. *British Journal of Developmental Psychology, 12*, 397-402.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a theory of mind? *Cognition*, *21*, 37–46.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The 'Reading the Mind in the Eyes' test revised version: A study of normal adults, and adults with Asperger Syndrome or High Functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241-251.

Bates, E. (1979). *The emergence of symbols: Cognition and communication in infancy*. New York: Academic Press.

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.

Bellini, S. (2006). *Building Social Relationships: A Systematic Approch to Teaching Social Interaction Skills to Children with Autism Spectrum Disorders and Other Social Difficulties.* Shawnee Mission, KS: Autism Asperger Publishing Co.

Bellini, S., Peters, J., Benner, L., & Hopf, A. (2007). A metaanalysis of school-based social skills interventions for children with autism spectrum disorders. *Journal of Remedial and Special Education 28*(3), 153-162.

Bernad-Ripoll, S. (2007), Using a self-as-model video combined with Social Stories™ to help a child with Asperger syndrome understand emotions. *Focus on Autism and Other Developmental Disabilities, 22* (2), 100-106.

Brennan, L.C. (2011). Teaching perspective taking to children with autism spectrum disorders. Austin, TX: PRO-ED

Brownell, M. D. (2002). Musically adapted social stories to modify behaviors in students with autism: Four case studies. *Journal of Music Therapy*, *39*(2), 117-144.

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

Buitelaar, J. K., Ven der wees, M., Swaab-barneveld, H., & Jan Van der gaag, R. (1999). Theory of mind and emotion-recognition functioning in autistic spectrum disorders and psychiatric control and normal children. *Development and Psychopathology, 11*, 39-58.

Carpendale, J. M., & Lewis, C. (2006). *How children develop social understanding.* Malden, MA: Blackwell.

Carpendale, J. M., & Chandler, M. J. (1996). On the distinction between false belief understanding and subscribing to an interpretive theory of mind. *Child Development, 67,* 1686-1706.

Carter, A., Messinger, D., Stone, W., Celimli, S., Nahmias, A., & Yoder, P. (2011). A randomized controlled trial of Hanen's More Than Words in toddler with autism. *Journal of Child Psychology and Psychiatry, 52* (4).

Charman, T., Ruffman, T., & Clements, W. (2002). Is there a gender difference in false belief development? *Social Development, 11*(1), 1-10.

Cornish, K., Burack, J. A., Rahman, A., Munir, F., Russo, N., & Grant, C. (2005). Theory of mind deficits in children with Fragile X Syndrome. *Journal of Intellectual Disability Research, 49*(5), 372-378.

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9.

Crozier, S., & Tincani, M. (2007). Effects of social stories on prosocial behavior of preschool children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 37* (9), 1803-1814.

D'Ateno, P., Mangiapanello, K., & Taylor, B. (2003). Using video modeling to teach complex play sequences to a preschooler with autism. *Journal of Positive Behavior Interventions 5*(10), 5-11.

Davies, M., & Stone, T. (2003). Synthesis: Psychological understanding of social skills. In B Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development* (pp. 305-352). New York: Psychology Press.

DeGangi, G. A. & Greenspan, S. I. (1997). The effectiveness of short-term interventions in treatment of inattention and irritability in toddlers. *The Journal of Developmental and Learning Disorders, 1*, 277-298.

Delano, M., & Snell, M. E. (2006). The effects of social stories on the social engagement of children with autism. *Journal of Positive Behavior Interventions, 8, 29*-42.

Dionne, M. & Martini, R. (2011). Floortime™ play with a child with autism. A single-subject study. *Canadian Journal of Occupational Therapy, 78*, 196-203.

Dodd, S., Stephen, D., Hupp, A., Jewell, J. D., & Krohn, E. (2008). Using parents and siblings during a social story intervention for two children diagnosed with PDD-NOS. *Journal of Developmental and Physical Disabilities, 20*, 217-229.

Doherty, M. J. (2009). *Theory of mind: How children understand others' thoughts and feelings.* New York: Psychology Press.

Drew, A., Baird, G., Baron-Cohen, S., Cox, A., Sionims, V., Wheelwright, S., Swettenham, J., Berry, B. & Charman, T. (2002). A pilot randomised control trial of a parent training intervention for pre-school children with autism: Preliminary findings and methodological challenges. *European Child & Adolescent Psychiatry, 11*, 266-272.

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., Marchman, V., Thal, D., Dale, P. Reznich, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and manual*. Baltimore: MD; Paul H. Brookes.

Flavell, J. H. (1992). Perspectives on perspective taking. In H. Beilin & P. B. Pufall (Eds.), *Piaget's theory: Prospects and possibilities* (pp. 107-139). Hillsdale, NJ: Lawrence Erlbaum.

Flavell, J. H., Miller, P. H., & Miller, S. A. (2002). *Cognitive development* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Fodor, J. (1992). A theory of the child's theory of mind. *Cognition, 44*, 281-296.

Foran, J., Hutchins, T. L., Prelock, P. A., & Murray-Close, D. (in preparation). Examining the benefit of including a typically developing sibling in story-based interventions with a child with Asperger Syndrome.

Friedman, O., Griffin, R., Brownell, H., & Winner, E. (2003). Problems with the seeing = knowing rule. *Developmental Science, 6*(5), 505-513.

Frith, C., & Frith, U. (2000). The physiological basis of theory of mind: Functional neuroimaging studies. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental neuroscience* (pp. 334–356). Oxford, UK: Oxford University Press.

Gerber, S. (2012). An introduction to the DIR Model (Developmental Individual-Difference Relationship Based) and its application to children with autism spectrum disorders. In P.A. Prelock and R.J. McCauley (Eds.), *Treatment of Autism Spectrum Disorders: Evidence-based intervention strategies for communication and social interaction.* Baltimore, MD: Paul H. Brookes.

Girolametto, L., Sussman, F., & Weitzman, E. (2007). Using case study methods to investigate the effects of interactive intervention for children with Autism Spectrum Disorders. *Journal of Communication Disorders, 40* (6). 470-492.

Glaeser, B. C., Pierson, M. R., & Fritschmann, N. (2003). Comic strip conversations: A positive behavioral support strategy. *Teaching Exceptional Children, 36 (2), 14-19.*

Gonzalez-Gadea, M. L., Baez, S., Torralva, T., Castellanos, F. X., Rattazzi, A, Bein, V., et al. (2013). Cognitive variability in adults with ADHD and AS: Disentangling the roles of executive functions and social cognition. *Research in Developmental Disabilities, 34*, 817-830.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development, 59*, 26-37.

Grant, C. M., Apperly I., & Oliver, C. (2007). Is theory of mind understanding impaired in males with Fragile X syndrome? *Journal of Abnormal Child Psychology, 35*, 17-28.

Gray. C. (2010). *The new Social Story book.* Arlington, TX: Future Horizons.

Gray, C. A. (1994). *Comic strip conversations*. Arlington, TX: Future Horizons.

Gray, C. (1995). Teaching children with autism to 'read' social situations. In K. A. Quill (Ed.), *Teaching children with autism: Strategies to enhance communication and socialization* (pp. 219-241). New York: Delmar Publishers Inc.

Gray, C. A. (1998). Social stories and comic strip conversations with students with Asperger syndrome and high functioning autism. In E. Schopler, G.B. Mesibov, & L.J. Kunce (Eds.) *Asperger syndrome or high functioning autism?* (pp. 167-194). New York: Plenum Press.

Gray, C. A. (2000). *The new social story book-illustrated edition*. Jenison, MI: Carol Gray.

Gray, C. A., & Garand, J. D. (1993). Social stories: Improving responses of students with autism with accurate social information. *Focus on Autistic Behavior, 8* (1), 1-10.

GreenBaum, R. L., Stevens, S. A., Nash, K., Korean, G., & Rovet, J. (2009). Social cognitive and emotion processing abilities of children with fetal alcohol spectrum disorders: A comparison with attention deficit hyperactivity disorder. *Alcoholism: Clinical and Experimental Research, 33*(10) 1656-1670.

Greenspan, S. I., & Wieder, S. (1998). *The child with special needs: Encouraging intellectual and emotional growth.* Reading, MA: Addison-Wesley.

Greenspan, S. I., & Wieder, S. (1997a). An integrated developmental approach to interventions for young children with severe difficulties in relating and communicating. *Zero to Three, 17*, 5-17.

Greenspan, S. I., & Wieder, S. (1997b). Developmental patterns and outcomes in infants and children with disorders in relating and communicating: A chart review of 200 cases of children with autistic spectrum diagnoses. *Journal of Developmental and Learning Disorders, 1*, 87-141.

Greenspan, S., & Wieder, S. (2001). *Floor Time Techniques and the DIR Model: For Children and Families with Special Needs.* Bethesda, MD. ICDL Publications.

Gresham, F. M., Sugai, G., & Horner, R. H. (2001). Interpreting outcomes of social skills training for students with high-incidence disabilities. *Exceptional Children, 67*(3), 331-344.

Grossman, M., Peskin, J., & San Juan, V. (in press). Thinking about a reader's mind: Fostering communicative clarity in the compositions of youth with autism spectrum disorders. *Journal of Autism and Developmental Disabilities*. DOI:10.1007/s10803-013-1786-y

Gutstein, S. E. (2000). *Autism/Aspergers: Solving the relationship puzzle*. Arlington, TX: Future Horizons.

Gutstein, S.E. (2009). Empowering families through Relationship Development Intervention: An important part of the biopsychosocial management of autism spectrum disorders. *Annals of Clinical Psychiatry, 21* (3), 174-182.

Gutstein, S.E., Burgess, A. F., & Montfort, K. (2007). Evaluation of the relationship development intervention program. *Autism, 11* (5), 397-411.

Gutstein, S. E., & Sheely, R. K. (2002a). *Relationship development intervention with children, adolescents & adults: Social and emotional development activities for Asperger syndrome, autism, PDD & NLD*. Philadelphia, PA: Jessica Kingsley Publishers.

Gutstein, S. E., & Sheely, R. K. (2002b). *Relationship development intervention with young children: Social and emotional development activities for Asperger syndrome, autism, PDD & NLD*. Philadelphia, PA: Jessica Kingsley Publishers.

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

Hans, S.L., Auerbach, J.G., Styr, B., & Marcus, J. (2004). Offspring of parents with schizophrenia: Mental disorders during childhood and adolescence. *Schizophrenia Bulletin, 30* (2), 303-315.

Happé, 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.

Happé, F. (1995). The role of age and verbal ability in theory of mind task performance in subjects with autism. *Child Development, 66*(3), 843-855.

Houssa, M., Nader-Grasbois, N., & Jacobs, E. (2014). Experimental study of short-term training in social cognition in preschoolers. *Journal of Education and Training Studies, 2*(1), 139-154.

Howlin, P., Baron-Cohen, S., & Hadwin, J. (1999). *Teaching children with autism to mind read: A practical guide*. NY: John Wiley & Sons.

Hutchins, T. L. (2012). Social Stories™. In P. A. Prelock & R. J. McCauley (Eds.). Treatment of Autism Spectrum Disorders: Evidence-Based Intervention Strategies for Communication and Social Interaction. Baltimore, MD: Brookes.

Hutchins, T., Bonazinga, L., Prelock, P. A., & Taylor, R. S. (2008). Beyond false beliefs: The development and psychometric evaluation of the Perceptions of Children's Theory of Mind Measure - Experimental Version (PCToMM-E). *Journal of Autism and Developmental Disorders, 38* (1), 143-155.

Hutchins, T. L., & Prelock, P. A. (2013a). Parents' perceptions of their children's social behaviors: The social validity of Social Stories™ and comic strip conversations. *Journal of Positive Behavior Interventions, 15*(3), 156-168.

Hutchins, T. L., & Prelock, P. A. (2013b). The social validity of Social Stories™ for supporting the behavioral and communicative functioning of children with autism spectrum disorder. *International Journal of Speech, Language, and Hearing.*

Hutchins, T. L., & Prelock, P. A. (2008). Supporting theory of mind development: Considerations and recommendations for professionals providing services to individuals with autism spectrum disorder. *Topics in Language Disorders, 28*(4), 338-362.

Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2012). Psychometric evaluation of the Theory of Mind Inventory: A study of typically developing children and children with autism spectrum disorders. *Journal of Autism and Developmental Disabilities*, *42*(3), 327-341.

Hutchins, T., Prelock, P. A., & Chace, W. (2008). Test-retest reliability of theory of mind tasks representing a range of content and complexity and adapted to facilitate the performance of children with ASD. *Focus on Autism & Other Developmental Disabilities, 23*(4), 195-206.

Ingersoll, B., Dvortcsak, A., Whalen, C., & Sikora, D. (2005). The effects of a developmental, social-pragmatic language intervention on rate of expressive language production in young children with ASD. *Focus on Autism and Other Developmental Disabilities. 20* (4) 213-222.

Jones, E., Carr, D., & Feeley, K. (2006). Multiple effects of joint attention intervention for children with autism. *Behavior Modification*, *30*, 782-834.

Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology & Psychiatry*, *47*, 611-620.

Kasari, C., Paparella, T, Freeman, S.N., & Jahromi, L. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology, 76*, 125-137.

Kasari, C., Gulsrud, A. C., Wong, C., Kwon, S., & Locke, J. (2010). Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of Autism and Developmental Disorders, 40*(9),1045-1056.

Kazak, S., Collins, G.M., & Lewis, V. (1997). Can young people with autism refer to knowledge states? Evidence from their understanding of 'know' and 'guess." *Journal of Child Psychology and Psychiatry, 38*(8), 1001-1009.

Keenan, T. (2003). The preschool years and beyond. In B. Repacholi, & V. Slaughter (Eds.), *Individual differences in theory of mind* (pp. 121-142). New York, Hove: Psychology Press.

Kincaid, J., Braby, R., & Mears, J. (1988). Electronic authoring and delivery of technical information. *Journal of Instructional Development*, 11, 8–13.

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.

Klin, A., Schultz, R., & Cohen, D. J. (2000). Theory of mind in action: Developmental perspectives on social neuroscience. In S. Baron-Cohen, H. Tager-Flusberg, & D. J. Cohen (Eds.), *Understanding other minds: Perspectives from developmental neuroscience* (pp. 357-390). Oxford, UK: Oxford University Press.

Koegel, R., & Koegel, L. (2006). *Pivotal Response Treatments for Autism*. Baltimore, MD: Brookes.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-2: A Developmental Neuropsychological Assessment Manual*. San Antonio, TX: The Psychological Corporation.

Kossewska, J. (in preparation). *How do deaf children understand other minds?* Unpublished master's thesis. Pedagogical University of Cracow.

Leekam, S. R., & Prior, M. (1994). Can autistic children distinguish lies from jokes? A second look at second-order belief attribution. *Journal of Children Psychology and Psychiatry, 35*(5), 901-915.

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.

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

Lorimer, P. A., Simpson, R. L,. Myles, B. S., & Ganz, J. B. (2002). The use of social stories as a preventative behavioral intervention in a home setting with a child with autism. *Journal of Positive Behavioral Interventions, 4*, 53-60.

McAfee, J. (2002). *Navigating the social world: A curriculum for individuals with Asperger's syndrome, high functioning autism and related disorders.* Arlington, TX: Future Horizons.

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

McConachie, H., Val Randle, V., Hammal, D., & LeCouteur, A. (2005). A controlled trial of a training course for parents of children with suspected autism spectrum disorders. *Journal of Pediatrics, 147,* 335-340.

McFee, J. (2002). *Navigating the social world: A curriculum for individuals with Asperger's syndrome, high functioning autism and related disorders*. Arlington, TX: Future Horizons.

McGhee, P. E. (1979). *Humor: Its origins and development*. San Francisco, W. H. Freeman and Company.

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.

Moore, C., & Furrow, D. (1991). The development of the language of belief: The expression of relative certainty. In D. Frye & C. Moore (Eds.), *Children's Theories of Mind: Mental States and Social Understanding* (173-193). Hillsdale, NJ: Erlbaum.

Mundy, P., & Burnette, C. (2005a). Joint attention and neurodevelopmental models of autism. In F. Volkmar, R. Paul, A. Klin & D. Cohen (Eds.), *Handbook of autism and pervasive developmental disorders* (Vol. 1, pp. 650-681). NY: Wiley.

Murray, D. S., Creaghead, N. A., Maning-Courtney, P., Shear, P. K., Bean, J., & Prendeville, J. (2008). The relationship between joint attention and language in children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 23* (1), 5-14.

Myles, B.S., Trautman, M.L., & Schelvan, R.L. (2004). The hidden curriculum: Practical solutions for understanding unstated rules in social situations. Shawnee Mission, KS: Autism Asperger Publishing.

National Autism Center (2009). The National Standards Report of *The National Standards Project—Addressing the need for evidence based practice guidelines for autism spectrum disorders*. Randolph, MA: National Autism Center.

Norris, C., & Dattilo, J. (1999). Evaluating effects of a social story on a young girl with autism. *Focus on Autism and Other Developmental Disabilities, 14,* 180-186.

Paul, R. (2007). *Language disorders from infancy through adolescence* (3rd ed.). Oxford, England: Mosby Elsevier.

Pajareya, K., & Nopmaneejumruslers, K. (2011). A pilot randomized controlled trial of DIR/Floortime™ parent training intervention for pre-school children with autistic spectrum disorders. *SAGE Publications and The National Autistic Society, 15*, 1-15.

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

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

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

Peterson, C., Wellman, H., & Liu, D. (2005). Steps in theory of mind development for children with deafness or autism. *Child Development*, *76*(2), 502-517.

Pillow, B. H. (1991). Children's understanding of biased social cognition. *Developmental Psychology*, *27*(4), 539-551.

Prelock, P. A. (2006). *Autism Spectrum Disorders: Issues in Assessment & Intervention*. Austin, TX: Pro-Ed Publishers.

Prelock, P.A. (2011). Understanding the National Standards Project: Evidence-based interventions to support the language and social communication of children with ASD. *Perspectives on School-Based Issues*, *12* (2), 68-76.

Prelock, P.A., Calhoun, J., Morris, H., & Platt, G. (2011). Supporting parents to facilitate communication and joint attention in their young children with ASD: Two pilot studies. *Topics in Language Disorders, 31* (3), 210-234.

Scattone, D. (2007). Social skills interventions for children with autism. *Psychology in the Schools, 44*(7), 717-726.

Scattone, D., Tingstrom, D. H., & Wilczynski, S. M. (2006). Increasing appropriate social interactions of children with autism spectrum disorders using Social Stories™. *Focus on Autism and Other Developmental Disabilities, 21* (4), 211-222.

Schertz, H. H., & Odom, S. L. (2007). Promoting joint attention in toddlers with autism: A parent-mediated developmental model. *Journal of Autism and Developmental Disorders, 37* (8), 1562-1575.

Schietecatte, I., Roeyers, H., & Warreyn, P. (2012). Exploring the nature of joint attention impairments in young children with ASD: Associated social and cognitive skills. *Journal of Autism and Developmental Disorders, 42* (1), 1-12.

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. *Cognition, 38*, 1–12.

Silliman, E. R., Diehl, S. F., Bahr, R. H., Hnath-Chisolm, T., Zenko, C. B., & 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.

Solomon, R., Necheles, J., Ferch, C., & Bruckman, D. (2007). Pilot study of a parent-training program for young children with autism: The PLAY Project Home Consultation model. *Autism: The International Journal of Research and Practice*, *11*, 205-224.

Sullivan, M., Finelli, J., Marvin, A., Garrett-Mayer, E., Bauman, M., & Landa, R. (2007). Response to joint attention in toddlers at risk for autism spectrum disorder: A prospective study. *Journal of Autism and Developmental Disabilities, 37*, 37-48.

Sullivan, K., Winner, E., & Hopfield, N. (1995). How children tell a lie from a joke: The role of second-order mental state attributions. *British Journal of Developmental Psychology, 13*, 191-204.

Sullivan, K., Zaitchik, D., & Tager-Flusberg, H. (1994). Preschoolers can attribute second-order beliefs. *Developmental Psychology, 30(3)*, 395–402.

Sussman, F. (1999). *More than words: Helping parents promote communication and social skills in children with autism spectrum disorders.* Toronto, Ontario: A Hanen Centre Publication.

Sussman, F. (2006). *Talkability: People skills for verbal children on the autism spectrum-A guide for parents.* Toronto, Ontario: A Hanen Centre Publication.

Swaggart, B. L., Gagnon, E., Bock, S. J., Earles, T. L., Quinn, C., Myles, B. S., & Simpson, R. L. (1995). Using social stories to teach social and behavioral skills to children with autism. *Focus on Autistic Behavior, 10* (1), 1-16.

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.

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.

Taylor, B. A., Levin, L., & Jasper, S. (1999). Increasing play-related statements in children with autism toward their siblings: Effects of video modeling. *Journal of Developmental and Physical Disabilities, 11*, 253-264.

Thiemann, K.S., & Goldstein, H. (2001). Social stories, written text cues, and video feedback effects on social communication of children with autism. *Journal of Applied Behavior Analysis, 34*, 425-446.

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore, & P. J. Dunham (Eds.), *Joint attention: It's origins and role in development* (pp. 103–130). Hillsdale, NJ: Erlbaum.

Uekermann, J., Kraemer, M., Abdel-Hamid, M., Schimmerlmann, B. G., Heberbrand, J., Daum, I., Wiltfang, J., & Kis, B. (2010). Social cognition in attention-deficit hyperactivity disorder (ADHD). *Neuroscience and Biobehavioral Reviews, 34*, 734-743.

Vivian, L., Hutchins, T. L., & Prelock, P. A. (2012). A family-centered approach for training parents to use comic strip conversations with their child with autism. *Contemporary Issues in Communication Science and Disorders, 39*, 30-42.

Vygotsky, L. (1934). *Thought and Language.* Cambridge, MA: MIT Press.

Walden, T., & Ogan, T. (1988). The development of social referencing. *Child Development, 59,* 1230-1240.

Wellman, H. M. (1988). First steps in the child's theorizing about the mind. In J. W. Astington, P. L. Harris, & D. R. Olson (Eds.), *Developing theories of mind* (pp. 64-92). New York: Cambridge University Press.

Wellman, H. M., & Banerjee, M. (1991). Mind and emotions: Children's understanding of the emotional consequences of beliefs and desires. *British Journal of Developmental Psychology, 9*, 191-214.

Wellman, H. M., & Bartsch, K. (1994). Before belief: Children's early psychological theory. In C. Lewis & P. Mitchell (Eds.), *Children's early understanding of mind: Origins and Development* (pp. 331-354). Hove, UK: Lawrence Erlbaum.

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.

Wetherby, A. (1986). Ontogeny of communicative functions in autism. *Journal of Autism and Developmental Disorders, 16,* 295-315.

Whalen, C., & Shreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry, 44*, 456-468.

White, P.J., O'Reilly, M., Streudsand, W., Levine, A., Sigafoos, J., Lancioni, G., Fragale, C., Pierce, N., & Aguilar, J. (2011). Best practices for teaching joint attention: A systematic review of the intervention literature. *Research in Autism Spectrum Disorders,* 5(4), 1283-1295.

Wieder, S., & Greenspan, S. I. (2005). Can children with autism master core deficits and become empathetic, creative, and reflective? *Journal of Developmental and Learning Disorders*, *9*, 39–61.

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

Winner, M. (2005) *Think social! A social thinking curriculum for school-aged students.* San Jose, CA: Think Social Publishing.

Winner, M. (2000) *Inside out: What makes a person with socio-cognitive deficits tick*? San Jose, CA: Think Social Publishing.

Winner, M. (2007a, 2nd ed.) *Thinking about you thinking about me.* San Jose, CA: Think Social Publishing.

Winner, M. (2007b) *Social behavior mapping- Connecting behavior, emotions and consequences across the day.* San Jose, CA: Think Social Publishing.

Winner, M. (2008). *Superflex: A superhero social thinking curriculum*. San Jose, CA: Think Social Publishing.

APPENDIX: Interpretation and Report Writing

One can interpret ToMI and ToM Task Battery scores in two ways: by considering the subtest scores in Early, Basic, and Advanced subtests and/or by looking at scores on individual items. Subtest scores can be used as an overall indicator of the individual's capacity for ToM across the three levels, which can be explicitly stated. For example, a student who scores within the clinically significant range for the Advanced and Basic ToM, but not on the Early ToM subtest is most likely demonstrating the need for ToM training within the Early ToM level and therefore may benefit from intervention on individual skills within that subtest. Once the level of ToM has been determined, scores on individual items within that subtest can be used to inform intervention targets.

Recall that item-level scores that reflect skills considered latent fall below 5, those considered emerging fall between 5 and 15, and those considered developed fall above 15. Once a general and/or specific level of ToM has been determined, an examination of relative strengths and challenges and adherence to a descriptive-developmental approach should provide a good foundation for intervention planning. Recall also that individual scores within subscales may reveal "splinter skills" within the more general level of ToM functioning.

For an example on interpretation of the ToMI and ToM Task Battery, please reference the sample report provided below, describing results of the ToMI for a young adult with ASD.

SAMPLE REPORT:

MF is an 18 year, 11 month old male with a diagnosis of Pervasive Developmental Disorder- Not Otherwise Specified (PDD-NOS). MF is primarily a non-verbal communicator; however, he does use intermittent one to four word utterances, with some level of prompting or incentive, which are often characterized by atypical use of tone, stress and volume. He communicates his basic wants and needs, initiates some communication related to preferred topics (e.g., Disney characters) and responds well to communicative partners who are familiar, clearly attending to him, and who show an interest in interacting with him. His father completed the ToMI and the clinician administered the ToM Task Battery as part of a comprehensive assessment.

COMPOSITE SCORE

Composite Mean = *9.77 Percentile = 1st

EARLY ToM SUBSCALE

Subscale Mean = *14.99 (of 20) Percentile =1st

Subscale Item Scores

Item	designed to assess	score
3.	affect recognition (complex)	14.8
6.	affect recognition (emotion-expression relationship)	15.1
24.	Intentionality	15.3
25.	affect recognition (basic)	15.2
29.	social referencing	5.50
37.	sharing attention – initiating	20.0
38.	sharing attention - responding	20.0

BASIC ToM SUBSCALE

Subscale Mean = *9.67 (of 20) Percentile = 1st

Subscale Item Scores

Item	designed to assess	score
1.	physiologically-based behavior	5.3
4.	emotion-based behavior	15.2
7.	mental state term comprehension (think)	0.0
8.	false beliefs in context of unexpected change of location	4.8

9.	seeing leads to knowing	14.8
10.	mental state term comprehension (know)	0.0
11.	appearance-reality distinction	20.0
12.	false beliefs in context of unexpected contents	0.0
15.	Certainty (knowing/guessing)	4.8
16.	mental-physical distinction	9.8
26.	Pretense	20.0
29.	counterfactual reasoning	14.8
30.	mental-physical distinction	15.0
31.	ability to deceive	15.5
32.	level 1 visual perspective-taking	5.6
33.	speech acts (promises)	15.3
35.	speech acts (secrets)	5.4
39.	mental state term comprehension (believe)	4.8
42.	attribute-based behavior	15.3

ADVANCED ToM SUBSCALE

Subscale Mean = *7.64 (of 20) Percentile = 1st

Subscale Item Scores

Item	designed to assess	score
2.	Sarcasm	4.5
5.	second-order false desire attribution	0.0
13.	idiomatic language	20.0
14.	use of language to intentionally deceive	4.8
17.	understanding display rules	4.8
18.	complex social judgment	9.5
19.	white lies	4.5
20.	understanding lies versus jokes	5.0

21.	level 2 visual perspective-taking	15.0
22.	second order understanding of belief	4.8
23.	second order understanding of emotion	9.5
27.	complex social judgment	20.0
34.	Empathy	20.0
36.	humor (play on words)	0.0
40.	biased cognition	0.0
41.	mind as active interpreter	0.0

Theory of Mind Inventory (ToMI). The results of the ToMI, completed by MF's father, revealed a composite mean score of 9.77 (possible range 0-20), placing him below the 10th percentile when compared to a typically developing population of 11 years. According to his ToMI scores, MF is most competent in Early ToM (subscale mean = 14.98) and demonstrates the development of joint or shared attention (in both initiation and responding to bids for shared attention) and emerging capacities in the areas of affect (emotion) recognition, intentionality (i.e., understanding that others have intentions and reading motivational states), and social referencing (looking at others to gauge their inner mental states).

ToMI scores also indicate some competence in Basic ToM (subscale score = 9.67) although splintered skill sets seem evident. More specifically, mental state term (i.e., 'know', 'think', 'believe') comprehension, the understanding of false beliefs, and the understanding of certainty (e.g., there is a difference between knowing and guessing), are characterized as 'latent' meaning they are not yet demonstrated. On the other hand, Basic ToM understandings that may be emerging are as follows: physiologically-based behavior (knowing that physiological states guide behavior; e.g., being cold drives one to seek warmth), emotion-based behavior (e.g., knowing that emotion states guide behavior; e.g., if someone is afraid of the dark, she will not want to go into a dark room), seeing-leads-to-knowing (understanding that people can know about things through visual observation), counterfactual reasoning (hypotheticals), the mental-physical distinction (the difference between mental and physical entities; e.g., an actual dog vs. the idea of a dog), the ability to deceive, level 1 visual perspective-taking (understanding line of regard), and speech acts (e.g., promises, secrets). Finally, Basic ToM skills that appear to be reliably developed include: pretense (pretending) and the appearance-reality distinction (e.g., understanding that objects can look like one thing but really be another thing; e.g., a candle that looks like an apple).

Advanced ToM is limited (subscale score = 7.64) and again, splintered skills seem evident. Specifically, understanding of sarcasm, second-order false beliefs, purposeful deceit by others, white lies, humor (play on words), biased cognition (understanding that previous knowledge about a person can color our interpretation of their behavior), and mind as active interpreter (understanding that the mind itself constructs experience) are latent and not yet developed. By contrast, some competence may be emerging in the areas complex social judgment, level 2 visual perspective-taking (understanding that others have different vantage points), and second-order understanding of emotion (having emotions about emotions). Finally Advanced ToM skills that appear to be reliably developed include the understanding of idiomatic expression and empathy.

Theory of Mind Task Battery. The Theory of Mind (ToM) Task Battery is a child performance measure designed to assess ToM competency in a variety of areas. Those areas that would be considered part of Early ToM skills and that are typically developed by age 2-3 include emotion recognition (happy, sad, mad, scared) and desire-based emotion (e.g., knowing that people are happy when they get what they want). Areas that would be considered part of Basic

ToM skills that are typically developed in the preschool years include: seeing-leads-to-knowing, visual perspective-taking, perception-based action (e.g., knowing that people will look for their keys based on where they last saw them), and first order beliefs. Areas that would be considered part of more Advanced ToM skills and that typically develop in middle and later childhood include: second-order beliefs (beliefs about beliefs), belief- and reality-based emotion (e.g,. predicting how someone will feel based on what they think and how things really are), and understanding the message-desire discrepancy (inferring belief of other when interpreting a statement of desire).

Test questions are embedded in a series of vignettes that are accompanied by color illustrations. Each story is read to the child who can respond to test questions verbally or by pointing to one of four pictures that depicts the answer. The ToM test questions increase in difficulty as the test proceeds. A total of 15 points (one for each question) are possible.

Most of MF's responses were obtained by him pointing to the answer and he obtained a score of 6/15. His response pattern indicated that he demonstrated skills consistent with Early ToM by passing the first two tasks (i.e., emotion recognition and desire-based emotion) for a total of 5 points. In addition, he passed one task (i.e., seeing-leads-to-knowing) consistent with Basic ToM skills. When ToM Task Battery items were failed, it was almost always the case that the associated memory control questions were also failed. Because MF appeared to enjoy and be engaged in the task, it is likely that failure on these items reflect limitations in language comprehension and/or working memory. Following five consecutive incorrect answers, testing was terminated as is required on this test.

The results of the ToM Task Battery were consistent with those of the ToMI in revealing general competency in Early ToM development and strength in the Basic ToM skill of seeing-leads-to-knowing. Beyond this, however, the results of the ToM Task Battery have limited utility. Thus, the results of the ToMI may be more useful in guiding treatment planning. In summary, the results of ToM assessment reveal emerging competency in Early ToM areas including intentionality, emotion recognition, and social referencing and Basic ToM areas including physiologically- and emotion-based behavior, seeing-leads-to-knowing, counterfactual reasoning, the mental-physical distinction, the ability to deceive, level 1 visual perspective-taking, and the understanding of speech acts. From a descriptive-developmental perspective, intervention should target these emerging competencies. Immediate treatment goals are as follows:

Goal Area 1:

MF will identify his emotional states in real-life social scenarios with ___ accuracy out of ___ trials.

Goal Area 2:

MF will recognize the emotional states of others in real-life social scenarios with ___ accuracy out of ___ trials.

Goal Area 3:

MF will demonstrate understanding of others' intentions by responding to nonverbal cues that signal a desire for an object within reach for ___out of ___opportunities.

Goal Area 4:

MF will match a character's mental state, specifically a basic physiologic or emotional state with a behavior in a picture-based story when presented with at least one situational cue with ___ accuracy.

Recommended Intervention Approaches:

- Joint Attention Training
- Relationship Development Intervention
- Social Stories™
- Comic Strip Conversations
- Teaching Children with Autism to Mind Read
- Teaching Perspective Taking to Children with Autism Spectrum Disorders