

Brief Report: An Independent Replication and Extension of Psychometric Evidence Supporting the Theory of Mind Inventory

Kathryn J. Greenslade^{1,2}  · Truman E. Coggins³

© Springer Science+Business Media New York 2016

Abstract This study presents an independent replication and extension of psychometric evidence supporting the *Theory of Mind Inventory (ToMI)*. Parents of 20 children with ASD (4; 1–6; 7 years; months) and 20 with typical development (3; 1–6; 5), rated their child’s theory of mind abilities in everyday situations. Other parent report and child behavioral assessments included the *Social Responsiveness Scale-2*, *Vineland Adaptive Behavior Scales-2*, *Peabody Picture Vocabulary Test-4*, and *Clinical Evaluation of Language Fundamentals-Preschool, 2*. Results revealed high internal consistency, expected developmental changes in children with typical development, expected group differences between children with and without ASD, and strong correlations with other measures of social and communication abilities. The *ToMI* demonstrates strong psychometrics, suggesting considerable utility in identifying theory of mind deficits in children with ASD.

Keywords Theory of mind · Assessment · Reliability · Validity

Some data from this manuscript has been presented as a poster at the 2015 International Meeting for Autism Research, and as an oral presentation at the 2015 American Speech-Language-Hearing Convention.

✉ Kathryn J. Greenslade
Greenslade@kennedykrieger.org

¹ Center for Autism and Related Disorders, Kennedy Krieger Institute, 3901 Greenspring Ave., Baltimore, MD 21211, USA

² Psychiatry and Behavioral Sciences, The Johns Hopkins University School of Medicine, Baltimore, MD, USA

³ Department of Speech and Hearing Sciences, University of Washington, 1417 NE 42nd St., Seattle, WA 98105, USA

Introduction

Deficits in theory of mind are a common manifestation of the social-communication challenges in individuals with autism spectrum disorder (ASD; Baron-Cohen et al. 1985; Perner et al. 1989). Depending on age and functioning, individuals with ASD demonstrate reduced joint attention (e.g., Mundy et al. 2009) and false belief understanding (e.g., Baron-Cohen et al. 1985), and poor performance on advanced theory of mind tasks, assessing nuanced social situation/cue interpretation (e.g., Happé 1994). Unfortunately, existing behavioral tasks have drawbacks, such as dichotomous scoring, ceiling effects, an inability to assess those with lower cognitive and verbal abilities, and questionable ecological validity (Hutchins et al. 2012; Tahiroglu et al. 2014).

The *Theory of Mind Inventory (ToMI)* (Hutchins et al. 2010) is a promising new tool that purports to measure theory of mind and social-cognitive functioning via parent report. Initial findings of measurement quality have revealed strong internal consistency (Hutchins et al. 2012; Lerner et al. 2011; Pujals et al. 2016). Furthermore, accumulated evidence supports the construct validity of *ToMI* scores. For example, Hutchins et al. (2012) and Pujals et al. (2016) reported significant score differences between children with ASD and those with typical development (TD). Hutchins and colleagues (Lerner et al. 2011; Hutchins et al. 2012) reported strong correlations between *ToMI* scores and scores on the *Social Communication Questionnaire* (Rutter et al. 2005; $r = -.55$, $p < .01$), *Social Skills Rating System-Parent* (Gresham and Elliot 1990; $r = .61$, $p < .001$), *Social Responsiveness Scale* (Constantino and Gruber 2005; $r = -.75$, $p < .001$), and *Peabody Picture Vocabulary Test-Fourth Edition* (Dunn and Dunn 2007; $r = .73$, $p < .05$) for participants with

ASD, and on the Theory of Mind Task Battery (Hutchins et al. 2008) for ASD and TD samples (ASD: $r = .66$, $p < .05$; TD: $r = .82$, $p < .05$). Finally, a principle components analysis revealed three factors corresponding to “Early Theory of Mind: Reading Affect and Sharing Attention”, “Basic Theory of Mind: Metarepresentation and Developmentally Related Understandings”, and “Advanced Theory of Mind: Complex Recursion, Mind as Active Interpreter, and Social Judgment” (Hutchins et al. 2012, pp. 334–335).

While the prefatory evidence is heartening, several non-trivial issues have yet to be experimentally addressed. Specifically, correlations between *ToMI* scores and broader measures of social abilities have only been reported for adolescents with ASD; no comparable data has been presented for a TD control group or in other age ranges (Lerner et al. 2011). Correlations with verbal abilities have only examined receptive vocabulary, not general language (Hutchins et al. 2012). To date, the vast majority of psychometric data have been collected by the *ToMI*'s creators. As in all scientific pursuits, replication by independent researchers is essential. Finally, and importantly, reliability and validity data for the English version of the *ToMI* has been based solely on children with parent-reported, community-based ASD diagnoses, rather than diagnoses confirmed with a gold-standard diagnostic measure (Hutchins et al. 2012; Lerner et al. 2011).

The current research directly addresses these concerns by providing construct validity evidence for using *ToMI* scores to evaluate children's theory of mind abilities. Specifically, it provides psychometric evidence from independent researchers using (1) children with ASD whose diagnoses were confirmed by a gold standard diagnostic measure, and (2) a matched control group of children with TD. Psychometric evidence will address internal consistency, developmental changes in children with TD, group differences between ASD and TD groups, and criterion-related validity based on correlations with social and communication measures.

ToMI scores were hypothesized to be internally consistent, indicating that all items measure a single construct: theory of mind abilities. Within the TD group, *ToMI* scores were expected to increase with increasing age, reflecting typical theory of mind development, and to be significantly greater after 4 years of age, as performance on false belief tasks go from chance levels to above chance levels around this age (Wellman et al. 2001). Because theory of mind deficits are common in children with ASD, *ToMI* scores were expected to be significantly lower in the ASD group as compared to the TD group. Finally, *ToMI* scores were expected to be strongly correlated with other measures of social and cognitive abilities. Expected results would

support the *ToMI*'s use for assessing theory of mind abilities in children with and without ASD.

Methods

Participants

Twenty verbal children with ASD and 20 children with TD were recruited in the Pacific Northwest and Southwest United States, with human subjects approval. All parents provided informed consent for themselves and their child; whenever able, children provided verbal assent. All participants were recruited for a larger study examining children's prosocial behaviors; this brief report presents the analysis of pre-existing data, collected for the larger study. Children were matched for gender (14 male, 6 female) and verbal mental age (± 6 months), to account for potential associations with theory of mind abilities (Happé 1995).

Children with ASD had community diagnoses, which were confirmed through administration of the *Autism Diagnostic Interview-Revised* (Lord et al. 1994). All children were English speakers with intelligible speech and no history of permanent hearing loss, seizure disorder, or brain injury/disease. Also, children with TD had no reported history of special education services or language disorder, or family history of ASD.

Measures

All parents completed the *ToMI* (Hutchins et al. 2010), *Social Responsiveness Scale-2* (*SRS-2*; Constantino and Gruber 2012), and Survey Interview of the *Vineland Adaptive Behavior Scales-2* (*Vineland-2*; Sparrow et al. 2005). A nationally-certified speech-language pathologist administered the *Peabody Picture Vocabulary Test-4* (*PPVT-4*; Dunn and Dunn 2007) and *Clinical Evaluation of Language Fundamentals-Prseshool, 2* (*CELF-P2*; Wiig et al. 2004).

The *ToMI* (Hutchins et al. 2010) measures parental confidence in their child's ability to demonstrate theory of mind abilities in everyday situations, using ratings on a visual analog scale with anchors from “definitely not” to “definitely.” Items measure abilities ranging from early emerging joint attention and social referencing, to advanced understanding of figurative language and subtle social distinctions (e.g., teasing vs. bullying; Hutchins et al. 2012). The *ToMI* generates a Composite score between 0 and 20, with higher scores indicating higher parental confidence in their child's understanding of targeted theory of mind abilities. Prior psychometric evidence for the *ToMI* was presented in the introduction.

The *SRS-2* (Constantino and Gruber 2012) measures parent perceptions of children's social awareness, social communication, social cognition, social motivation, and restricted interests/repetitive behaviors. It generates *T*-scores ($M = 50$, $SD = 10$), where higher scores indicate greater social impairment. The *SRS-2* demonstrates good internal consistency, inter-rater reliability, concurrent validity, and sensitivity/specificity. The *SRS-2* confirmed age appropriate social abilities ($\leq 59T$) in participants with TD; no criterion was set for those with ASD.

The *Vineland-2* (Sparrow et al. 2005) assesses adaptive functioning in everyday contexts in the domains of socialization, communication, motor, and daily living skills. It generates standard scores ($M = 100$, $SD = 15$); lower scores reflect lower adaptive functioning. The *Vineland-2* demonstrates good test–retest reliability, internal consistency, construct validity (expected group differences), and criterion-related validity (concurrent, discriminant) as well as minimal bias as measured by differential item functioning. The *Vineland-2* confirmed age appropriate adaptive functioning in all participants with TD (≥ 85); no criterion was set for those with ASD.

The *PPVT-4* (Dunn and Dunn 2007) is a standardized measure of receptive vocabulary, with good test–retest, split half, and alternate form reliability and strong criterion-related validity. It generates standard scores ($M = 100$, $SD = 15$); lower scores reflect lower verbal abilities. The *PPVT-4* age equivalence score was used to match verbal abilities across groups. Standard scores confirmed age appropriate receptive vocabularies in children with TD (≥ 85); no criterion score was set for the ASD group, though all were required to have an age equivalent of ≥ 32 months.

The *CELF:P2* (Wiig et al. 2004) measures general language abilities, and has good test–retest reliability, internal consistency, concurrent validity, and sensitivity/specificity. It generates standard scores ($M = 100$, $SD = 15$), with lower scores reflecting lower verbal abilities. (Note: Two children with TD and two with ASD did not complete all subtests comprising the *CELF-P2*'s Core Language Index.) The Core Language Index confirmed age appropriate general language abilities in participants with TD (≥ 85); no criterion score was set for those with ASD.

Results

Descriptive Statistics

The groups were roughly equivalent in terms of ethnicity ($X^2 = .63$, $p = .43$), non-Caucasian race ($X^2 = 1.29$, $p = .26$), household income ($X^2 = 10.15$, $p = .07$), and caregiver education ($X^2 = 4.68$, $p = .10$). Table 1

summarizes children's chronological age and standardized test scores, along with the results of significance testing. Because children were matched on their verbal age equivalent on the *PPVT-4*, chronological age was significantly lower in the TD group than the ASD group. Because the standardized parent report and child behavioral assessment all examined social, communication, and language abilities that are characteristically impaired in children with ASD, significantly poorer performance was observed in the ASD group across measures. Despite these group differences, 18 of 20 children with ASD received standard scores within 1 SD of the mean on the *PPVT-4*, and 15 of 18 did so on the *CELF-P2* Core Language Index.

Psychometric Analyses

To examine whether *ToMI* items are internally consistent, Cronbach's alpha was calculated, revealing a high degree of internal consistency ($\alpha = .96$). No improvement was noted with the removal of any item(s), indicating that all 42 items contribute to the *ToMI* Composite score and its measurement of the construct of theory of mind abilities.

To examine whether *ToMI* scores reflect expected developmental changes in theory of mind abilities, two analyses were conducted using *ToMI* Composite scores from the TD group. First, a Pearson's correlation r between chronological age and *ToMI* Composite score was strong, $r = .59$, $p < .01$, supporting the prediction that *ToMI* scores would increase as age increased. Approximately 35 % of the variability in TD group's *ToMI* scores was shared with variability in age. Second, to account for non-normal distributions, a Mann–Whitney test was used to compare the *ToMI* Composite scores of younger (< 48 months, $n = 10$; $M = 12.95$, $SD = 2.91$) and older children (≥ 48 months, $n = 10$; $M = 15.56$, $SD = 1.66$). Results confirmed significantly lower scores in younger than older children, $Z = 2.34$, $p = .02$, supporting predictions of increased *ToMI* scores following expected improvements in false belief understanding around 4 years of age.

To examine whether *ToMI* scores reflect expected group differences, a Mann–Whitney test was used to compare the non-normally distributed *ToMI* Composite scores of children with ASD and those with TD. Results were consistent with expected group differences: *ToMI* Composite scores of children with ASD ($M = 10.83$, $SD = 2.03$) were significantly lower than those of children with TD ($M = 14.25$, $SD = 2.66$), $Z = 3.84$, $p < .001$. Descriptively, 18 of 20 children with ASD (90 %) and only 1 of 20 with TD (5 %) received scores corresponding to a percentile rank of 16 or lower (i.e., ≥ 1 SD below the mean). A Pearson Chi squared test confirmed that significantly more children with ASD received a percentile rank below this level, $X^2 = 28.97$, $p < .001$.

Table 1 Standardized test results

	Group				Significance	
	ASD		TD		Z	p
	Mean (SD)	Range	Mean (SD)	Range		
Age (months)	63.85 (9.53)	48–79	50.75 (13.39)	37–78	3.16	.002
<i>ToMI</i>	10.83 (2.03)	5.55–13.71	14.25 (2.66)	7.50–18.64	3.84	<.001
<i>Vineland-2</i>	87.35 (11.51)	68–108	104.95 (7.91)	92–119	4.41	<.001
<i>SRS-2</i>	72.30 (9.76)	58–90	45.70 (4.21)	37–54	5.41	<.001
<i>PPVT-4</i>	100.55 (15.07)	77–139	116.75 (10.10)	96–137	3.47	.001
VMA (months)	65.20 (19.13)	32–104	65.35 (18.88)	38–106	.12	.90
<i>CELF:P2</i> (N = 18 per group)	92.72 (10.59)	73–116	109.61 (6.37)	96–121	4.28	<.001

ASD autism spectrum disorder, *TD* typically developing, *ToMI* Theory of Mind Inventory, *Vineland-2* vineland adaptive behavior scales-second edition, *SRS-2* social responsiveness scale-second edition, *PPVT-4* Peabody Picture Vocabulary Test-4th Edition, *VMA* verbal mental age (i.e., *PPVT-4* Age Equivalent), *CELF:P2* Clinical Evaluation of Language Fundamentals: Preschool, Second Edition

To examine whether *ToMI* scores demonstrate expected relationships with scores on measures of social and communication abilities (criterion-related validity), Pearson's correlations were calculated between *ToMI* Composite scores and standard scores on the *SRS-2*, *Vineland-2* Socialization and Communication subscales, *PPVT-4*, and *CELF-P2*. Table 2 summarizes these correlations and corresponding *p* values. Correlations between *ToMI* scores and measures of social abilities were strong: *SRS-2* Total score, $r = -.67$, $p < .001$; *Vineland-2* Socialization score, $r = .62$, $p < .001$. Variability in *ToMI* Composite scores was shared with approximately 45 and 39 % of the variability in *SRS-2* Total scores and *Vineland-2* Socialization scores, respectively. Furthermore, the five *SRS-2* subscales were all strongly correlated with *ToMI* Composite scores. Note that as expected, these correlations were all negative, because higher *SRS-2* scores indicate a higher degree of social impairment, whereas a higher *ToMI* scores correspond with better theory of mind abilities. Strong correlations were also found between *ToMI* Composite scores and measures of communication abilities: *Vineland-2* Communication subscale score, $r = .58$, $p < .001$; *PPVT-4* standard scores, $r = .53$, $p < .001$; *CELF-P2* Core Language Index, $r = .58$, $p < .001$. Variability in *ToMI* scores was shared with approximately 34, 28, and 34 % of the variability in *Vineland-2* Communication, *PPVT-4*, and *CELF-P2* scores, respectively.

Discussion

The present study provides an independent replication and extension of previous reliability and validity evidence, supporting the use of *ToMI* scores to identify theory of

mind deficits in young children with ASD. Evidence indicated that items on the *ToMI* are internally consistent, with all items contributing to the measurement of a common construct, namely theory of mind abilities. Moreover, developmental changes were clearly reflected in *ToMI* scores, such that older children who were expected to have better theory of mind abilities received higher scores. Group differences were also evident: children with confirmed ASD diagnoses received lower *ToMI* scores than children with TD, indicating expected theory of mind deficits in the ASD group. *ToMI* percentile ranks, likewise, revealed group differences, with 90 % of children with ASD scoring at least one standard deviation below the mean, and 95 % of children with TD scoring above this cut-off. Finally, strong correlations were found between *ToMI* scores and all measures of social and communication abilities examined, providing excellent criterion-related validity evidence supporting the use of *ToMI* scores to measure theory of mind abilities.

This strong reliability and construct validity evidence underscores the diagnostic utility of the *ToMI*. The findings represent the first empirical evidence of the strong psychometric properties of the *ToMI* (English version) based on a sample of children whose ASD diagnoses were confirmed through administration of a gold-standard diagnostic tool, the *ADI-R*. It also extends previous research by providing criterion-related validity evidence in preschool-aged children with ASD and a control group of children with TD. Moreover, the evidence also supports the relationship between *ToMI* scores and an omnibus measure of language abilities (*CELF-P2*). Thus, by addressing previous limitations, the present results considerably increase our confidence in the *ToMI*'s reliability and validity for measuring theory of mind abilities.

Table 2 Correlations between scores on the theory of mind inventory and measures of social, communication, and language abilities

Measure	Theory of Mind Inventory (ToMI)	
	Pearson's <i>r</i>	<i>p</i> value
1. <i>Social Responsiveness Scale-2 (SRS-2)</i>	-.67	<.001
Social awareness subscale	-.62	<.001
Social cognition subscale	-.71	<.001
Social communication subscale	-.63	<.001
Social motivation subscale	-.65	<.001
Restricted interests/repetitive behaviors subscale	-.61	<.001
2. <i>Vineland-2 Socialization subscale</i>	.62	<.001
3. <i>Vineland-2 Communication subscale</i>	.58	<.001
4. <i>Peabody Picture Vocabulary Test-4 (PPVT-4)</i>	.53	<.001
5. <i>Clinical Evaluation of Language Fundamentals-Preschool, 2nd Edition (CELF-P2)</i>	.58	<.001

It should be noted that the present study did not provide independent replication of test–retest reliability, or reliability and validity evidence in children older than 6 years, 7 months. Thus, independent replication in older children and adolescents will be an important aim of future research. Furthermore, the present evidence was based on samples from urban and suburban regions in the Pacific Northwest and Southwest United States. In general, parents were well-educated, and household incomes were relatively high. Independent replication in a more geographically and economically diverse sample would be beneficial. Reliability and validity evidence addressing these limitations would further contribute to the quality and quantity of evidence supporting the use of *ToMI* scores.

The current study advances our confidence in the use of *ToMI* scores as effective and efficient indicators of children's understanding, appreciation, and representation of their own and others' mental states as a means of making sense of everyday social situations. Because of its continuous measurement, the *ToMI* offers a marked methodological advantage over many behavioral tasks. Scores on individual items are able to differentiate between abilities that are firmly grasped, emerging, or beyond the child's current zone of proximal development, and emerging abilities can be flagged as potential treatment targets. Like other parent report measures being developed (Tahiroglu et al. 2014), the *ToMI*'s breadth provides a comprehensive view of the child's theory of mind abilities, which can reveal uneven profiles and identify potential targets (e.g., social referencing, understanding humor) within and across domains (early, basic, advanced). In addition, its design enables the identification of treatment targets for children with ASD who have more advanced theory of mind abilities as well as those with lower verbal and cognitive abilities. Thus, the *ToMI* may help clinicians design appropriate, individualized treatments to address the needs of children across the autism spectrum.

Acknowledgments The current study and its first author were supported by the University of Washington's Department of Speech and Hearing Sciences, University of Washington Gatzert Fellowship, and the American Speech-Language-Hearing Foundation. We greatly appreciate our recruitment sources, including Dr. Annette Estes and studies at the University of Washington Autism Center as well as Anita Werner and the staff of Pediatric Speech and Language Specialists, Scottsdale, Arizona. We would also like to acknowledge the research assistants who contributed to this work. Finally, we would like to thank the families and children who participated in this study, without whom this work would not have been possible.

Author Contribution KJG participated in conceiving of the study and its design; she also coordinated the study; collected, analyzed, and interpreted the data; and drafted/revised the manuscript. TEC participated in conceiving of the study and its design, interpreting the data, and drafting/revising the manuscript.

Compliance with Ethical Standards

Ethical Approval All procedures performed in this study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of interest The authors declare that they have no conflict of interest.

References

- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind?". *Cognition*, 21(1), 37–46.
- Constantino, J. N., & Gruber, C. P. (2005). *The social responsiveness scale*. Los Angeles, CA: Western Psychological Services.
- Constantino, J. N., & Gruber, C. P. (2012). *The social responsiveness scale, second edition (SRS-2)*. Los Angeles, CA: Western Psychological Services.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test, fourth edition (PPVT-4)*. Minneapolis, MN: Pearson Assessments.
- Gresham, F., & Elliot, S. (1990). *Social skills rating system manual*. Circle Pines, MN: American Guidance Service.
- Happé, F. G. E. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able

- autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, 24(2), 129–154.
- Happé, F. (1995). The role of age and verbal ability in the theory of mind task performance of subjects with autism. *Child Development*, 66(3), 843–855.
- Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2010). *The theory of mind inventory (ToMI)*. Available from <http://www.theoryofmindinventory.com/download-tests-test-materials/>.
- Hutchins, T. L., Prelock, P. A., & Bonazinga, L. (2012). Psychometric evaluation of the theory of mind inventory (ToMI): A study of typically developing children and children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 42, 327–341.
- Hutchins, T. L., 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 and Other Developmental Disabilities*, 23(4), 195–206.
- Lerner, M. D., Hutchins, T. L., & Prelock, P. A. (2011). Brief report: Preliminary evaluation of the theory of mind inventory and its relationship to measures of social skills. *Journal of Autism and Developmental Disorders*, 41, 512–517.
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659–685.
- Mundy, P., Sullivan, L., & Mastergeorge, A. M. (2009). A parallel and distributed processing model of joint attention, social-cognition and autism. *Autism Research*, 2(1), 2–21.
- Perner, J., Frith, U., Leslie, A. M., & Leekham, S. R. (1989). Exploration of the autistic child's theory of mind: Knowledge, belief, and communication. *Child Development*, 60, 689–700.
- Pujals, E., Batlle, S., Camprodon, E., Pujals, S., Estrada, X., Aceña, M., et al. (2016). Brief report: Translation and adaptation of the theory of mind inventory to Spanish. *Journal of Autism and Developmental Disorders*, 46(2), 685–690.
- Rutter, M., Bailey, A., & Lord, C. (2005). *SCQ: The social communication questionnaire manual*. Los Angeles, CA: Western Psychological Services.
- Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland adaptive behavior scales: Second edition (Vineland-2)*. Livonia, MN: Pearson Assessments.
- Tahiroglu, D., Moses, L. J., Carlson, S. M., Mahy, C. E. V., Olofson, E. L., & Sabbagh, M. A. (2014). Children's social understanding scale: Construction and validation of a parent-report measure for assessing individual differences in children's theory of mind. *Developmental Psychology*, 50(11), 2485–2497.
- Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655–684.
- Wiig, E. H., Secord, W. A., & Semel, E. (2004). *Clinical evaluation of language fundamentals: Preschool* (2nd ed.). San Antonio, TX: Psychological Corporation.