Research Article

Validation of the Persian Theory of Mind Inventory-2 (ToMI-2-P): Psychometric Properties in Adolescents with Autism Spectrum Disorder

Negin Borjianboroujeni (iD) [1], Javad Hatami (iD) [1, 2, *], Mehdi Alizadeh Zarei (iD) [3], Hadi Parhoon (iD) [4], Shafagh Saei (iD) [3]

[1] Institute for Cognitive Sciences Studies, Tehran, Iran

[2] Department of Psychology, University of Tehran, Tehran, Iran

[3] Rehabilitation Research Center, Department of Occupational Therapy, School of Rehabilitation Sciences, Iran University of Medical Sciences, Tehran, Iran

[4] Department of Psychology, Razi University, Kermanshah, Iran

Corresponding Author: Institute for Cognitive Sciences Studies, Tehran, Iran. Email: Hatamijm@ut.ac.ir

Abstract

Background: The Theory of Mind (ToM), which refers to the capacity to infer the thoughts, beliefs, and emotions of others, plays a crucial role in facilitating successful social interactions. This capacity is often impaired in individuals with autism spectrum disorder (ASD). The Theory of Mind Inventory-2 (ToMI-2) is a commonly utilized caregiver-report measure for evaluating ToM; however, a validated Persian version suitable for clinical and research applications is currently unavailable.

Objectives: This study translated and culturally adapted the Theory of Mind Inventory-2 for Persian speakers (ToMI-2-P) and evaluated its psychometric properties in adolescents with ASD.

Methods: Following established cross-cultural adaptation guidelines, the ToMI-2 was translated into Persian. Psychometric testing involved 70 parents of adolescents with ASD and a matched typically developing control group. Assessments included face and content validity, convergent validity (via correlation with the Social Responsiveness Scale-2 [SRS-2]), discriminative validity using multivariate analysis of variance (MANOVA), internal consistency (Cronbach's α), and test-retest reliability [intra-class correlation coefficient (ICC)].

Results: Face and content validity were strong, with content validity ratio (CVR) values ranging from 0.50 to 1.00 and Content Validity Index (CVI) from 0.80 to 0.95. The MANOVA revealed significant group differences [Wilks' Lambda F(3, 86) = 138.02, P < 0.001, η^2 = 0.820], with medium to large effect sizes (Cohen's d = 0.19 - 0.75). Convergent validity was supported by a significant negative correlation with SRS-2 scores (R = -0.43, P < 0.01). The ToMI-2-P showed good internal consistency (α = 0.70 - 0.77) and excellent test-retest reliability (ICC = 0.81 - 0.85).

Conclusions: The ToMI-2-P exhibits strong psychometric properties, supporting its use as a reliable and valid tool for assessing ToMI-2-P with ASD. Further research is needed to explore cultural adaptations for specific items.

Keywords: Adolescent, Autism Spectrum Disorder (ASD), Psychometrics, Theory of Mind, Validity and Reliability

1. Background

Theory of Mind (ToM) refers to the cognitive ability to attribute mental states — such as beliefs, intentions, desires, and emotions — to oneself and others. This skill is fundamental for navigating social situations, forming relationships, and regulating behavior within a social context. Deficits in ToM are frequently observed in individuals with autism spectrum disorder (ASD), contributing to challenges in communication, social interaction, and overall quality of life ([1]-[3]). The ASD is a developmental condition characterized by difficulties in social interaction and communication, along with the presence of repetitive behaviors and narrowly focused

interests ([4]). The ToM impairments are a core feature of ASD and are associated with social maladjustment, difficulties in interpersonal relationships, and reduced social participation ([3], [5], [6]). These deficits may manifest as challenges in understanding sarcasm, irony, deception, and in predicting behavior based on others' mental states ([1], [2]).

Various tools have been developed to assess ToM in children and adolescents. Traditional tasks like the Sally and Anne task ([2]) and the Smarties task ([7]) measure first-order false beliefs, while the ToM Scale ([8]) and the ToM Task Battery ([9]) offer a developmental progression. More advanced tools, such as Happe's Strange Stories ([10]) and the advanced ToM test ([11]), evaluate higher-order ToM skills like lying, sarcasm, and pretending. Recently developed measures like the ToM Battery ([12]) and the ToM Booklet Task ([13]) aim to provide more comprehensive assessments. In addition to these tasks, proxy-report questionnaires such as the Theory of Mind Inventory-2 (ToMI-2) ([9]) and the Everyday Mindreading Skills and Difficulties Scale (EMSD) have been created to evaluate ToM in everyday contexts. Completed by caregivers, these tools complement task-based measures by offering insights into real-world ToM functioning. Notably, the ToMI-2 also includes a self-report version suitable for adolescents and adults with typical or near-typical language skills, allowing individuals to reflect on their own ToM abilities.

Despite its advantages, including reduced cognitive and linguistic demands and real-life contextual relevance, many ToM assessments, including the ToMI-2, face limitations such as underrepresentation of developmental stages and insufficient cross-cultural adaptation ([14]). Recent evidence also highlights the co-occurrence of motor impairments with social and behavioral symptoms in ASD, suggesting that assessments of social cognition may benefit from broader neurodevelopmental evaluation ([15]). The ToMI-2 helps address these issues by offering a broader, ecologically valid evaluation of ToM performance. Validated in multiple languages such as English, Dutch, and Chinese ([9], [16], [17]), the ToMI-2 has not yet been adapted for Persian-speaking populations. Persian, spoken by over 100 million people in Iran, Afghanistan, and Tajikistan ([18]), lacks a validated version of this tool. Developing a ToMI-2-P is essential for accurately assessing ToM abilities in adolescents with ASD within this linguistic and cultural context.

2. Objectives

Accordingly, the objectives encompassed the cross-cultural translation and adaptation of the ToMI-2-P and the rigorous assessment of its validity and reliability metrics in adolescents with ASD.

3. Methods

3.1. Study 1: Translation and Cultural Adaptation of the Theory of Mind Inventory-2

3.1.1. Participants

An expert panel of 20 professionals — each holding a PhD in neuroscience or occupational therapy and having over five years of experience with neurodevelopmental disorders and ASD — was recruited using expert sampling. The panel included 12 occupational therapists, 6 neuropsychologists, and 2 developmental pediatricians to ensure comprehensive evaluation of cognitive, social, and developmental aspects.

3.1.2. Measure

The ToMI-2 is a 60-item caregiver-report questionnaire that assesses ToM across three developmental subscales: Early, basic, and advanced. Responses use a 5-point Likert scale (1 = strongly disagree to 5 = strongly agree). Subscale averages over 15 indicate typical development, 5 - 15 suggest possible issues, and below 5 reflect developmental concerns. The original ToMI-2 has demonstrated sound psychometric properties in American populations ([19]).

3.1.3. Procedures

The translation and cultural adaptation of the ToMI-2 followed Beaton et al.'s guidelines ([20]). Initially, the inventory was forward-translated into Persian by a bilingual occupational therapist, then back-translated into English by a native speaker unfamiliar with the original. Discrepancies were resolved through synthesis and consensus among the research team.

3.1.4. Expert Panel Review

The resulting Persian version was evaluated by the expert panel for content relevance, clarity, simplicity, and cultural appropriateness. Panelists rated each item using Content Validity Index (CVI) and content validity ratio (CVR) metrics and provided qualitative feedback on potential cultural or linguistic ambiguities. Items not meeting thresholds for relevance (CVR > 0.42) or clarity (CVI > 0.79) were revised in discussion with the research team.

3.1.5. Pilot Study

A pilot study was conducted with a small group of Persian-speaking parents of adolescents with ASD to ensure item comprehensibility and cultural relevance. Feedback from parents was used to identify potential ambiguities, confusing wording, or culturally unfamiliar expressions. Items requiring modification were flagged for review, but specific item changes are reported in the Results section.

3.2. Study 2: Psychometric Evaluation of the Theory of Mind Inventory-2 for Persian Speakers

3.2.1. Participants

This study included 70 parents of adolescents with ASD aged 12 - 18 years, diagnosed based on DSM-5 criteria (4), with a minimum IQ of 70. Those with comorbid conditions such as ADHD, intellectual disability, or cerebral palsy were excluded. Participants were recruited from Parto Rehabilitation Center in Tehran. A control group of parents of typically developing adolescents was recruited from a local school and matched to the ASD group on demographic variables, excluding diagnosis.

3.2.2. Measures

Social Responsiveness Scale (SRS-2) was used in addition to the ToMI-2-P to assess ASD-related social difficulties. It has 65 items rated on a 4-point scale and produces *t*-scores, with higher scores reflecting more severe deficits (21). Its Persian version has been validated with strong psychometric indices: The CVI = 0.82, CVR = 0.86, intra-class correlation coefficient (ICC) = 0.80, and Cronbach's α = 0.93 (22).

3.2.3. Procedures

Participants provided written consent following an explanation of the study's aims. The ToMI-2-P and SRS-2 were administered to both groups (ASD and control) to examine convergent validity. Discriminative validity was evaluated by comparing ToMI-2-P scores between groups. Cronbach's α was used to evaluate internal consistency in both the ASD and control groups. Test-retest reliability was determined by the ICC after a subset of the ASD group only completed the ToMI-2-P again three weeks later, as the measure is primarily intended for clinical application and monitoring in the ASD population. All procedures were approved by the Research Ethics Committee of the Institute for Higher Education in Cognitive Sciences (ethics code: IR.UT.IRICSS.REC.1404.010).

3.3. Statistical Analysis

Descriptive statistics were calculated for demographic and psychometric variables. Face validity was assessed based on participants' ratings of item importance, with acceptable items showing an impact score above 1.5 (23). Content validity was evaluated using Lawshe's CVR method, accepting items with CVR > 0.42 (24), and CVI values based on the Waltz and Bausell criteria, with CVI > 0.79 considered satisfactory (25, 26). To assess reliability, internal consistency was measured with Cronbach's α (acceptable if α ≥ 0.70), and test-retest reliability was examined using the ICC. To assess convergent validity, Pearson's correlation was used to compare ToMI-2-P scores with the total scores of the SRS-2. Discriminative validity was assessed using multivariate analysis of variance (MANOVA) to compare the ASD and control groups. Assumptions of normality, homogeneity of variances, and equality of variance-covariance matrices were checked using Shapiro-Wilk, Levene's, and Box's M tests. Cohen's d was used to calculate effect sizes, where 0.2, 0.5, and 0.8 represent small, medium, and large effects, respectively (27). Data analysis was carried out in SPSS version 26, considering results significant if p was less than 0.05.

4. Results

4.1. Face and Content Validity

The face and content validity of the ToMI-2-P were evaluated through participant reviews. Table 1 displays the results, including the CVR, Content Validity indices for Relevance (CVIR), Simplicity (CVIS), Clarity (CVIC), and impact scores.

The CVR scores ranged from 0.50 to 1.00, reflecting the proportion of participants who viewed each item as essential. Items scoring above 0.75 indicated strong agreement on relevance to ToM. The CVIR values ranged from 0.80 to 0.95, confirming item alignment with ToM constructs.

The CVIS and CVIC scores (0.80 - 1.00) demonstrated that items were perceived as both simple and clear. Impact scores ranged from 2.70 to 3.87, with all items exceeding the 1.5 threshold, supporting their retention. High-performing items included: "My child distinguishes between teasing done playfully by a friend and mean teasing by a bully" and "My child notices when a listener is not interested." Items with lower

scores may need further review for conceptual alignment and clarity.

4.2. Pilot Study Adaptations

Based on the pilot study conducted with a small group of Persian-speaking parents of adolescents with ASD, several items of the ToMI-2-P were slightly modified to enhance cultural and linguistic clarity while preserving their theoretical intent. Item 2, which originally involved sarcasm, was adapted to a culturally familiar joking expression about the weather. Item 7, referring to the meaning of 'think', was clarified to indicate considering ideas or possibilities in the mind. Item 30, describing the hand-as-bird metaphor, was revised to explicitly indicate that it reflects pretend play rather than reality. Item 36, which involved wordplay or riddle humor, was generalized to encompass understanding riddles or jokes with double meanings that are culturally appropriate in Persian.

4.3. Discriminant Validity

The MANOVA assumptions were satisfied ($P > 0.05$). Results showed a significant group effect on ToMI-2-P scores and all subscales [Wilks' Lambda = 0.180, $F(3, 86)$ = 138.02, $P < 0.001$, $\eta^2 = 0.820$]. Adolescents with ASD scored significantly lower than controls on Early [$F(1, 88)$ = 20.75, $P < 0.001$, $\eta^2 = 0.19$], Basic [$F(1, 88)$ = 85.03, $P < 0.001$, $\eta^2 = 0.49$], and Advanced [$F(1, 88)$ = 275.27, $P < 0.001$, $\eta^2 = 0.75$] subscales, confirming the scale's discriminant validity (Table 2).

4.4. Convergent Validity

The ToMI-2-P total scores showed a significant negative correlation with SRS-2 total scores ($R = -0.43$, $P < 0.01$), indicating that higher ToM abilities were associated with fewer social responsiveness difficulties. Similar negative correlations were observed between corresponding ToMI-2-P and SRS-2 subscales, supporting convergent validity. Importantly, this pattern was consistent across both the ASD and typically developing control groups (Table 3), confirming the robustness of these relationships across populations.

4.5. Reliability

Cronbach's α values demonstrated adequate internal consistency for the total scale and subscales (α = 0.70 - 0.77). For the typically developing control group, internal consistency was also good to excellent (α = 0.75 - 0.85), supporting the structural reliability of the tool

across both populations. Test-retest reliability, assessed via the ICC in the ASD group, ranged from 0.81 to 0.87, showing good temporal stability (Table 4).

5. Discussion

This study was conducted to assess the reliability and validity of the ToMI-2-P. The findings provide evidence for the tool's applicability in a Persian-speaking context, while also highlighting areas requiring further refinement to address cross-cultural nuances. When scrutinizing the results pertaining to face and content validity, alignment is discerned with antecedent research by Hutchins et al. and Lerner et al. Commensurate with Hutchins et al., the current study observes the salient CVR values for items within the ToMI-2, such as "My child notices when a listener is not interested" and "My child distinguishes between teasing done playfully by a friend and mean teasing by a bully," substantiating the transcultural and linguistic pertinence of these items across Western and Persian-speaking demographics (9, 16).

Furthermore, confluence between the study's findings on item simplicity and clarity, appraised through the CVIS and CVIC scores, is mirrored by the emphasis placed on lucid and comprehensive phrasing in the ToM assessments as prescribed by Richardson et al. High CVIS and CVIC scores attributed to items such as "My child understands that people can lie on purpose to fool others" and "My child understands the word "know"" are congruent with Richardson et al.'s counsel to incorporate items that adeptly convey intricate cognitive constructs to heterogeneous populations, amplifying the centrality of linguistic precision and ambiguity mitigation in ToM assessment items (13).

In the sphere of item selection and significance, the present research's application of impact scores echoes strategies undertaken by Chen et al. and Hutchins et al. to prioritize items of heightened theoretical import and participant resonance. The discerned parallelism between items demonstrating substantial impact scores in this study, such as "My child notices when others are sad" and "My child notices when others are angry," with those underscored in Chen et al.'s Chinese rendition and Hutchins et al.'s original version accentuates the semblance of cross-cultural universality in acknowledging the prominence of these emotions as cardinal components of ToM competence (9, 17).

Nevertheless, a noteworthy deviation from precedent scholarship surfaces in the study's identification of select items possessing diminished Validity indices and impact scores, underscoring conceivable intricacies in cross-cultural adaptation. Evidently, the item "If I say,

Table 2. Discriminant Validity of the Theory of Mind Inventory-2 for Persian Speakers [a]

Variables	Group		Test Statistic F2,177	P-Value	Effect Size
	ASD (N = 30)	Control (N = 60)			
ToMI-2-P	12.51 ± 2.01	16.98 ± 1.40	150.30	< 0.001	0.63
Early	15.64 ± 1.25	16.87 ± 1.18	20.75	< 0.001	0.19
Basic	13.46 ± 2.51	17.33 ± 1.46	85.03	< 0.001	0.49
Advanced	8.45 ± 2.79	16.74 ± 1.89	275.27	< 0.001	0.75

Abbreviations: ASD, autism spectrum disorder; ToMI-2-P, Theory of Mind Inventory-2 for Persian speakers.
[a] Values are expressed as mean ± SD.

Table 3. Pearson Correlations Between Theory of Mind Inventory-2 for Persian Speakers and Social Responsiveness Scale-2 in Adolescents with Autism Spectrum Disorder (N = 70) and Controls (N = 70)

Variables	ToMI-2-P Total	Early	Basic	Advanced	SRS-2 Total	Awareness	Cognition	Communication	Motivation	Restricted Interests and Repetitive Behavior
ToMI-2-P total										
ASD	1	-	-	-	-	-	-	-	-	-
Control	1	-	-	-	-	-	-	-	-	-
Early										
ASD	0.64 [a]	1	-	-	-	-	-	-	-	-
Control	0.69 [a]	1	-	-	-	-	-	-	-	-
Basic										
ASD	0.86 [a]	0.37 [a]	1	-	-	-	-	-	-	-
Control	0.89 [a]	0.41 [a]	1	-	-	-	-	-	-	-
Advanced										
ASD	0.85 [a]	0.32 [a]	0.62 [a]	1	-	-	-	-	-	-
Control	0.88 [a]	0.33 [a]	0.59 [a]	1	-	-	-	-	-	-
SRS-2 total										
ASD	-0.43 [a]	-0.25 [a]	-0.36 [a]	-0.40 [a]	1	-	-	-	-	-
Control	-0.47 [a]	-0.28 [a]	-0.38 [a]	-0.42 [a]	1	-	-	-	-	-
Awareness										
ASD	-0.33 [a]	0.01	-0.28 [a]	-0.46 [a]	0.85 [a]	1	-	-	-	-
Control	-0.36 [a]	-0.03	-0.31 [a]	-0.50 [a]	0.87 [a]	1	-	-	-	-
Cognition										
ASD	-0.32 [a]	-0.28 [a]	-0.24 [b]	-0.28 [a]	0.84 [a]	0.59 [a]	1	-	-	-
Control	-0.35 [a]	-0.25 [b]	-0.27 [b]	-0.32 [a]	0.85 [a]	0.61 [a]	1	-	-	-
Communication										
ASD	-0.38 [a]	-0.11	-0.34 [a]	-0.42 [a]	0.86 [a]	0.90 [a]	0.53 [a]	1	-	-
Control	-0.41 [a]	-0.15	-0.36 [a]	-0.46 [a]	0.88 [a]	0.91 [a]	0.55 [a]	1	-	-
Motivation										
ASD	-0.27 [a]	-0.23 [b]	-0.19	-0.24 [b]	0.88 [a]	0.68 [a]	0.80 [a]	0.66 [a]	1	-
Control	-0.30 [a]	-0.20	-0.22 [b]	-0.26 [b]	0.90 [a]	0.70 [a]	0.83 [a]	0.68 [a]	1	-
Restricted interests and repetitive behavior										
ASD	-0.53 [a]	-0.39 [a]	-0.51 [a]	-0.37 [a]	0.81 [a]	0.54 [a]	0.64 [a]	0.62 [a]	0.65 [a]	1
Control	-0.55 [a]	-0.41 [a]	-0.54 [a]	-0.39 [a]	0.83 [a]	0.56 [a]	0.67 [a]	0.64 [a]	0.67 [a]	1

Abbreviations: ToMI-2-P, Theory of Mind Inventory-2 for Persian speakers; SRS-2, Social Responsiveness Scale-2; ASD, autism spectrum disorder.
[a] P < 0.05.
[b] P < 0.01.

'What is black, white, and 'read' all over? A newspaper!' my child understands the wordplay humor" evinced

lower Validity indices, warranting the consideration of linguistic idiosyncrasies, cultural context, and humor

Table 4. Reliability of the Theory of Mind Inventory-2 for Persian Speakers

| Variables | Cronbach's α Group | | ICC (ASD Group) | 95% CI (Lower - Upper) | P-Value |
	ASD	Control			
ToMI-2-P total	0.75	0.85	0.87	0.79 - 0.95	< 0.001
Early	0.77	0.78	0.83	0.75 - 0.94	< 0.001
Basic	0.70	0.75	0.81	0.73 - 0.92	< 0.001
Advanced	0.73	0.80	0.85	0.77 - 0.93	< 0.001

Abbreviations: ASD, autism spectrum disorder; ICC, intra-class correlation coefficient; ToMI-2-P, Theory of Mind Inventory-2 for Persian speakers.

interpretation during the process of adapting assessments across distinct languages and cultures (17). Moreover, several items with lower CVR scores (e.g., 0.50) may reflect partial conceptual misalignment or limited cultural familiarity among Persian-speaking parents. Items involving sarcasm, figurative expressions, or abstract reasoning (e.g., irony, idioms, or metaphorical play) might be less salient or differently interpreted in Persian sociolinguistic contexts. This is consistent with findings from previous cross-cultural adaptations, where abstract or idiomatic ToM items required modification to achieve semantic equivalence (9, 17). Such results highlight that while core ToM concepts are universal, their linguistic and contextual representations vary across cultures. Consequently, items with lower CVR scores should be re-evaluated through cognitive interviewing and pilot testing to ensure both theoretical and cultural relevance in Persian contexts.

In addition to face and content validity, convergent validity findings further support the applicability of the Persian ToMI-2. The significant correlations between the ToMI-2 scores and external measures of social cognition are consistent with the results reported by Hutchins et al. and Chen et al. (9, 17), where similar associations were observed in the Spanish and Chinese adaptations of the tool. These results indicate that the ToMI-2-P effectively captures ToM constructs that are aligned with related social-cognitive competencies, further corroborating its validity. Importantly, convergent validity was demonstrated in both the ASD and typically developing control groups, indicating that the ToMI-2-P maintains conceptual coherence across distinct populations. Although the correlation patterns were consistent, adolescents with ASD showed generally weaker associations between ToMI-2-P and SRS-2 subscales compared to controls. This attenuation may reflect variability in social-cognitive functioning and heterogeneity in symptom expression within the ASD population. The presence of significant, albeit weaker, correlations in the ASD group nonetheless reinforces the measure's sensitivity to individual differences in ToM

and its related constructs. These findings support the ToMI-2-P's discriminative capacity while underscoring the importance of examining group-specific patterns when interpreting convergent validity.

The moderate correlation between ToMI-2-P and SRS-2 total scores (R = -0.43) suggests that while both measures assess overlapping domains of social understanding, they also capture distinct constructs. The ToMI-2-P focuses specifically on inferential and representational aspects of social cognition (how individuals attribute mental states to others), whereas the SRS-2 encompasses broader dimensions of social behavior and interactional challenges (28). Therefore, a strong but not excessive correlation aligns with theoretical expectations, supporting discriminant validity between cognitive ToM and observable social behaviors. Moreover, variability in parents' insight into their adolescents' internal mental-state reasoning versus externally observable behaviors may further attenuate this relationship (9, 29). Cultural norms regarding emotion expression and social reciprocity in Persian-speaking families might also contribute to this moderate correlation, as social communication difficulties may be perceived differently across contexts (30). Overall, these results affirm that the ToMI-2-P captures a theoretically narrower, cognitively grounded construct distinct from — but related to — the broader social impairments assessed by the SRS-2.

The reliability findings further substantiate the robustness of the ToMI-2-P. Cronbach's α coefficients indicated good to excellent internal consistency in both groups, demonstrating structural reliability across clinical and non-clinical samples. The slightly higher internal consistency observed in the control group suggests that typically developing adolescents may interpret and respond to ToMI-2-P items more uniformly than those with ASD, possibly due to more consistent language comprehension and social experience. Test-retest reliability values in the ASD group were strong, confirming good temporal stability for clinical monitoring. The High Reliability indices reported in this study, including intra-rater and inter-rater reliability, are

consistent with findings from other cross-cultural adaptations of ToMI-2. Studies by Hutchins et al. and Chen et al. reported similarly robust reliability for their Spanish and Chinese adaptations, highlighting the reliability of this tool across diverse languages and settings when rigorous standardization processes are applied (9, 17).

Despite these strengths, differences in item performance between this study and prior research warrant further exploration. Items with lower reliability or ambiguous interpretation, such as those involving advanced linguistic constructs, highlight potential areas for refinement. De Los Reyes et al. similarly emphasized the importance of iterative piloting to enhance item clarity and cultural fit, reinforcing the need for ongoing refinement in cross-cultural adaptations (31). In terms of discriminative validity, the ToMI-2-P successfully differentiated between children with typical development and those with ASD, consistent with findings from Hutchins et al. and Chen et al. (9, 17). However, subtle differences in item performance suggest potential cultural variations in how ToM is expressed and prioritized. These findings align with Miller et al., who noted that cultural norms significantly influence social and emotional behaviors, which may affect test performance (32).

A key limitation of this study is that participants were recruited from a single rehabilitation center in Tehran, which may introduce selection bias and limit generalizability. Families in such centers may differ from broader Persian-speaking populations in socioeconomic or clinical characteristics. Future studies should include participants from multiple regions and community settings to enhance representativeness and external validity. Overall, while the ToMI-2-P demonstrates strong psychometric properties, the observed variations in item performance compared to previous studies highlight the complexities of cross-cultural adaptation. Future research should incorporate mixed-methods approaches, including qualitative feedback from parents and caregivers, to further refine the tool and ensure its cultural responsiveness. Additionally, longitudinal studies could provide valuable insights into the developmental trajectory of ToMI-2-P, contributing to a more nuanced understanding of cultural influences on these competencies.

5.1. Conclusions

This study confirms the validity and reliability of the ToMI-2-P for adolescents with ASD. The results align with previous validations, demonstrating its effectiveness in measuring ToM and distinguishing between ASD and typically developing peers. Significant correlations with the SRS-2 support its role in assessing social cognition. Some items showed cultural sensitivity, highlighting the need for refinement. Future research should incorporate mixed-methods and longitudinal studies to enhance cultural alignment and applicability in clinical and research settings.

Footnotes

References

1. Baron-Cohen S. Theory of mind and autism: A review. In: Baron-Cohen S, editor. *International Review of Research in Mental Retardation*. **23**. Cambridge, USA: Academic Press; 2000. p. 169-84. https://doi.org/10.1016/s0074-7750(00)80010-5.

2. Baron-Cohen S, Leslie AM, Frith U. Does the autistic child have a "theory of mind" ? *Cognition*. 1985;**21**(1):37-46. https://doi.org/10.1016/0010-0277(85)90022-8.

3. O'Brien JC; Case-Smith J. *Occupational therapy for children and adolescents-e-book. Elsevier Health Sciences*. St. Louis, USA: Mosby; 2014.

4. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, USA: American Psychiatric Publishing; 2013. https://doi.org/10.1176/appi.books.9780890425596.

5. Orsmond GI, Shattuck PT, Cooper BP, Sterzing PR, Anderson KA. Social participation among young adults with an autism spectrum disorder. *J Autism Dev Disord*. 2013;**43**(11):2710-9. [PubMed ID: 23615687]. [PubMed Central ID: PMC3795788]. https://doi.org/10.1007/s10803-013-1833-8.

6. Yoder PJ, Bottema-Beutel K, Woynaroski T, Chandrasekhar R, Sandbank M. Social communication intervention effects vary by dependent variable type in preschoolers with autism spectrum disorders. *Evid Based Commun Assess Interv*. 2013;**7**(4):150-74. [PubMed ID: 25346776]. [PubMed Central ID: PMC4206946]. https://doi.org/10.1080/17489539.2014.917780.

7. Perner J, Frith U, Leslei AM, Leekam SR. Exploration of the Autistic Child's Theory of Mind: Knowledge, Belief, and Communication. *Child Develop*. 1989;**60**(3):689-700. https://doi.org/10.2307/1130734.

8. Wellman HM, Liu D. Scaling of theory-of-mind tasks. *Child Dev*. 2004;**75**(2):523-41. [PubMed ID: 15056204]. https://doi.org/10.1111/j.1467-8624.2004.00691.x.

9. Hutchins TL, Prelock PA, Bonazinga L. Psychometric evaluation of the Theory of Mind Inventory (ToMI): a study of typically developing children and children with autism spectrum disorder. *J Autism Dev Disord*. 2012;**42**(3):327-41. [PubMed ID: 21484516]. https://doi.org/10.1007/s10803-011-1244-7.

10. O'Hare AE, Bremner L, Nash M, Happe F, Pettigrew LM. A clinical assessment tool for advanced theory of mind performance in 5 to 12 year olds. *J Autism Dev Disord*. 2009;**39**(6):916-28. [PubMed ID: 19205858]. https://doi.org/10.1007/s10803-009-0699-2.

11. Scheeren AM, de Rosnay M, Koot HM, Begeer S. Rethinking theory of mind in high-functioning autism spectrum disorder. *J Child Psychol Psychiatry*. 2013;**54**(6):628-35. [PubMed ID: 23072222]. https://doi.org/10.1111/jcpp.12007.

12. Siller M, Swanson MR, Serlin G, George A. Internal State Language in the Storybook Narratives of Children with and without Autism Spectrum Disorder: Investigating Relations to Theory of Mind Abilities. *Res Autism Spectr Disord*. 2014;**8**(5):589-96. [PubMed ID: 24748899]. [PubMed Central ID: PMC3988534]. https://doi.org/10.1016/j.rasd.2014.02.002.

13. Richardson H, Lisandrelli G, Riobueno-Naylor A, Saxe R. Development of the social brain from age three to twelve years. *Nat Commun*. 2018;**9**(1):1027. [PubMed ID: 29531321]. [PubMed Central ID: PMC5847587]. https://doi.org/10.1038/s41467-018-03399-2.

14. Fu IN, Chen KL, Liu MR, Jiang DR, Hsieh CL, Lee SC. A systematic review of measures of theory of mind for children. *Develop Rev*. 2023;**67**. https://doi.org/10.1016/j.dr.2022.101061.

15. Shiri V, Hosseini A, Pishyareh E. Relationship Between Behavioral Symptoms of Autism Spectrum Disorder and Motor Development Impairments: A Systematic Review. *Arch Neurol Neurosci*. 2024;**11**(3). https://doi.org/10.5812/ans-143284.

16. Lerner MD, Calhoun CD, Mikami AY, De Los Reyes A. Understanding parent-child social informant discrepancy in youth with high functioning autism spectrum disorders. *J Autism Dev Disord*. 2012;**42**(12):2680-92. [PubMed ID: 22456819]. https://doi.org/10.1007/s10803-012-1525-9.

17. Chen KL, Jiang DR, Yu YT, Lee YC. Development and psychometric evidence of the Chinese Version of the Theory of Mind Inventory-2 (ToMI-2) in children with autism spectrum disorder. *Res Autism Spectr Disord*. 2023;**103**. https://doi.org/10.1016/j.rasd.2023.102132.

18. Gordon RG. Gordon R, editor. *Ethnologue: Languages of the World, fifteenth edition*. Dallas, USA: SIL International; 2005. 1272 p.

19. Hutchins TL. *Technical Manual for the Theory of Mind Inventory-2*. Burlington, USA: The Theory of Mind Inventory; 2016. Available from: https://www.theoryofmindinventory.com/.

20. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;**25**(24):3186-91. [PubMed ID: 11124735]. https://doi.org/10.1097/00007632-200012150-00014.

21. Constantino JN, Gruber CP. Social responsiveness scale: SRS-2. *J Psycho Asses*. 2012;**32**(4):365-9.

22. Mirzakhani Araghi N, Alizadeh Zarei M, Saei S, Dibajnia P. Psychometric properties of the Persian version of social responsiveness scale-second edition (SRS-2). *Adv Autism*. 2021;**8**(3):264-71. https://doi.org/10.1108/aia-06-2021-0027.

23. Vakili MM, Jahangiri N. Content Validity and Reliability of the Measurement Tools in Educational, Behavioral, and Health Sciences Research. *J Med Educ Dev*. 2018;**10**(28):106-18. https://doi.org/10.29252/edcj.10.28.106.

24. Lawshe CH. A Quantitative Approach to Content Validity1. *Pers Psychol*. 1975;**28**(4):563-75. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x.

25. Waltz CF, Bausell RB. *Nursing Research: Design, Statistics, and Computer Analysis*. Philadelphia, USA: F.A. Davis Company; 1981.

26. Munro BH. *Statistical Methods for Health Care Research*. Philadelphia, USA: Lippincott Williams & Wilkins; 2005.

27. Cohen J. Statistical Power Analysis. *Curr Dir Psychol Sci*. 1992;**1**(3):98-101. https://doi.org/10.1111/1467-8721.ep10768783.

28. Bruni TP. Test Review: Social Responsiveness Scale–Second Edition (SRS-2). *J Psychoeduc Assess*. 2014;**32**(4):365-9. https://doi.org/10.1177/0734282913517525.

29. Peterson CC, Garnett M, Kelly A, Attwood T. Everyday social and conversation applications of theory-of-mind understanding by children with autism-spectrum disorders or typical development. *Eur Child Adolesc Psychiatry*. 2009;**18**(2):105-15. [PubMed ID: 18810310]. https://doi.org/10.1007/s00787-008-0711-y.

30. Mirzakhani Araghi N, Pashazadeh Azari Z, Alizadeh Zarei M, Akbarzadeh Baghban A, Saei S, Yousefi Nodeh HR, et al. The Relationship Between Sensory Processing Patterns and Participation in Childhood Leisure and Play Activities: A Systematic Review and Meta-analysis. *Iran Rehab J*. 2023;**21**(1):17-38. https://doi.org/10.32598/irj.21.1.1277.2.

31. De Los Reyes A, Thomas SA, Goodman KL, Kundey SM. Principles underlying the use of multiple informants' reports. *Annu Rev Clin Psychol*. 2013;**9**:123-49. [PubMed ID: 23140332]. [PubMed Central ID: PMC4103654]. https://doi.org/10.1146/annurev-clinpsy-050212-185617.

32. Miller JG, Engelbrecht J, Wang Z, Tsudaka G. Toward greater cultural sensitivity in developmental psychology. *Appl Dev Sci*. 2020;**25**(1):1-13. https://doi.org/10.1080/10888691.2020.1789348.

Table 1. Face and Content Validity Scores of the Assessment of the Persian Version of the Theory of Mind Inventory-2: Results of Expert Review [a]

Item No.	Items	CVR	CVIR	CVIS	CVIC	Impact Score
1	My child knows that wearing a jacket usually means someone feels cold.	0.62	0.85	0.80	0.80	2.80
2	If it is raining and I say jokingly, 'What a wonderful sunny day,' my child understands that I do not really mean it	0.50	0.80	0.80	0.85	2.94
3	My child notices when someone requires help.	0.75	0.80	0.80	0.90	3.70
4	My child understands that if someone is afraid of darkness, they would avoid going into a dark room.	0.50	0.80	0.90	0.80	2.85
5	My child realizes that people can be mistaken about what others want.	0.50	0.80	0.85	0.85	2.76
6	My child knows that frowning means feeling different than smiling.	0.75	0.80	0.85	0.80	3.52
7	My child understands that 'thinking' means considering ideas or possibilities in the mind.	0.50	0.85	0.90	0.85	3.13
8	If I put my keys on a table and leave, and my child moves them to a drawer, my child understands I will first look for them on the table when I come back.	0.62	0.85	0.80	0.90	3.16
9	My child knows that to find out what is inside a box without a label, one needs to see or hear about it.	0.75	0.80	0.80	0.85	2.88
10	My child understands the word 'know'.	0.50	0.80	0.80	0.80	3.29
11	Sometimes things don't look as they really are. For example, when seeing a candle shaped like an apple, people might first think it's an apple but then realize it's a candle. My child understands it is the idea that changes, not the object.	0.50	0.90	1	0.80	2.94
12	If I showed my child a cereal box filled with cookies and asked what someone who hasn't seen inside would think is in the box, my child would answer cereal.	0.75	0.80	0.80	0.80	3.19
13	When I say "Let's hit the road!" my child understands I mean "Let's go!"	0.62	0.85	0.80	0.80	2.71
14	My child understands that people can lie on purpose to fool others.	0.75	0.80	0.80	0.80	2.88
15	My child knows that when someone guesses, they are less sure than when they say they know something.	0.75	0.80	1	0.90	2.94
16	My child understands that thinking about a cookie is not the same as smelling, eating, or sharing it.	0.50	0.80	0.90	0.85	2.85
17	My child understands that people may smile even when they are not happy.	0.75	0.85	0.80	0.80	3.20
18	My child distinguishes between teasing done playfully by a friend and mean teasing by a bully.	1	0.90	0.90	0.85	3.69
19	My child knows that people may not say what they really think to avoid hurting others' feelings.	062	0.80	0.85	0.80	2.88
20	My child understands the difference between lies and jokes.	0.75	0.85	0.80	0.90	3.69
21	My child knows that if two people look at the same object from different places, they may see it differently.	0.50	0.80	0.80	0.80	2.92
22	My child understands that people often think about what others are thinking.	0.50	0.80	0.80	0.80	3.76
23	My child understands that people often consider how others feel.	0.50	0.85	0.90	0.80	3.13
24	My child understands whether someone hurts another person intentionally or by accident.	1	0.95	0.90	0.85	3.05
25	My child notices when others feel happy.	1	0.80	0.80	0.80	3.62
26	My child can pretend that one thing is something else (for example, pretending a banana is a phone).	0.62	0.80	0.80	0.80	2.85
27	My child notices when a listener is not interested.	1	0.80	0.95	0.80	3.68
28	My child knows that when I show fear, it means something is unsafe or dangerous.	0.62	0.80	0.80	0.85	3.87
29	My child understands the word "if" used in hypothetical situations, like "If I had money, I would buy a house."	0.50	0.85	0.95	0.80	3.56
30	My child understands that pretending their hands are a bird is pretend play, not real.	0.62	0.85	0.85	0.80	2.70
31	My child knows how to invent stories to get what they want.	0.62	0.85	0.80	0.90	3.56
32	My child notices when others feel disgusted.	1	0.90	0.95	0.80	2.96
33	My child understands that when someone makes a promise, they are expected to keep it.	0.62	0.80	0.80	0.80	3.42
34	My child can empathize and understand how others feel by imagining their perspective.	1	0.90	0.90	0.80	3.02
35	My child knows that when someone shares a secret, it should not be told to others.	062	0.85	0.80	0.80	3.10
36	If I tell a riddle or joke where words have double meanings, my child understands the humor.	0.50	0.80	0.95	0.85	2.79
37	My child is able to show me objects.	0.50	0.80	0.80	0.80	3.10
38	My child pays attention when I show something to them.	0.50	0.80	0.80	0.80	2.88
39	My child understands the meaning of the word "believe".	0.50	0.80	0.80	0.80	2.93
40	My child understands that liking someone can influence how we interpret their behavior positively, while disliking someone can cause more negative interpretations.	0.75	0.85	0.80	0.80	2.82
41	My child knows that two people can see the same picture but interpret it differently, like one seeing a rabbit and another seeing a duck.	0.75	0.80	1	0.80	2.76
42	My child understands that Bruce, who is mean, is more likely than John, who is kind, to behave hurtfully.	0.50	0.80	0.80	0.95	3.56
43	If I look up and stare at the sky, my child will also look up to see what I'm looking at.	0.50	0.80	0.85	0.80	2.94
44	If my child sees a strange new object, they look to me to check my reaction before touching it.	0.50	0.85	0.80	0.85	2.76
45	My child talks differently to young children than to adults (for example, using simpler words or a higher pitch).	1	0.90	0.90	0.85	3.48

Item No.	Items	CVR	CVIR	CVIS	CVIC	Impact Score
46	My child understands that it is possible to feel two opposite emotions at the same time, like being sad a pet died but happy it's no longer suffering.	0.50	0.80	0.80	1	3.10
47	My child knows that an unfamiliar adult can make good guesses about their likes and dislikes, such as not liking to clean their room.	0.62	0.85	0.80	0.80	2.85
48	My child notices when others are sad.	1	0.80	0.80	0.90	3.42
49	My child notices when others are angry.	1	0.85	0.90	0.80	3.42
50	My child notices when others are scared.	1	0.80	0.80	0.85	3.42
51	My child notices when others are surprised.	0.88	0.90	0.80	0.90	3.19
52	My child notices when others feel embarrassed.	1	0.80	0.80	0.85	3.42
53	My child understands the meaning of the word "need".	0.62	0.80	0.90	0.90	2.85
54	My child understands the meaning of the word "want".	0.62	0.85	0.80	1	2.94
55	My child notices when others feel guilty.	0.62	0.80	0.80	0.80	2.88
56	My child can recognize and reflect accurately on their own emotions.	0.50	0.80	0.80	0.80	3.19
57	My child can predict how they will feel in future situations to plan accordingly (for example, bringing a favorite blanket when staying overnight away from home).	0.50	0.80	0.80	0.80	2.94
58	My child understands people's thoughts and feelings based on the context (for example, crying from losing a game is different from crying after winning an award).	0.62	0.85	0.85	0.90	3.48
59	My child knows that people feel happy when they get what they want.	0.75	0.80	0.80	0.90	3.42
60	My child understands that beliefs can influence emotions (for example, Patty is happy because she believes she will win an award).	1	0.90	0.85	0.80	3.69

Abbreviations: CVR, content validity ratio; CVIR, Content Validity indices for Relevance; CVIS, Content Validity indices for Simplicity; CVIC, Content Validity indices for Clarity.

[a] These items are back-translated from Persian to English and have been verified by the original developer of the Theory of Mind Inventory-2 (ToMI-2).