

AN ECOLOGICAL APPROACH TO THEORY-OF-MIND MEASUREMENT: CREATION OF THE EV-TOMI FROM OPEN-ENDED REPORTS

Famira Racy^{1*}, Makayla Vermette^{2*}, Jennifer Schorn³, & Alain Morin⁴

¹Inner Speech Research Lab, Mount Royal University (Canada)

²Inner Speech Research Lab, MRU; University of Windsor Psychology Graduate Student (Canada)

³Inner Speech Research Lab, MRU Psychology Undergraduate Student (Canada)

⁴Department of Psychology, Inner Speech Research Lab (PI), MRU (Canada)

*Equal authorship

Abstract

Theory-of-Mind (ToM), or mentalizing about what other people might be thinking about, is an important part of conscious experience that facilitates social cognition and navigation of our perceived worlds (e.g., predicting other people's thoughts and behaviours). In a 2023 study, our research group began exploring Theory-of-Mind in relation to inner speech, and we found that the selected ToM measures left something to be desired. For example, measures claiming to capture ToM had items that seemed to describe understanding of one's own time perception or one's own episodic memory rather than thinking about other people's mentalizations. Furthermore, existing ToM questionnaires are typically based on *a priori* notions of what researchers think ToM is, for example, as informed by literature reviews and judged by a panel of experts. To fill this gap, and in seeking ecological validity for ToM measurement, our team took an open-format approach to ask Canadian students, "if you are trying to infer what other people are thinking or experiencing, what comes into your mind?" We have used these responses to create the "Ecologically Valid Theory-of-Mind Inventory" (EV-ToMI). Here we present preliminary results of endorsement, reliability, and validity of this measure, and relationships with other self-processes.

Keywords: Theory-of-Mind (ToM), inner speech, self-processes, self-awareness, measurement.

1. Introduction

Theory-of-Mind (ToM) is the ability to attribute mental states such as beliefs, desires, and intentions to oneself and others. ToM is essential for social interaction, empathy, and moral reasoning (Premack & Woodruff, 1978). In essence, having an accurate ToM is part of one's internal life and knowing the self, as well as of paramount importance to navigating the social world. ToM facilitates smooth social interactions, predicting others' behaviours, helping behaviours like cooperation, empathy, deception use and detection, and avoidance of others in the case of threat (Krych-Appelbaum, Law, Jones, Barnacz, Johnson, & Keenan, 2007).

However, most of the qualitative research done on ToM is with clinical samples. Work has been done on ToM and patients experiencing anorexia nervosa, and autism (e.g., Frith & Happé, 1999; Sedgewick, Leppanen, Goh, Hayward, Happé, & Tchanturia, 2019; Holt, Bounekhla, Welch, & Polatajko, 2022), and patients experiencing dementia (e.g., Giovagnoli, Bell, Erbetta, Paterlini, & Bugiani, 2019).

Some ToM studies attempting ecological validity, or real world contexts in questionnaire development, include case-studies (2-5-year-olds; e.g., Bosco, Gabbatore, & Tirassa, 2014), and experience sampling (predefined notions, small samples, e.g., Bryant, Coffey, Povinelli, & Pruett Jr., 2013; larger samples, narrow scope like empathy and emotion sharing, *a priori* notions (e.g., Depow et al., 2021).

1.1. The current study

Our Inner Speech Research Lab has been primarily focused on inner speech and self-related processes, which until recently, have not included ToM, a crucial part of the intricate web of the self, especially in light of the simulation view – knowing the self helps predict others' internal lives (Focquaert, Braeckman, & Platek, 2008). The current study begins to test our new ToM measure (EV-ToMI) attempting to address ecological validity. We also test our revised measure of inner speech (GISQ-R-30). We include a range of tools conceptualized to measure self-related processes in order to gain a preliminary

understanding of understudied relationships between ToM for self and others, self-talk, self-reflection, cognitive flexibility, self-rumination, self-absorption, and socially aversive traits.

1.2. Construction of the EV-ToMI

To consider ecological validity in creating the EV-ToMI, we thought it beneficial to ask participants from non-clinical samples, using every-day context, about their ToM use via an online LimeSurvey open-ended format (Racy, Vermette, & Morin, 2024; $N = 1969$ instances from 343 participants), without the constraints of a priori notions and predefined prompts. In essence we asked, what thoughts do you engage in when trying to understand other people's thoughts?

Bottom up thematic analysis, allowing themes to emerge, showed that uncommon themes in the literature were more commonly present in our results. Participants often reported using others' behaviours, contexts, attributes, and appearances to inform their own ToM about what others may be thinking, thinking about others' emotions, and thoughts about "me". They reported use of ToM to ascertain others' possible beliefs, intentions, goals, desires, and values less often than the above.

At least 3 items were constructed for each participant-based theme on the EV-ToMI. We also consulted the prominent literature on ToM and confirmed that themes present in that literature were also present in our item construction. The items were reviewed by 3 researchers (A.M., F.R., M.V.), making sure to consider and evaluate our own expertise and biases. The items were randomized and piloted. The resulting preliminary EV-ToMI tool is presented in the 2.1 Measures section.

1.3. Current hypotheses

Our current hypotheses are (H): The EV-ToMI will (H1) exhibit high reliability, (H2) will have weak to moderate positive concurrent validity (H3); will exhibit divergent validity using measures of Rumination, Short Dark Triad, and Self-Absorption Private (but not Public) subscale responses; (H4) the GISQ-R-30 will exhibit high reliability, and when compared to responses on the Self-Talk-Scale, will exhibit moderate concurrent validity; (H5) the relationships between ToM and other self-processes will be mixed, as there are so many varying components involved. There should also be some significant positive relationship between ToM measurement and adaptive self-processes, as knowing oneself is supposed to facilitate knowing others.

2. Methods

We recruited Mount Royal University undergraduate psychology students via SONA Systems which allowed online access to the Research Ethics Board (REB) approved survey package on LimeSurvey. Students received course credit for participating. We pilot tested the study before going live. All measures were counterbalanced. To clean the data, we removed participants with large amounts of incomplete responses, no consent, overall response times under 10 minutes, and who self-identified as receiving an ASD diagnosis. Analyses were conducted using SPSS Version 29.0.2.0 (20).

2.1. Measures

ToMI:SR-A. The 60-item Theory of Mind Inventory: Self Report – Adult (Hutchins, Lewis, Prelock, & Brien, 2020) is intended to assess social cognition in adults "at risk for" Autism Spectrum Disorder (ASD). Developed using clinical samples with non-clinical samples as a control or a standardization sample, this measure uses a total score including various sub-aspects such as Sympathy, Affective Empathy, Cognitive Empathy, Emotional Intelligence, and more. The ToMI:SR-A has exhibited acceptable reliability and validity.

TEIQUE-SF. The 30-item Trait Emotional Intelligence Questionnaire – Short Form is intended to measure global trait emotional intelligence in non-clinical samples over a range of facets associated with personality (Petrides, 2009) to obtain a global score. Although it is possible to obtain scores for factors of Well-being, Self-control, Emotionality, and Sociability, using a total score is more reliable.

MentS. The 28-item Mentalization Scale (Dimitrijević, Hanak, Dimitrijević, & Marjanović, 2017) is intended to measure the degree to which participants from clinical and non-clinical samples agree with statements about their mentalization, be it for the self or about others, and uses a total score as well as subscale scores for the factors of Self-Related Mentalization, Other-Related Mentalization, and Motivation to Mentalize. The subscales exhibit acceptable reliability for non-clinical samples and lower reliability for clinical samples.

EQ-10. The 10-item Emotional Quotient (Greenberg, Warrier, Allison, & Baron-Cohen, 2018) is intended to measure participant agreement with statements about use of cognitive and affective aspects of empathy (predicting others' thoughts, feelings, actions; having appropriate emotions in response to others' emotions) using a total score ($0 = \text{strongly disagree AND } 0 = \text{slightly disagree}$, $1 = \text{slightly agree}$,

2 = *strongly agree*) which is flipped for negatively worded items. Although the original point system is intended to emphasize empathy (not a lack of it), in the current study we simply used a Likert scale and reverse-scored negatively worded items, as we are not screening for ASD, but trying to understand empathy in general.

EV-ToMI. The 83-item Ecologically Valid Theory of Mind Inventory was developed to measure the degree to which participants agree with statements about their every-day use of social information to attribute others' mental states in general; it is based on a previous open-format investigation (Racy et al., 2024), construction of which is described above, and uses a total score (factor analysis underway). The current study is the first attempt at collecting reliability and validity information.

GISQ-R-30. The 30-item General Inner Speech Questionnaire-Revised (Racy, Kolesnikova, Vermette, & Morin, 2023) is intended to measure the degree to which participants agree with statements about their general inner speech use in every-day life. It is a revised version of the ecologically-driven GISQ (Racy, Morin, & Duhnych 2019). For now, a total score is used, although there is potential for factors of Self-Reinforcement, Cognition, Self-Criticism, Relationships, Activities, and Goals.

STS. The 16-item Self-Talk Scale (Brinthaupt, Hein, & Kramer, 2009) is intended to measure the degree of agreement with statements about their self-talk for self-regulatory aspects. It uses a total score and factors (Social Assessment, Self-Criticism, Self-Management, Self-Reinforcement). The STS is widely used (e.g., Hurlburt, 2019), especially in sport psychology, exhibiting strong reliability and moderate validity (e.g., Racy, Morin, & Hagerty, 2022).

RRQ. The 24-item Reflection and Rumination Questionnaire (Trapnell & Campbell, 1999) is intended to measure the degree to which participants agree with statements about self-attentiveness associated with personality characteristics of neuroticism and openness to experience. Only subscale scores for Reflection (healthy epistemic curiosity about the self) and Rumination (recurrent, impeding self-attentiveness) are used. The questionnaire exhibits acceptable reliability and validity.

SAS. The 17-item Self-Absorption Scale (McKenzie & Hoyle, 2008) is intended to measure how participants endorse statements about their "excessive, sustained, and rigid focus on the self". It uses a total score and subscale scores for Private Self-absorption (e.g., focus on your own thoughts) and Public Self-absorption (e.g., focus on what others think of you), exhibiting acceptable reliability and validity.

SD3. The 27-item Short Dark Triad (Jones & Paulhus, 2014) is intended to measure the degree to which participants agree with statements about their use of socially aversive traits, using a total score and scores for factors (Machiavellianism, Narcissism, Psychopathy).

CFI. The 22-item Cognitive Flexibility Inventory (Dennis & Vander Wal, 2010) is meant to measure agreement with statements about abilities to question their maladaptive thoughts and replace them with adaptive thoughts. With a total score and subscale scores derived from the factors Alternatives (i.e., generate alternative explanations) and Control (i.e., perceive difficult situations as controllable), it exhibits acceptable internal consistency, test-retest reliability, convergent and concurrent validity.

3. Results and discussion

After data cleaning, we were left with 396 participants (65 male, 8 preferred not to identify gender). The EV-ToMI distribution appeared normal and wide-ranging (range min = 13; range max = 414; $M = 259.45$; $SD = 62.32$; skew = $-.18$; kurtosis = 1.07). Participants endorsed items 19, 55, and 64 the least ($M = 2.39, 2.37, 2.38$ respectively) and items 1, 12, and 76 the most ($M = 3.70, 3.71, \text{ and } 3.70$ respectively; actual minimum = 0; actual maximum = 5). The lowest endorsements had to do with considering others' body shape, others' personal histories, and others' pasts, respectively. The highest endorsements were to do with considering others' emotions, others' honesty, and others' facial expressions respectively in formulating ToM.

The EV-ToMI exhibited acceptable reliability ($r = .98$) and weak to moderate concurrent validity (r ToMI:SR-A = $.27^{**}$; r MenS-S = $-.15^{**}$; r MenS-O = $.47^{**}$; r MenS-M = $.47^{**}$; r TEIQue-SF = $-.02$; r EQ10 = $.28^{**}$), perhaps because of the inclusion of non-traditional aspects of ToM found in the open-ended method used to generate the EV-ToMI items. Further, the EV-ToMI did not contain items about mentalizing about the self (like MenS-S and TEIQue-SF) which could account for those weak correlations. Our hypotheses about discriminant validity were mostly supported (r RRQ rumination = $.32^{**}$; r SASpub = $.36^{**}$; r SASpriv = $.19^{**}$; r SD3mach = $.09$; r SD3nar = $.02$; r SD3psych = $.01$).

The low and non-significant correlations with the SAS Private subscale and SD3 subscales (Dark Triad) adds support to claims that self-absorption (Joireman, Parrott, & Hammersla, 2002) and socially aversive traits (Moghaddam, Mojtabaie, & Bashardoust, 2022) may impede ToM. We did not fully expect the relationship with the RRQ Rumination subscale; we found rumination as a predictor of ToM in a past unpublished study, so we knew it was a possibility. Rumination should impede accurate ToM, but folks may ruminate about what others are thinking about anyway, possibly accounting for this relationship.

The revised GISQ-R-30 exhibited acceptable reliability ($r = .96$) and moderate concurrent validity (r STSscrit = .57**; r STSsreinforce = .47**; r STSsmanage = .51**; r STSsocassess = .50**), supporting our hypotheses, and similar to our past results with the full GISQ (Racy et al., 2019). This measure offers potential for researchers interested in an everyday inner speech measurement tool that considers ecological validity.

The EV-ToMI responses correlated significantly and positively with responses on measures of self-processes (r GISQ-R-30 = .44**; r STSscrit = .37**; r STSsreinforce = .25**; r STSsmanage = .33**; r STSsocassess = .37**; r RRQreflection = .31**; r CFIalt = .35**; r CFIcontrol = -.24**), supporting our hypotheses. These results suggest that talking to oneself, self-reflection, and cognitive flexibility probably have something to do with ToM – perhaps folks talk to themselves when trying to reflect on, be flexible with, or ascertain ToM, whether related to understanding and predicting one's own, or others' inner lives.

4. Strengths, limits, conclusions and future research

This study represents a unique attempt at considering ecological validity in designing the new EV-ToMI, allowing respondents to participate in the discovery of unprompted, underrepresented, everyday ToM aspects using a large, adult, non-clinical sample using a novel open-ended method.

Further, this study is a first attempt at understanding the validity and reliability of our GISQ-R-30, developed using a similar method. Limitations include relying on retrospection, and the potential for biased or uninformed ToM and inner speech conceptions. Further, the EV-ToMI includes some items that do not represent ToM *per se*, nor do they measure ToM “accuracy”, but instead represent participant thought processes in ascertaining ToM. However, staying true to what participants reported in the open-ended portion of the development, we kept the items.

This study offers preliminary insights into relationships between ToM and inner speech, as well as other self-processes. Our next steps include more in-depth correlational and regression analyses on the whole dataset, factor analysis to refine the EV-ToMI, and replications. For now, our results offer at least a glimpse into relationships between ToM and self-processes that may be useful in informing further research and exploration of educational, relationship, communication, workplace, personality, developmental, and sports contexts.

References

- Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34(2), 163-175. <https://doi.org/10.1023/B:JADD.0000022607.19833.00>
- Bosco, F. M., Gabbatore, I., & Tirassa, M. (2014). A broad assessment of theory of mind in adolescence: The complexity of mindreading. *Consciousness and Cognition*, 24, 84-97. <https://doi.org/10.1016/j.concog.2014.01.003>
- Brinthaup, T. M., Hein, M. B., & Kramer, T. E. (2009). The Self-Talk Scale: Development, factor analysis, and validation. *Journal of Personality Assessment*, 91(1), 82-92. <https://doi.org/10.1080/00223890802484498>
- Bryant, L., Coffey, A., Povinelli, D. J., & Pruett Jr., J. R. (2013). Theory of mind experience sampling in typical adults. *Consciousness and Cognition*, 22(3), 697-707. <https://doi.org/10.1016/j.concog.2013.04.005>
- Dennis, J.P., Vander Wal, J.S. (2010). The Cognitive Flexibility Inventory: Instrument development and estimates of reliability and validity. *Cognitive Therapy and Research*, 34, 241-253. <https://doi.org/10.1007/s10608-009-9276-4>
- Dimitrijević, A., Hanak, N., Altaras Dimitrijević, A., & Jolić Marjanović, Z. (2017). The Mentalization Scale (MentS): A self-report measure for the assessment of mentalizing capacity. *Journal of Personality Assessment*, 100(3), 268-280. <https://doi.org/10.1080/00223891.2017.1310730>
- Depow, G. J., Francis, Z., & Inzlicht, M. (2021). The experience of empathy in everyday life. *Psychological Science*, 32(8), 1198-1213. <https://doi.org/10.1177/0956797621995202>
- Focquaert, F., Braeckman, J., & Platek, S. M. (2008). An evolutionary cognitive neuroscience perspective on human self-awareness and Theory of Mind. *Philosophical Psychology*, 2, 47-68. <https://doi.org/10.1080/09515080701875156>
- Giovagnoli, A. R., Bell, B., Erbetta, A., Paterlini, C., & Bugiani, O. (2019). Analyzing theory of mind impairment in patients with behavioral variant frontotemporal dementia. *Neurological Sciences*, 40, 1893-1900. <https://doi.org/10.1007/s10072-019-03911-6>

- Greenberg, D. M., Warrier, V., Allison, C., & Baron-Cohen, S. (2018). Testing the empathizing–systemizing theory of sex differences and the extreme male brain theory of autism in half a million people. *Proceedings of the National Academy of Sciences*, 115(48), 12152-12157. <https://doi.org/10.1073/pnas.1811032115>
- Holt, A., Bounekhla, K., Welch, C., & Polatajko, H. (2022). “Unheard minds, again and again”: Autistic insider perspectives and theory of mind. *Disability and Rehabilitation*, 44(20), 5887-5897. <https://doi.org/10.1080/09638288.2021.1949052>
- Hutchins, T. L., Lewis, L., Prelock, P. A., & Brien, A. (2020). The development and preliminary psychometric evaluation of the Theory of Mind Inventory: Self Report—Adult (ToMI:SR-Adult). *Journal of Autism and Developmental Disorders*, 50(10), 3545-3557. <https://doi.org/10.1007/s10803-020-04654-6>
- Joireman, J. A., Parrott, L., & Hammersla, J. (2002). Empathy and the self-absorption paradox: Support for the distinction between self-rumination and self-reflection. *Self and Identity*, 1, 53-65. <https://doi.org/10.1080/152988602317232803>
- Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of Dark Personality Traits. *Assessment*, 21(1), 28-41. <https://doi.org/10.1177/1073191113514105>
- Krych-Appelbaum, M., Banzon Law, J., Jones, D., Barnacz, A., Johnson, A., & Keenan, J. P. (2007). "I think I know what you mean": The role of theory of mind in collaborative communication. *Interaction Studies*, 8(2), 267-280. <https://doi.org/10.1075/is.8.2.05kry>
- McKenzie, K. S., & Hoyle, R. H. (2008). The Self-Absorption Scale: Reliability and validity in non-clinical samples. *Personality and Individual Differences*, 45(8). <https://doi.org/10.1016/j.paid.2008.07.020>
- Moghaddam, M. Z., Mojtabaie, M., & Bashardoust, S. (2022). Study of Dark Triad personality and its relationship with warm empathy through the mediating role of affective-cognitive theory of mind in individuals with borderline personality disorder symptoms. *Journal of Positive School Psychology*, 17(65), 9640-9649. <https://doi.org/10.22034/jmpr.2022.14882>
- Petrides, K. V. (2009). Psychometric Properties of the Trait Emotional Intelligence Questionnaire (TEIQue). In J. Parker, D. Saklofske, & C. Stough (Eds.), *Assessing Emotional Intelligence. The Springer Series on Human Exceptionality*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-88370-0_5
- Premack, D., & Woodruff, G., (1978). Chimpanzee theory of mind: Part I. Perception of causality and purpose in the child and chimpanzee. *Behavioural and Brain Sciences*, 1(4), 616-629. <https://doi.org/10.1017/S0140525X00077050>
- Racy, F., Morin, A., & Duhnynch, C. (2019). Using a thought listing procedure to construct the General Inner Speech Questionnaire: An ecological approach. *Journal of Constructivist Psychology*, 33(4), 385-405. <https://doi.org/10.1080/10720537.2019.1633572>
- Racy, F., Morin, A., & Hagerty, J. (2022). Concurrent validity and reliability of representative inner speech questionnaires. *Psychological Reports*, 127(2). <https://doi.org/10.1177/00332941221109109>
- Racy, F., Kolesnikova, N., Vermette, M., & Morin, A., (2023). Exploration of the General Inner Speech Questionnaire – Revised 30 (GISQ-R-30). [talk and questionnaire brief; 2 files]. Talk presented at the *Science of Consciousness Conference*, Taormina, Italy. <http://dx.doi.org/10.13140/RG.2.2.10119.78243>
- Racy, F., Vermette, M., & Morin, A. (2024). Through the mind of the beholder: Ecological validity and student reports of Theory of Mind. [talk]. Presented at the *Science of Consciousness Conference*, Arizona, USA. <http://dx.doi.org/10.13140/RG.2.2.20637.52966>
- Sedgewick, F., Leppanen, J., Goh, F., Hayward, H., Happé, F., & Tchanturia, K. (2019). Similarities and differences in theory of mind responses of patients with anorexia nervosa with and without autistic features. *Frontiers in Psychiatry*, 10, 318. <https://doi.org/10.3389/fpsyt.2019.00318>
- Trapnell, P., & Campbell, J. (1999). Private self-consciousness and the five-factor model of personality: Distinguishing rumination from reflection. *Journal of Personality and Social Psychology*, 76(2), 284-304. <https://psycnet.apa.org/doi/10.1037/0022-3514.76.2.284>