Routledge
Taylor & Francis Group

Check for updates

The Brief Version of the Mentalization Scale (MentS-12): Evidence-Based Assessment of Mentalizing Capacity

Alberto Stefana[1] , Zorana Jolić Marjanović[2] and Aleksandar Dimitrijević[3]

[1]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy; [2]Department of Psychology, Faculty of Philosophy, University of Belgrade, Belgrade, Serbia; [3]Psychoanalyst in Private Practice, Berlin, Germany

ABSTRACT

Recognizing the need for a concise self-report measure of mentalizing capacity, we developed a 12-item iteration of the well-established Mentalization Scale (MentS). Using college student and community samples of Serbian adults ($N = 566$), we performed a precise selection of items and then examined the psychometric attributes of the shortened scale (MentS-12). The new scale maintains the original three-dimensional structure: self-related mentalization, other-related mentalization, and motivation to mentalize. MentS-12 proves to be both reliable and structurally consistent. To improve its utility in therapeutic contexts, we determined clinical change thresholds for both the complete and abbreviated forms. We hope that its feasibility stimulates the integration of the MentS-12 in longitudinal research projects and real-world clinical settings.

The construct of mentalization has garnered substantial attention within psychology and psychiatry in the preceding decades. The construct that was initially grounded in psychoanalytic theory (Fonagy & Bateman, 2010) is nowadays defined as "an individual's awareness of mental states in himself or herself and in other people, particularly in explaining their actions. It involves perceiving and interpreting the feelings, thoughts, beliefs, and wishes that explain what people do" (Fonagy & Bateman, 2019, p. 3). Mentalization is recognized as one of the vital facets of emotional and cognitive development since the lack of it has been associated with aggression, delinquency, substance abuse, and mental disorders (Luyten et al., 2020). This, along with potential benefits for treatment outcomes (Fonagy et al., 2019), has initiated meticulous attempts to adequately assess mentalization.

The most prominent among these attempts was made by Fonagy et al. (1998), who developed the Reflective Function Scale (RFS), an instrument that evaluates one's ability to reflect on attachment experiences as assessed by coding narratives elicited by the Adult Attachment Interview (George et al., 1996). As a performance-based measure with remarkable reliability, validity, and structural attributes (Taubner et al., 2013), the RFS has in the meantime become the gold standard in mentalization assessment. Nevertheless, its use is not without challenges, which are related to both the stringent training and certification prerequisites, and the time-consuming administration that comprises interviewing, transcription, and coding. Given these challenges, as well as the need for rapid mentalization assessment in larger cohorts of examinees, several self-report measures, including the Mentalization Questionnaire (MZQ; Hausberg et al., 2013),

the Reflective Functioning Questionnaire (RFQ; Fonagy et al., 2019), and the Mentalization Scale (MentS; Dimitrijević et al., 2018) were more recently conceived. Regarding the assessment process, it is important to remember that the first phase of an evidence-based approach to psychological diagnosis and treatment involves assessment to build a case formulation and a treatment plan (Hayes et al., 2020; Youngstrom et al., 2017). Once the target psychological entities or processes have been clarified and benchmarked, the role of assessment changes to monitoring processes and evaluating progress (Stefana, Vieta, et al., 2023; Youngstrom et al., 2015). Measuring the rate of change informs the clinician if they are making good progress or stuck; helping to identify the need for a modification of the treatment plan (Youngstrom et al., 2017). Therefore, an efficient assessment process will produce not only more accurate diagnoses but also better patient engagement and improved outcomes (Youngstrom & Van Meter, 2016).

In this paper, we will focus on the MentS by inspecting the possibility of introducing its shorter version. MentS is a 28-item self-report measure of mentalization in adults that includes three subscales: Self-related Mentalization (MentS-S), Other-related Mentalization (MentS-O), and Motivation to Mentalize (MentS-M). The first two subscales capture the ability to understand mental states of oneself (MentS-S) and others (MentS-O), while the last one (MentS-M) taps into mentalizing propensity. Since its publication, the MentS has achieved considerable international recognition and use, all of its translations and validations supporting its utility and validity in both community and clinical populations (Ahmadian & Ghamarani, 2021; Asgarizadeh et al., 2023; Jańczak, 2021; Matsuba et al., 2022; Müller et al., 2023;

Surim & Munhee, 2018; Törenli-Kaya et al., 2021; Wen et al., 2022). Notably, MentS has robust–moderate to high–positive correlations with the RFS (Richter et al., 2021) and effectively differentiates individuals with borderline personality disorder from healthy controls (Dimitrijević et al., 2018).

Although the administration of the MentS is already relatively time-efficient, the complexity of studying the therapeutic process typically necessitates the simultaneous administration of multiple measures (Wampold & Flückiger, 2023; Wampold & Imel, 2015). This is why an even swifter, yet reliable mentalization assessment is still called upon (Smith et al., 2000; Wampold & Flückiger, 2023).

Hence, the present study aimed to meaningfully shorten the assessment of mentalization by devising a brief version of MentS (henceforth *MentS-12*) and investigating its psychometric properties *vis-à-vis* the long version of the scale. Our general expectation was that the brief version will fully mimic the established properties of its mother scale in terms of reliability and structural properties (H_1). The selection of measures against which convergent and discriminant validity of the original version were tested was steered by several theoretical links. These links also served to formulate our current hypotheses. As a concept, mentalizing is understood to flourish in early secure attachment relationships and to be impaired or absent where attachment is disorganized (Fonagy & Target, 2008). This is why we expected that MentS-12 will have a negative association with attachment dimensions (H_2). Likewise, mentalization is understood as overlapping with a host of concepts: it is similar to empathy and social cognition, except that it also deals with one's own emotions (Bateman & Fonagy, 2019); it is like ability emotional intelligence and theory of mind except that it also deals with the affective component of human experiences (*ibid.*). At the same time, mentalization is inconsistent with alexithymia and impulsiveness (considered as constructs overlapping with that of emotional intelligence; Parker et al., 2001). Thus, we predicted that MentS-12 will have positive associations with empathy (H_3) and both trait (H_4) and ability emotional intelligence (H_5). Finally, we also tested whether the brief version of MentS would replicate the pattern and the strength of relationships with the Big Five personality dimensions and demographic data established for the original scale. We hypothesized that it would (H_6 and H_7, respectively).

To test our hypotheses, we draw on data originally used to assess the psychometric scrutiny of the 28-item version of the scale (Dimitrijević et al., 2018).

Methods

Participants and procedure

Original data were gathered on a non-clinical sample of 566 Serbian adults, consisting of workers (51%) and university students (49%). Within this sample, 58% were female, while the age range was between 18 and 61 years ($M = 30.0$, $SD = -12.1$). Employees were recruited and administered measures at their workplace during working hours, and students were recruited and administered measures during lecture hours. Examinees were briefed about the purpose of the study and further use of data, upon which they signed informed consent forms. Participation was fully voluntary, and no compensation was offered for the involvement in the study.

Measures

All participants completed the above described 28-item version of the MentS (Dimitrijević et al., 2018), and were administered measures of attachment and empathy. Additionally, data on personality, trait, and ability emotional intelligence were available for the working part of the sample.

Attachment

The 36-item Modified Experiences in Close Relationships-Revised (ECR-R; Fraley et al., 2000) examines attachment Avoidance (Av) and Anxiety (Ax) scores in close relationships in general. Responses are given on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Participants completed a Serbian adaptation of the modified ECR–R, which has previously been shown to reliably assess the two attachment dimensions (Hanak & Dimitrijević, 2013), which was confirmed in the current sample with alphas being .84 and .89 for Av and Ax, respectively.

Empathy

The Empathy Quotient (EQ; Baron-Cohen & Wheelwright, 2004) is a self-report measure of cognitive and affective aspects of empathy. Responses are given on a 4-point scale ranging from 1 (strongly disagree) to 4 (strongly agree). In the scoring procedure, nonempathic responses receive 0 points, while empathic responses are worth 1 or 2 points, which is why the overall score can range from 0 to 44. In this study, the Serbian translation of the 22-item version of EQ (Dimitrijević et al., 2012) was used, and its internal consistency in this sample was .60.

Personality

The 60-item NEO Five-Factor Inventory (NEO-FFI; McCrae & Costa, 2004) is a short version of the Revised NEO Personality Inventory (NEO PI-R), supplying scores for Neuroticism, Extraversion, Openness, Agreeableness, and Conscientiousness. Responses are given on a 5-point Likert scale ranging from 1 (completely incorrect) to 5 (completely correct). NEO family of inventories are widely used and exhibit generally good psychometric properties, which also pertain to their Serbian translation (e.g., Knežević et al., 2004). Alpha coefficients established in the current sample were in the .57–.80 range.

Trait emotional intelligence

The Trait Emotional Intelligence Questionnaire (TEIQue; Petrides, 2009) measures emotional intelligence defined as a constellation of emotional self-perceptions located at lower levels of personality hierarchies (Petrides et al., 2007).

It comprises 153 items rated on a 7-point scale, ranging from 1 (strongly disagree) to 7 (strongly agree), yielding one global and four-factor level scores of well-being, self-control, emotionality, and sociability. The Serbian version of TEIQue has been shown to have good reliability and factorial validity (Jolić Marjanović & Altaras Dimitrijević, 2014); alphas in the current sample ranged from .78 to .96.

Ability emotional intelligence

The MSCEIT v2.0 (Mayer et al., 2002) is a 141-item performance-based measure of ability emotional intelligence; its tasks examine capacities in the four areas of Perceiving Emotions, Using Emotions, Understanding Emotions, and Managing Emotions. The MSCEIT also yields an overall emotional intelligence score. Test items use a multiple-choice format, asking participants to select the best possible solution to the problem, or the rate-the-extent format, where examinees need to judge the appropriateness of several proposed options. Responses were scored by the test publisher (Multi-Health Systems), using the general consensus scoring method. The Serbian translation of the instrument exhibits good reliability and validity (Altaras Dimitrijević & Jolić Marjanović, 2010); internal consistencies in the current sample were between .67 and .92.

Statistical analyses

The development and validation of the MentS-12 adhered to the best practice guidelines for scale development (Smith et al., 2000; Stefana et al., 2024). First, data suitability for factor analysis was evaluated using the Kaiser-Meyer-Olkin test and the Bartlett test of sphericity. The exploratory factor analysis (EFA) was then performed to analyze the matrix of inter-item polychoric correlations. Factor loadings on hypothesized factors that were <.30 were considered unsatisfactory. Next, the item discrimination (α) and difficulty (β) parameters were calculated using the Item Response Theory (IRT) Graded Response Model (GRM) (Samejima, 2010). Discrimination parameter values were interpreted based on Baker and Kim (2017) as very low (.01–.34), low (.50–.64), moderate (.65–1.34), high (1.35–1.69), or very high (≥1.70). The selection of items for the brief version was based on those providing abundant information across a broad range of theta (θ). The fit of the final Standard threshold values were used as rough guidelines for the overall model fit: CFI ≥ 0.95, TLI ≥ 0.95, RMSEA ≤ 0.06, SRMR ≤ 0.08 (Hu & Bentler, 1999). To predict potential values for Cronbach's alpha, McDonald's omega (ω) total, and the content scope, the methodologies proposed by Smith et al. (2000) were applied. Smith et al.'s methodology for projecting the internal consistency of a short-form scale employs a proportional reduction factor, which is the quotient of the number of items in the short scale relative to the full scale. The projected internal consistency is derived by applying this factor to the known reliability coefficient of the long scale (it can be alpha, ω, or another reliability estimate). The calculation of projected internal consistency adjusts for the reduction in the number of items, acknowledging that internal consistency

typically decreases as the number of items on a scale decreases. For the projected correlation between scores on the full scale and the short scale, the formula takes the product of the reliability of the full scale and the projected reliability of the short scale. This resultant value provides an estimate of the expected correlation, reflecting the degree of linear relationship one could anticipate between the scores of the two scales, based on the assumption of parallel forms reliability. These statistical projections are integral to psychometrically validating the short-form scale, ensuring that, despite a reduced item set, the scale remains a reliable and valid measure of the construct it intends to assess, maintaining its psychometric fidelity to the original, longer instrument.

Internal consistency was primarily measured using ω (Revelle & Zinbarg, 2009), as it has been shown to be superior to alpha in assessing the reliability of short scales. Additionally, the average item correlation was used, as it is not dependent on the length of the scale (Streiner et al., 2015). To assess the accuracy of scores derived from the MentS-12 compared to the full 28-item version, Bland-Altman plots were employed (Bland & Altman, 1986). These plots provided an evaluation of score bias and helped at establishing "limits of agreement." The MentS-12 validity was further scrutinized by inspecting its correlations with theoretically tangent constructs and the Big Five personality. These correlations were compared to those obtained for the 28-item scale, with the Cohen's q (Cohen, 1992) as an operationalization of meaningful differences in correlations (Youngstrom et al., 2019). In this study, effect sizes ≤.32 are considered small, between .33 and .55 moderate, and ≥.56 large (Lipsey & Wilson, 1993).

All analyses were performed on an existing dataset and as such received approval from the IRB of the Department of Psychology, Faculty of Philosophy, University of Belgrade, Serbia (Protocol #2023-76). The study design and its analysis were not preregistered. Both the data and the analysis code that support the findings of this study are available from the corresponding author upon reasonable request.

Results

MentS factor structure

The Bartlett sphericity test of sphericity [$X^2(378) = 3778$, $p < .001$] and the Kaiser–Meyer–Olkin test (.843) maintained the suitability of our data obtained with the MentS for factor analysis. Working with the assumption that the MentS has three dimensions, EFA was set to extract a three-factor solution. Item loading divided into the hypothesized factors is given in Table 1, with median loadings being .55, .59, and .42. for MentS-O, MentS-S, and MentS-M, respectively.

MentS item discrimination and difficulty

The three MentS factors were separately analyzed using GRM IRT to obtain discrimination (α) and difficulty (β) parameters at the item level (Table 1). For the MentS-12, four items per subscale having moderate to very high

Table 1. EFA item loadings and GRM-model item parameters for MentS.

Item	Factor loadings			GRM-model item parameters				
	F1	F2	F3	a	$\beta1$	$\beta2$	$\beta3$	$\beta4$
MentS others								
2. When I make conclusions about other people's personality traits I carefully …	.36	.16	0.09	0.81	−7.25	−5.51	−2.90	0.43
3. I can recognize other people's feelings.	.60	.09	−0.03	1.65	−3.94	−3.04	−1.00	1.23
5. Usually I can recognize what makes people feel uneasy.	.63	.02	0.15	2.25	−3.18	−1.93	−0.43	1.44
6. I can sympathize with other people's feelings.	.43	−0.04	0.33	1.41	−2.54	−1.71	−0.26	1.50
10. I can make good predictions of other people's behavior when I know their …	.63	.07	−0.12	1.35	−4.17	−2.64	−0.64	1.77
12. Sometimes I can understand someone's feelings before s/he tells me anything.	.67	−0.08	0.02	1.74	−4.07	−2.28	−0.64	1.31
20. I can describe significant traits of people who are close to me with precision …	.49	.19	0.05	1.19	−3.95	−3.36	−1.29	1.00
23. People tell me that I understand them and give them sound advice.	.61	.02	0.08	1.56	−3.61	−2.77	−0.95	0.89
25. I can easily describe what I feel.	.41	.32	−0.09	0.82	−4.95	−2.86	−0.72	2.06
28. One of the most important things that children should learn is to express …	.31	−0.04	0.14	0.63	−7.13	−5.67	−2.58	0.66
MentS self								
8. When I get upset, I am not sure whether I am sad, afraid, or angry.	−0.02	.66	−0.08	1.57	−2.05	−0.78	0.18	1.34
11. Often I cannot explain, even to myself, why I did something.	.04	.66	−0.01	1.73	−1.82	−0.73	0.13	1.37
14. I do not want to find out something about myself that I will not like.	−0.02	.42	0.14	0.93	−3.68	−1.90	−0.49	1.23
18. I find it difficult to admit to myself that I am sad, hurt, or afraid.	−0.11	.52	0.21	1.15	−3.07	−1.61	−0.31	1.55
19. I do not like to think about my problems.	.01	.34	0.28	0.79	−4.07	−2.38	−0.61	1.77
21. I am often confused about my exact feelings.	.10	.79	−0.19	2.44	−1.78	−0.93	−0.03	1.17
22. It is difficult for me to find adequate words to express my feelings.	.15	.69	−0.06	1.85	−1.97	−0.89	−0.02	1.28
26. While people talk about their feelings and needs my thoughts often drift away.	.10	.39	0.36	0.99	−4.12	−2.76	−1.34	1.17
MentS motivation								
1. I find it important to understand reasons for my behavior.	.22	−0.04	.29	0.90	−4.91	−4.30	−2.60	0.33
4. I often think about other people and their behavior.	.11	−0.15	.61	1.46	−3.06	−1.55	−0.14	1.47
7. When someone annoys me, I try to understand why I react in that way.	.24	−0.11	.30	0.88	−4.36	−2.38	−0.76	2.06
9. I do not like to waste time trying to understand in detail other people's behavior.	−0.20	.24	.71	1.33	−2.45	−1.45	−0.13	1.51
13. I find it important to understand what happens in my relationships with …	.31	.03	.42	1.41	−3.81	−3.30	−1.88	0.23
15. To understand someone's behavior, we need to know her/his thoughts, …	.35	−0.02	.30	1.17	−4.52	−3.54	−1.69	0.58
16. I often talk about emotions with people that I am close to.	.21	−0.12	.42	1.11	−3.48	−1.83	−0.14	1.64
17. I like reading books and newspaper articles about psychological subjects.	.27	−0.04	.30	0.95	−3.29	−1.60	−0.11	1.48
24. I have always been interested in why people behave in certain ways.	.20	−0.14	.57	1.90	−2.79	−1.88	−0.69	0.72
27. Since we all depend on life circumstances, it is meaningless to think of other …	−0.17	.29	.70	1.19	−3.94	−2.21	−0.59	0.96

Note. The table presents the average loadings obtained from various factor extraction and rotation methods, conducted using the *EFA tools* package in R (Steiner & Grieder, 2020).

discrimination ability across a wide range of θ levels were selected: items 5, 10, 12, 23 for MentS-12-O; items 8, 11, 18, 21 for MentS-12-S; and items 13, 15, 16, 24 for MentS-12-M.

MentS-12 factor structure

CFA was run for each four-item factor separately and for the three factors together. One factor solution was tested for each subscale, indicating an excellent fit for the data: X^2 $(df=2)=.42$, CFI = 1.00, TLI = 1.01, RMSEA =.00 (90% CI [.00, .05]), and SRMR =.01 for the MentS-12-O; X^2 $(df=2)=1.32$, CFI = 1.00, TLI = 1.00, RMSEA =.00 (90% CI [.00, .08]), and SRMR =.01 for the MentS-12-M; and X^2 $(df=2)=4.24$, CFI =.99, TLI =.97, RMSEA =.05 (90% CI [.00, .11]), and SRMR =.02 for the MentS-12-M. In addition, the three-factor model for the MentS-12 had a good fit to the data: X^2 $(df=51)=111.03$, CFI =.95, TLI =.93, RMSEA =.05 (90% CI [.04, .06]), and SRMR =.04. Figure 1 shows the measurement model that presents a fully standardized solution using robust maximum likelihood estimation.

MentS-12 distributional properties, reliability, and score precision

Table 2 presents a full range of distributional and reliability indices for the MentS-12. For comparison, all reported properties are given for the MentS as well. On the MentS-12, the mean total score was 44.4 $(SD=5.47)$ out of the theoretical maximum of 60, pointing to a tendency to choose answers at the higher end of the 5-point Likert scale. The same trend was present at the subscale level, as well as for the full version of the scale. However, all reported Skewness/Kurtosis values (Table 2) were in an acceptable −2 to +2 range (George & Mallery, 2010), suggesting the data is fairly normal and suitable for psychometric purposes.

Table 2 also provides information on the standard error of measurement, standard error of differences, internal consistency, and average item correlations for both brief and full versions of the MentS. Of note, the estimated internal consistency of the full-length MentS-12 approached McDonald's ω of .71, while the observed reliability was .78, thus exceeding the projected value and approaching a high level.

Finally, Table 2 presents four clinical change benchmarks: critical changes (CC) at confidence levels of 90 and 95%, the minimally important difference (MID), and the minimum change for a reliable change (MCRC) based on the Jacobson (Jacobson & Truax, 1991) benchmark threshold. These clinical change benchmarks serve as practical tools for clinicians to evaluate whether changes in a patient's score are significant beyond measurement error and clinically meaningful. CC at 90% and 95% confidence levels indicates the minimum score increase or decrease needed to be confident that the change is not due to chance or measurement imprecision. MID reflects the smallest change in score that patients perceive as beneficial, which is crucial for determining
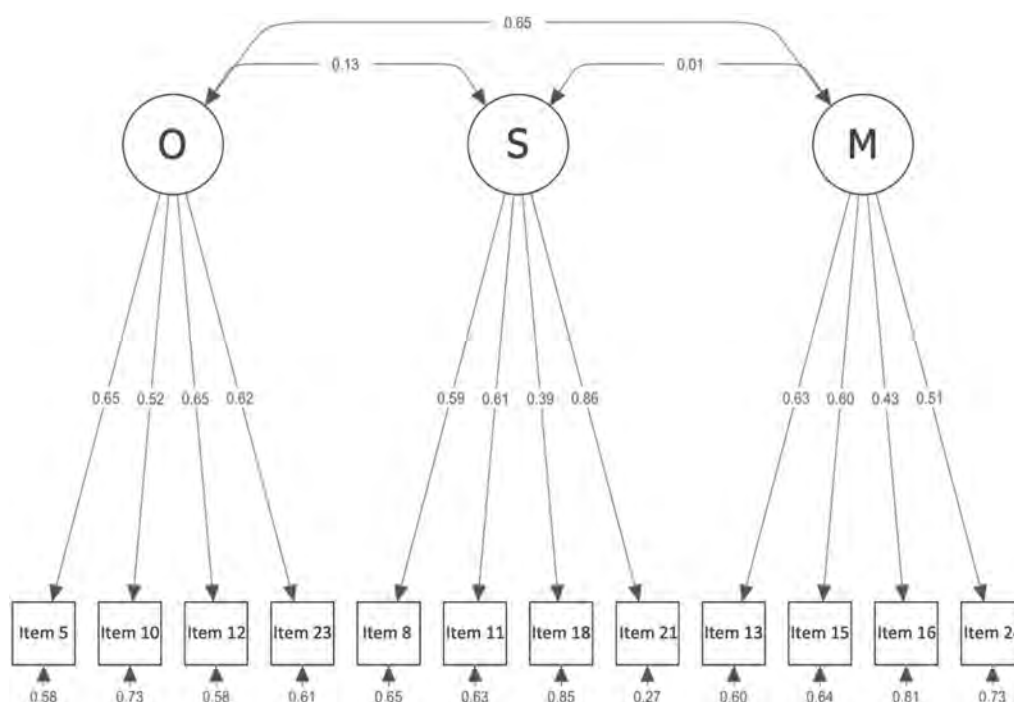
Figure 1. Measurement model from CFA (*N* = 566) presenting fully standardized solution using robust maximum likelihood estimation.

Table 2. Descriptive statistics, critical change benchmarks, and reliabilities for MentS-12 and MentS.

| Scale | Descriptive statistics | | | | | | | | Critical benchmarks | | | | Reliability | | | | |
| | | | | | | | | | | | | | Observed | | | Projected | |
	M (*SD*)	Mdn	Range	Sk	Ku	*SE*	SEm	SEd	90% CC	95% CC	MID	MCRC	α	ω	AIC	α	ω
MentS-12	44.4 (5.47)	44	30–60	.26	.00	.24	2.57	3.63	4.22	5.03	2.73	7.11	.68	.78	.15	.69	.71
MentS-12 self	13.5 (3.12)	14	5–20	−0.15	−0.28	.14	1.90	2.68	3.12	3.72	1.56	5.26	.62	.63	.29	.57	.59
MentS-12 others	15.2 (2.33)	15	8–20	−0.17	−0.10	.10	1.28	1.80	2.10	2.50	1.17	3.54	.69	.70	.37	.63	.63
MentS-12 motivation	15.8 (2.43)	16	4–20	−0.43	.74	.11	1.48	2.09	2.43	2.90	1.22	4.10	.61	.63	.28	.56	.57
MentS	104.6 (11.91)	104	68–140	.04	.00	.53	4.61	6.52	7.59	9.04	5.96	12.80	.84	.85	.16		
MentS self	27.9 (5.61)	28	10–40	−0.18	−0.27	.25	2.69	3.80	4.43	5.27	2.80	7.46	.77	.77	.29		
MentS others	38.8 (4.78)	39	18–50	−0.42	.61	.21	2.24	3.17	3.69	4.39	2.39	6.21	.77	.78	.25		
MentS motivation	38.0 (5.46)	38	14–50	−0.24	.31	.24	2.62	3.70	4.31	5.13	2.73	7.26	.76	.77	.24		

MentS-12: brief version of Mentalization Scale; MentS: Mentalization Scale; *M*: mean; *SD*: standard deviation; Mdn: median; Sk: skewness; Ku: Kurtosis; *SE*: standard error; SEm: standard error of measurement; SEd: standard error of difference; CC: critical change; MID: minimal important difference; MCRC: minimum change for a reliable change; α: Cronbach' alpha; ω: McDonald's omega total; AIC: average item correlation.

therapeutic impact. MCRC is based on the Jacobson benchmark threshold and represents the score change required to be considered statistically reliable, ensuring that the observed difference is not a random variation. Clinicians can use these benchmarks to monitor progress and make informed decisions about the effectiveness of interventions, signifying when a change in the patient's mentalization ability is substantial and when treatment adjustments may be necessary.

MentS-12 retained content coverage, agreement, and bias

Table 3 reports projected and observed correlation coefficients for the total scale and its individual subscales. The expected correlation between the total score of the MentS-12 and the MentS was *r* = .60. The observed correlation highly exceeded these expectations with *r* = .91, which is also the case for all subscores of the two versions of the scale.

Table 3 in addition presents the pattern of observed correlation between the total and subscale scores for both the MentS-12 and the MentS. These ranged from .65 to .69 for the former and were between .70 and .79 for the latter. The results of Steiger's *Z* test revealed no significant disparities in the correlations that subscale scores have with their corresponding total score for the two versions of the mentalization scale. This implies that the relationships between the total score and their corresponding subscales exhibit a consistent degree of association regardless of the MentS form used.

The coherence between the scores of the brief and the long version of the scale was further scrutinized using Bland-Altman plots and regression analyses.

After proportionate scaling to match the full length, the total score showed an average disparity of 1.10 points in favor of the full item set. Despite statistical significance, this disparity had a negligible clinical slope of .07 (Figure 2). For scores where most participants landed, the difference

approached zero, although the full item set leaned toward producing higher scores at the lowest end of the range. A comparable trend was observed at the subscale level (Figure 2), with average differences being .85 points for other-mentalization, 1.19 points for self-mentalization, and −1.65 points for motivation-mentalization, and slopes of −0.21, −0.20, and −0.11, respectively (both $p < .001$). These outcomes suggest a high concordance and minimal bias between the total scores of the brief and the full version of the MentS.

MentS-12 relationship with other study variables

Table 4 provides correlations between MentS-12 and other included variables in the study. For comparison, the same correlations are given for the long version as well. Overall, the pattern of correlations established with the full version of the scale was fully replicated using its brief form, showing: (a) significant small-to-moderate negative correlations with attachment Avoidance and generally somewhat lower correlations with attachment Anxiety, excluding the moderately high negative association between MentS-12-S and total score with ECR-R Ax; (b) significant small-to-moderate correlations with empathy, with the strongest associations retrieved for the MentS-12-S and the total score; (c) mainly significant, weak-to-moderate correlations with the Big Five, which were positive for all traits, excluding Neuroticism that

was negatively associated with MentS-12 scores; (d) significant weak-to-strong correlations with trait emotional intelligence scales, with the lowest values established for the MentS-12-M, and the highest for the MentS-12-S and the total score; (e) weak-to-moderate correlations with ability EI, with correlations generally being significant and stronger for capacities related to understanding and managing of emotions.

For all the variables considered, Cohen's q was calculated (Table 4), contrasting the correlations of each variable with the brief and the full version of the scale. Across the board, Cohen's q values did not reach statistical significance, indicating that the strength of the correlation of MentS-12 with all variables considered is the same as for the full version of the scale.

Discussion

Independent studies performed using different samples in diverse cultural contexts have recurrently emphasized the reliability and validity of the MentS as an instrument gauging mentalization capacity (Ahmadian & Ghamarani, 2021; Asgarizadeh et al., 2023; Jańczak, 2021; Matsuba et al., 2022; Richter et al., 2021; Surim & Munhee, 2018; Törenli-Kaya et al., 2021; Wen et al., 2022). As these studies show, the 28-item MentS is suitable for research purposes. However, for the scale's practical application in real-time psychotherapeutic environments, where quick and frequent administrations are crucial, a more concise version would be desirable.

In response to this need, our current study set out to formulate a brief version of the MentS, which was achieved by applying the factor analysis combined with item response theory. Our approach in item selection for the MentS-12 prioritized items that (a) presented the strongest loadings on the anticipated factor, without any overlapping loadings, whilst also (b) displaying the most desirable difficulty and discrimination attributes. Based on these criteria, 12 items, 4 per subscale, were retained for the brief version of MentS.

Overall, current findings demonstrate that the brief version is a competent substitute for its extensive counterpart. To start with, the MentS-12 fully replicated structural properties, reliability, and coherence of the full-length scale, confirming H_1. Of significance is the near-perfect alignment between the scores of the original and the truncated version, with correlations between their scores highly exceeding our initial expectations, and thus reinforcing the credibility of the brief form. Furthermore, a profound concordance between the two versions of the scale, displaying scarcely any bias, was confirmed in analyses using Bland-Altman plots coupled with regression. It is worth mentioning that minor score variations at the absolute lower spectrum using the full item set carry limited clinical consequences. This means that slight variations in lower-end scores, observed when the complete set of items is used, are not significant enough to affect clinical interpretations or treatment decisions. Finally, there were no notable disparities in criterion correlations, being that the pattern of correlations of the MentS-12 with demographic and external

Table 3. Intercorrelations between MentS-12 and MentS.

	1	2	3	4	5	6	7
1. MentS-12	2−						
2. MentS-12 self	.69	2−					
3. MentS-12 others	.69	.15	2−				
4. MentS-12 motivation	.65	.16	.44	2−			
5. MentS	.91[a]	.65	.64	.66	2−		
6. MentS self	.70	.85[b]	.19	.12	.70	2−	
7. MentS others	.73	.25	.87[c]	.52	.79	.31	2−
8. MentS motivation	.63	.32	.44	.86[d]	.78	.22	.53

MentS-12: brief version of the Mentalization Scale; MentS: Mentalization Scale.
Note. All correlations significant at $p < .01$.
[a]Expected correlation .58 (Cronbach alpha)/.60 (McDonald's omega total).
[b]Expected correlation .44 (Cronbach alpha)/.46 (McDonald's omega total).
[c]Expected correlation .48 (Cronbach alpha)/.48 (McDonald's omega total).
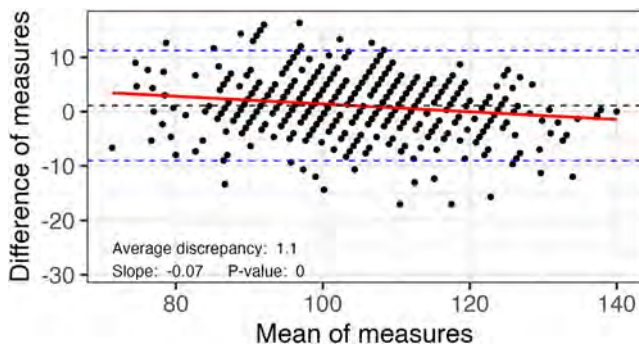[d]Expected correlation .56 (Cronbach alpha)/.44 (McDonald's omega total).

Figure 2. Bland-Altman plots comparing accuracy of MentS and MentS-12 scores.

Table 4. Correlations between MentS-12/MentS and external variables and Cohen's q.

	N	MentS-12 Self/MentS self	MentS-12 Others/MentS others	MentS-12 Motivation/MentS motivation	MentS12/MentS
Age	508	.04/−0.03 ($q=-0.01$)	−0.02/.02 ($q=-0.02$)	−0.14*/−0.22** ($q=.06$)	−0.04/−0.10 ($q=.03$)
Gender	512	−0.08/−0.13** ($q=-0.05$)	−0.18***/−0.21*** ($q=.03$)	−0.33***/−0.35*** ($q=.02$)	−0.10/−0.18*** ($q=.08$)
Attachment					
Experience in close relationships-avoidance	514	−0.29**/−0.32** ($q=.02$)	−0.25**/−0.32** ($q=.03$)	−0.34**/−0.31** ($q=.06$)	−0.42**/−0.42** ($q=.00$)
Experience in close relationships-anxiety	514	−0.45**/−0.53** ($q=.05$)	−0.18**/−0.22** ($q=.02$)	.06/.06 ($q=.06$)	−0.38**/−0.31** ($q=-0.03$)
Empathy					
Empathy quotient	512	.40**/.43** ($q=-0.01$)	.28**/.36** ($q=-0.04$)	.27**/.34** ($q=.06$)	.46**/.50** ($q=-0.02$)
Big Five					
Neuroticism	261	−0.55**/−0.53** ($q=-0.01$)	−0.20**/−0.23** ($q=.02$)	−0.09/−0.16* ($q=.09$)	−0.47**/−0.42** ($q=-0.03$)
Extraversion	261	.20**/.20** ($q=.00$)	.39**/.41** ($q=-0.01$)	.30**/.32** ($q=.08$)	.41**/.39** ($q=.01$)
Openness	261	.32**/.31** ($q=-0.02$)	.29**/34** ($q=.00$)	.31**/.41** ($q=-0.03$)	.42**/.46** ($q=.08$)
Agreeableness	261	.22**/.17** ($q=.02$)	.06/.08 ($q=-0.01$)	.10/.14* ($q=.09$)	.16*/.18** ($q=-0.01$)
Conscientiousness	261	.30**/.35** ($q=-0.03$)	.35**/.40** ($q=-0.03$)	.31**/.35** ($q=.08$)	.47**/.48** ($q=-0.01$)
Trait emotional intelligence					
Well-being	258	.39**/.45** ($q=-0.03$)	.25**/.33** ($q=-0.04$)	.20**/.24** ($q=.09$)	.44**/.45** ($q=-0.01$)
Self-control	258	.53**/.56** ($q=-0.02$)	.26**/.30** ($q=-0.02$)	.14*/.22** ($q=.09$)	.51**/.48** ($q=.01$)
Emotionality	258	.54**/.57** ($q=-0.02$)	.43**/.53** ($q=-0.08$)	.35**/.44** ($q=-0.06$)	.63**/.67** ($q=-0.03$)
Sociability	258	.43**/.50** ($q=-0.03$)	.41**/.46** ($q=-0.04$)	.27**/.30** ($q=.08$)	.57**/.55** ($q=.00$)
Total score	258	.57**/.62** ($q=.00$)	.39**/.47** ($q=-0.03$)	.28**/.35** ($q=-0.04$)	.63**/.63** ($q=.08$)
Ability emotional intelligence					
Perceiving emotions	258	.21**/.22** ($q=-0.01$)	.04/.11 ($q=-0.03$)	05/.07 ($q=09$)	.16*/.18** ($q=-0.01$)
Using emotions	258	.28**/.31** ($q=-0.01$)	.05/.10 ($q=-0.03$)	.09/.12 ($q=.09$)	.23**/.24** ($q=-0.01$)
Understanding emotions	259	.33**/.37** ($q=-0.02$)	.06/.13* ($q=-0.03$)	.13*/.13* ($q=.09$)	.30**/.29** ($q=.01$)
Managing emotions	257	.42**/.42** ($q=.00$)	.30**/.38** ($q=-0.04$)	.25**/.34** ($q=.08$)	.46**/.49** ($q=-0.02$)
Total score	256	.41**/.44** ($q=-0.01$)	.15*/.24** ($q=-0.04$)	.16*/.21** ($q=.09$)	.37**/.39** ($q=-0.01$)

MentS-12: brief version of the Mentalization Scale; MentS: Mentalization Scale; q: Cohen's q.
Note. *$p < .05$; **$p < .01$.

variables fully mirrored the one that was established for its full-length predecessor: both exhibited minimal correlations with demographic data, while presenting a spectrum of weak to robust correlations with pertinent metrics associated with attachment, empathy, and emotional intelligence. Therefore, our hypotheses H_2–H_7 were also confirmed.

An added value to our current findings is brought on by establishing clinical change thresholds for both versions of the Mentalization Scale, which enhances their applicability in therapeutic settings. Furthermore, the substantial sample size we used, as recommended by Comrey and Lee (2013), stands as a significant asset in our research.

Limitations and future directions

Although the short version of the MentS-12 has achieved our main goals, there are some limitations that should be addressed. The first is the almost exclusive reliance on the examinee as the informant since the MSCEIT was the only performance-based measure used in the study. While self-perceptions are crucial, they represent just one facet of a complex system; therefore, adopting a multi-informant approach could yield a more holistic understanding, capturing nuances potentially overlooked by self-report due to introspective limitations or biases. In this regard, to enhance the research and clinical utility of the MentS-12, future studies could aim to develop a clinician-rated version, which could offer valuable insights from a professional perspective and complement self-report data with observational assessments. Second, psychometrics of items that were retained for the brief version were possibly affected by the context, since they were examined while these items were embedded in the larger original item pool. Although these context-effects tend to be small with homogeneous scales that have strong factor

loadings (Stefana, Langfus, et al., 2023), as is the case of the MentS, our current findings still need to be corroborated by additional studies using the 12-item version itself. Furthermore, the generalizability of the MentS-12 may be limited, as the psychometric properties validated in a nonclinical sample may not accurately reflect mentalization processes in clinical populations, necessitating further validation and potential adaptation for clinical use. Finally, although we used an adequate sample, our CFA models were still based on the same sample as the exploratory analyses. Now that a reduced item set has been identified, systematic exploration of its dependability, retest stability, and sensitivity to treatment effects will be an important next step in the validation of the MentS-12.

Practical implications

In its long form, MentS has been used in various settings and with diverse populations. For instance, it was validated in a sample of patients with schizophrenia in China (Wen et al., 2022), and used in studying social cognition defects in individuals with Borderline Personality Disorder (Anupama et al., 2018), autistic traits and eating disorders (Fithall et al., 2023), childhood maltreatment and suicide risk (Huang & Hou, 2023), and for the assessment of mentalization in war veterans with and without PTSD (Berleković & Dimitrijević, 2020), and in cardiac heart disease patients (Mosavi et al., 2022).

Likewise, the 28-item form was also used in psychotherapy process research and studies with mental health professionals, examining countertransference reactions to BPD patients (Bhola & Mehrotra, 2021), resilience in health care workers (Hosgoren Alıcı et al., 2023), and psychology students' improvement in mentalizing skills (Fagerbakk et al., 2023).

Our hope is that the short version will be even more widely used in studies with clinical samples and within psychotherapy training, as it is less demanding for respondents and has equally sound psychometric properties. To conclude, the 12-item version of the MentS serves as a brief measurement tool for evaluating mentalization capacity in both research and clinical domains. Its straightforwardness and rapid administration will probably make it the instrument of choice for regular patient assessment in real-world therapeutic settings. It can also be integrated into routine outcome monitoring systems.

Author contributions

Conceptualization: AS, ZJM, AD. Formal analysis: AS. Supervision: AD. Writing—original draft: AS, ZJM, AD. Writing—review and editing: AS, ZJM, AD.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

ORCID

Alberto Stefana 🆔 http://orcid.org/0000-0002-4807-7184
Zorana Jolić Marjanović 🆔 http://orcid.org/0000-0002-7061-3588
Aleksandar Dimitrijević 🆔 http://orcid.org/0000-0001-6034-9595

Data availability statement

Both the data and the analysis code that support the findings of this study are available from the corresponding author upon reasonable request.

References

Ahmadian, Z., & Ghamarani, A. (2021). Reliability and validity of Persian version of mentalization scale in university students. *Journal of Fundamentals of Mental Health*, 23(4), 233–240.

Altaras Dimitrijević, A., & Jolić Marjanović, Z. (2010). The Mayer-Salovey-Caruso Emotional Intelligence Test: Psychometric properties of the Serbian version. *Psihologija*, 43(4), 411–426. https://doi.org/10.2298/PSI1004411A

Anupama, V., Bhola, P., Thirthalli, J., & Mehta, U. M. (2018). Pattern of social cognition deficits in individuals with borderline personality disorder. *Asian Journal of Psychiatry*, 33, 105–112. https://doi.org/10.1016/j.ajp.2018.03.010

Asgarizadeh, A., Vahidi, E., Seyed Mousavi, P. S., Bagherzanjani, A., & Ghanbari, S. (2023). Mentalization Scale (MentS): Validity and reliability of the Iranian version in a sample of nonclinical adults. *Brain and Behavior*, 13(8), e3114. https://doi.org/10.1002/brb3.3114

Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R*. Springer.

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

Bateman, A. W., & Fonagy, P. (Eds.). (2019). *Handbook of mentalizing in mental health practice*. American Psychiatric Publishing.

Berleković, V., & Dimitrijević, A. (2020). Attachment and mentalization in war veterans with and without post-traumatic stress disorder. In A. Hamburger (Ed.), *Trauma, trust, and memory* (pp. 151–159). Routledge.

Bhola, P., & Mehrotra, K. (2021). Associations between countertransference reactions towards patients with borderline personality disorder and therapist experience levels and mentalization ability. *Trends in Psychiatry and Psychotherapy*, 43(2), 116–125. https://doi.org/10.47626/2237-6089-2020-0025

Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, 327(8476), 307–310. https://doi.org/10.1016/S0140-6736(86)90837-8

Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Comrey, A. L., & Lee, H. B. (2013). *A first course in factor analysis*. Psychology Press.

Dimitrijevic, A., Hanak, N., Vukosavljevic-Gvozden, T., & Opacic, G. (2012). Psychometric properties of the Serbian version of the Empathy Quotient (S-EQ). *Psihologija*, 45(3), 257–276. https://doi.org/10.2298/PSI1203257D

Dimitrijević, A., Hanak, N., Altaras Dimitrijević, A., & Jolić Marjanović, Z. (2018). 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

Fagerbakk, S. A., Sørhøy, S. H., Nilsen, T., & Laugen, N. J. (2023). Does clinical training improve mentalization skills in future therapists? A

comparison of first and last year students of clinical psychology and of engineering. *Frontiers in Psychology*, *14*, 1066154. https://doi.org/10.3389/fpsyg.2023.1066154

Fithall, K., Gray, I. E., Linardon, J., Phillipou, A., Donaldson, P. H., Albein-Urios, N., Enticott, P. G., Fuller-Tyszkiewicz, M., & Kirkovski, M. (2023). Exploring the role of autistic traits and eating disorder psychopathology on mentalising ability in the general population. *BMC Psychology*, *11*(1), 269. https://doi.org/10.1186/s40359-023-01306-z

Fonagy, P., & Bateman, A. (2010). A brief history of mentalization-based treatment and its roots in psychoanalytic theory and practice. In M. B. Heller & S. Pollet (Eds.), *The work of psychoanalysts in the public health sector* (pp. 156–176). Routledge.

Fonagy, P., & Bateman, A. W. (2019). Introduction. In A. W. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 3–20). American Psychiatric Publishing, Inc.

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective Functioning Scale (RF)* [Database record]. APA PsycTests. https://doi.org/10.1037/t03490-000

Fonagy, P., & Target, M. (2008). Attachment, trauma, and psychoanalysis. Where psychoanalysis meets neuroscience. In E. L. Jurist, A. Slade, & S. Bergner (Eds.), *Mind to mind: Infant research, neuroscience, and psychoanalysis* (pp. 15–49). Other Press.

Fonagy, P., Luyten, P., Allison, E., & Campbell, C. (2019). Mentalizing, epistemic trust and the phenomenology of psychotherapy. *Psychopathology*, *52*(2), 94–103. https://doi.org/10.1159/000501526

Fraley, R. C., Waller, N. G., & Brennan, K. A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology*, *78*(2), 350–365. https://doi.org/10.1037/0022-3514.78.2.350

George, C., Kaplan, N., & Main, M. (1996). *The adult attachment interview* [Developed 1984-1996]. Unpublished manuscript. University of California at Berkeley.

George, D., & Mallery, M. (2010). *SPSS for windows step by step* (10th ed.). Pearson.

Hanak, N., & Dimitrijević, A. (2013). A Serbian version of modified and Revised Experiences in Close Relationships scale (SM–ECR–R). *Journal of Personality Assessment*, *95*(5), 530–538. https://doi.org/10.1080/00223891.2013.778271

Hayes, S. C., Hofmann, S. G., & Ciarrochi, J. (2020). A process-based approach to psychological diagnosis and treatment: The conceptual and treatment utility of an extended evolutionary meta model. *Clinical Psychology Review*, *82*, 101908. https://doi.org/10.1016/j.cpr.2020.101908

Hausberg, M. C., Schulz, H., Piegler, T., Happach, C. G., Klöpper, M., Brütt, A. L., Sammet, I., & Andreas, S. (2012). Is a self-rated instrument appropriate to assess mentalization in patients with mental disorders? Development and first validation of the Mentalization Questionnaire (MZQ). *Psychotherapy Research*, *22*(6), 699–709. https://doi.org/10.1080/10503307.2012.709325

Hosgoren Alıcı, Y., Hasanli, J., Saygılı, G., & Koçak, O. M. (2023). The importance of mentalization, coping mechanisms, and perceived stress in the prediction of resilience of healthcare workers. *Psychology, Health & Medicine*, *28*(9), 2635–2646. https://doi.org/10.1080/13548506.2022.2131855

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Huang, M., & Hou, J. (2023). Childhood maltreatment and suicide risk: The mediating role of self-compassion, mentalization, depression. *Journal of Affective Disorders*, *341*, 52–61. https://doi.org/10.1016/j.jad.2023.08.112

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. https://doi.org/10.1037/0022-006x.59.1.12

Jańczak, M. O. (2021). Polish adaptation and validation of the Mentalization Scale (MentS) – A self-report measure of mentalizing [Polska adaptacja i walidacja Skali mentalizacji (MentS) – samoopisowego narzędzia do pomiaru mentalizowania]. *Psychiatria Polska*, *55*(6), 1257–1274. https://doi.org/10.12740/PP/125383

Jolić Marjanović, Z., & Altaras Dimitrijević, A. (2014). Reliability, construct and criterion-related validity of the Serbian adaptation of the Trait Emotional Intelligence Questionnaire (TEIQue). *Psihologija*, *47*(2), 249–262. https://doi.org/10.2298/PSI1402249J

Knežević, G., Džamonja-Ignjatović, T., & Đurić-Jočić, D. (2004). *Petofaktorski model ličnosti*. Centar za primenjenu psihologiju.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *The American Psychologist*, *48*(12), 1181–1209. https://doi.org/10.1037/0003-066x.48.12.1181

Luyten, P., Campbell, C., Allison, E., & Fonagy, P. (2020). The mentalizing approach to psychopathology: State of the art and future directions. *Annual Review of Clinical Psychology*, *16*(1), 297–325. https://doi.org/10.1146/annurev-clinpsy-071919-015355

Matsuba, Y., Haraguchi, Y., Iwasaki, M., Otsuki, T., & Katsuragawa, Y. (2022). Development of the Japanese version of Mentalization Scale (MentS-J) and examination of reliability and validity. *Japanese Journal of Development Psychological Research*, *33*(3), 137–145. https://doi.org/10.11201/jjdp.33.137

Mayer, J. D., Salovey, P., & Caruso, D. R. (2002). *Mayer-Salovey-Caruso Emotional Intelligence Test: User's manual*. MHS.

McCrae, R. R., & Costa, P. T. (2004). A contemplated revision of the NEO Five-Factor Inventory. *Personality and Individual Differences*, *36*(3), 587–596. https://doi.org/10.1016/S0191-8869(03)00118-1

Mosavi, S. M. H., Ebrahimi, A., Sanei, H., Abedi, M. R., Feizi, A., & Bagherian-Sararoudi, R. (2022). The relationship between illness perception, mentalization and personality type D with the quality of life of coronary heart disease patients: An application of structural equation model. *Journal of Isfahan Medical School*, *40*(678), 490–497.

Müller, S., Wendt, L. P., & Zimmermann, J. (2023). Development and Validation of the Certainty About Mental States Questionnaire (CAMSQ): A self-report measure of mentalizing oneself and others. *Assessment*, *30*(3), 651–674. https://doi.org/10.1177/10731911211061280

Parker, J. D. A., Taylor, G. J., & Bagby, R. M. (2001). The relationship between emotional intelligence and alexithymia. *Personality and Individual Differences*, *30*(1), 107–115. https://doi.org/10.1016/S0191-8869(00)00014-3

Petrides, K. V. (2009). *Technical manual for the Trait Emotional Intelligence Questionnaires (TEIQue)*. London Psychometric Laboratory.

Petrides, K. V., Pita, R., & Kokkinaki, F. (2007). The location of trait emotional intelligence in personality factor space. *British Journal of Psychology*, *98*(Pt 2), 273–289. https://doi.org/10.1348/000712606X120618

Revelle, W., & Zinbarg, R. E. (2009). Coefficients Alpha, Beta, Omega, and the glb: Comments on Sijtsma. *Psychometrika*, *74*(1), 145–154. https://doi.org/10.1007/s11336-008-9102-z

Richter, F., Steinmair, D., & Löffler-Stastka, H. (2021). Construct validity of the Mentalization Scale (MentS) within a mixed psychiatric sample. *Frontiers in Psychology*, *12*, 608214. https://doi.org/10.3389/fpsyg.2021.608214

Samejima, F. (2010). The general graded response model. In M. L. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models* (pp. 77–107). Routledge.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102–111. https://doi.org/10.1037/1040-3590.12.1.102

Stefana, A., Fusar-Poli, P., Vieta, E., Gelso, C. J., & Youngstrom, E. A. (2024). Development and validation of an 8-item version of the Real Relationship Inventory–Client Form. *Psychotherapy Research*. https://doi.org/10.1080/10503307.2024.2320331

Stefana, A., Langfus, J. A., Vieta, E., Fusar-Poli, P., & Youngstrom, E. A. (2023). Development and initial validation of the In-Session Patient Affective Reactions Questionnaire (SPARQ) and the Rift In-Session Questionnaire (RISQ). *Journal of Clinical Medicine*, *12*(15), 5156. https://doi.org/10.3390/jcm12155156

Stefana, A., Vieta, E., Fusar-Poli, P., & Youngstrom, E. A. (2023). Enhancing psychotherapy outcomes by encouraging patients to regularly self-monitor, reflect on, and share their affective responses toward their therapist: Protocol for a randomized controlled trial. *JMIR Research Protocols*. https://doi.org/10.2196/55369

Steiner, M., & Grieder, S. (2020). EFAtools: An R package with fast and flexible implementations of exploratory factor analysis tools. *Journal of Open Source Software*, 5(53), 2521. https://doi.org/10.21105/joss.02521

Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford University Press. https://doi.org/10.1093/med/9780199685219.001.0001

Surim, L., & Munhee, L. (2018). Korean version of the Mentalization Scale. *Korean Journal of Counseling*, 19(5), 117–135.

Taubner, S., Hörz, S., Fischer-Kern, M., Doering, S., Buchheim, A., & Zimmermann, J. (2013). Internal structure of the Reflective Functioning Scale. *Psychological Assessment*, 25(1), 127–135. https://doi.org/10.1037/a0029138

Törenli-Kaya, Z., Alpay, E. H., Türkkal, Ş., & Özçürümez, G. (2021). Validity and reliability of the Turkish version of the Mentalization Scale (MentS). *Turkish Journal of Psychiatry*, 34(2), 118–124. https://doi.org/10.5080/u25692

Wampold, B. E., & Flückiger, C. (2023). The alliance in mental health care: Conceptualization, evidence and clinical applications. *World Psychiatry*, 22(1), 25–41. https://doi.org/10.1002/wps.21035

Wampold, B. E., & Imel, Z. E. (2015). *The great psychotherapy debate* (2nd ed.). Routledge.

Wen, Y., Fang, W., Wang, Y., Du, J., Dong, Y., Zu, X., Wang, H., Zhu, C., Zhang, L., & Wang, K. (2022). Reliability and validity of the Chinese version of the Mentalization Scale in the general population and patients with schizophrenia: A multicenter study in China. *Current Psychology*, 42(35), 30747–30756. https://doi.org/10.1007/s12144-022-04093-9

Youngstrom, E. A., & Van Meter, A. (2016). Empirically supported assessment of children and adolescents. *Clinical Psychology: Science and Practice*, 23(4), 327–347. https://doi.org/10.1037/h0101738

Youngstrom, E. A., Choukas-Bradley, S., Calhoun, C. D., & Jensen-Doss, A. (2015). Clinical guide to the evidence-based assessment approach to diagnosis and treatment. *Cognitive and Behavioral Practice*, 22(1), 20–35. https://doi.org/10.1016/j.cbpra.2013.12.005

Youngstrom, E. A., Salcedo, S., Frazier, T. W., & Perez Algorta, G. (2019). Is the finding too good to be true? Moving from "more is better" to thinking in terms of simple predictions and credibility. *Journal of Clinical Child & Adolescent Psychology*, 48(6), 811–824. https://doi.org/10.1080/15374416.2019.1669158

Youngstrom, E. A., Van Meter, A., Frazier, T. W., Hunsley, J., Prinstein, M. J., Ong, M. L., & Youngstrom, J. K. (2017). Evidence-based assessment as an integrative model for applying psychological science to guide the voyage of treatment. *Clinical Psychology: Science and Practice*, 24(4), 331–363. https://doi.org/10.1111/cpsp.12207

Appendix A

MentS-12

This questionnaire consists of 12 items. Please, read each of the items carefully and CIRCLE a number on the scale from 1 to 5, depending on how much the item is CORRECT FOR YOU PERSONALLY.

1	2	3	4	5
Completely incorrect	Mostly incorrect	Both correct and incorrect	Mostly correct	Completely correct
1. Usually I can recognize what makes people feel uneasy.	1	2	3	4	5
2. When I get upset I am not sure whether I am sad, afraid, or angry.	1	2	3	4	5
3. I can make good predictions of other people's behavior when I know their beliefs and feelings.	1	2	3	4	5
4. Often I cannot explain, even to myself, why I did something.	1	2	3	4	5
5. Sometimes I can understand someone's feelings before s/he tells me anything.	1	2	3	4	5
6. I find it important to understand what happens in my relationships with people close to me.	1	2	3	4	5
7. To understand someone's behavior, we need to know her/his thoughts, wishes, and feelings	1	2	3	4	5
8. I often talk about emotions with people that I am close to.	1	2	3	4	5
9. I find it difficult to admit to myself that I am sad, hurt, or afraid.	1	2	3	4	5
10. I am often confused about my exact feelings.	1	2	3	4	5
11. People tell me that I understand them and give them sound advice	1	2	3	4	5
12. I have always been interested in why people behave in certain ways.	1	2	3	4	5

Notes. MentS-12 Self: items 2, 4, 9, and 10; MentS-12 Others: items 1, 3, 5, and 11; MentS-12 Motivation: items 6, 7, 8, and 12.