Clinical Psychology Review 122 (2025) 102657

Contents lists available at ScienceDirect

Clinical Psychology Review

journal homepage: www.elsevier.com/locate/clinpsychrev

Review

Measurement of mentalizing: A systematic review and development of a construct validity framework

Hannah Hawksdale [a,*], Beichen Meng [a], Janet Feigenbaum [a], Peter Fonagy [a], Tobias Nolte [a,b]

[a] Research Department of Clinical, Educational and Health Psychology, University College London, UK
[b] Anna Freud, London, UK

ARTICLE INFO

Keywords:
Mentalizing
Measurement
Text-based

ABSTRACT

Mentalizing—holding in mind, appraising and reappraising mental states—is evidenced across research disciplines as a key psychological capacity of developmental and clinical importance. Due to this interdisciplinary context, there are a broad range of mentalizing measures available. The primary aim of this review was to systematically review existing mentalizing measures. MEDLINE, PSYCHINFO, HAPI and PSYCHTESTS were searched to identify English-language, text-based, general adult population mentalizing measures.

Mentalizing is established as a multidimensional concept that varies across relational contexts. However, explicit dimensional definition and systematic contextualisation of mentalizing measures is rare. The secondary aim of the review was to develop and pilot a novel framework for classifying the mentalizing dimensions and relational context operationalised by each measure. Following reference to the literature and expert discussion, a dimensional construct framework of mentalizing with 18 components was proposed and applied to existing instruments.

57 text-based measures were identified by the review. Inter-rater agreement using the novel framework was 75%. No measures completely overlapped on dimensions and relational context. Relational contexts of measures were largely mentalizing the self (52) and/or hypothetical unspecified others (33). The review demonstrates that measures of mentalizing can be compared on the basis of dimensions operationalised and relational context, increasing utility in measure selection. Divergent construct validity highlights a need for careful selection of appropriate measures on the basis of mentalizing components of interest. Increased clarity of construct validity could be used to illuminate conflicting findings in the mentalizing literature and to advance the development of new assessment instruments.

1. Introduction

1.1. Background

Mentalizing is the activity of holding in mind, appraising and re-appraising the mental states (e.g. thoughts and feelings) of others and the self (Allen, 2003; Fonagy et al., 1998). There are many terms derived from different disciplines that describe mentalizing or aspects of it, including reflective functioning, theory of mind, and metacognition (Quesque et al., 2024). The term 'mentalizing' is used here, due to the broad application of the term in considering the qualities of appraisal of mental states.

Mentalizing is a foundational psychological capacity and activity that is instantly recognisable to human beings. A great breadth of work across different traditions aims to define and measure mentalizing, with differing foci including consciousness, sense of self, affect and behavioural regulation, intentionality, co-operation and deception (Baron-Cohen, 2000; Bateman & Fonagy, 2004; de Waal, 2019; Dennett, 1987; Gallup, 1982; Schwarzer et al., 2021). Developmental, evolutionary and ethological studies of mentalizing have focused on whether and how mentalizing is acquired and develops in children, generally and in comparison to non-human species, and includes consideration of caregiver mentalizing and epistemic trust (Call & Tomasello, 2008; Duschinsky & Foster, 2021; Fonagy et al., 1991; Fonagy & Campbell, 2016; Fonagy & Target, 1996; Rubio-Fernández & Geurts, 2013; Slade et al., 2005; Wobber et al., 2014). There is a consensus that mentalizing is important clinically, with evidence for differences in mentalizing being correlated with secure attachment, personality structure and

* Corresponding author.
E-mail address: Hannah.hawksdale.21@ucl.ac.uk (H. Hawksdale).

https://doi.org/10.1016/j.cpr.2025.102657
Received 20 March 2025; Received in revised form 19 September 2025; Accepted 2 October 2025
Available online 10 October 2025

psychiatric diagnosis (Allen et al., 2017; Camerlinck et al., 2024; Kivity et al., 2024; Fonagy & Target, 1997). There are links between less accurate, less complex or less adaptive mentalizing profiles and specific psychiatric diagnoses, including Borderline Personality Disorder (BPD) (Euler et al., 2021; Rifkin-Zybutz et al., 2021), Schizophrenia (Lysaker et al., 2005; Mitchell et al., 2012) and Anorexia Nervosa (Fonagy & Leigh., 1996; Ward et al., 2001) as well as evidence for mentalizing differences between autistic and neurotypical people (Baron-Cohen, 1997; Baron-Cohen et al., 2001; Moessnang et al., 2020).

Understanding differences in mentalizing that may operate as transdiagnostic mechanisms in maintaining or preventing mental health difficulties can improve support for people with these diagnoses, enabling developmental trajectories that promote mental health resilience. A number of studies, including those of the efficacy of Mentalization Based Therapy, suggest that changes in mentalizing abilities are both a clinical outcome in their own right and a possible transdiagnostic mechanism of change across psychological therapies (Katznelson, 2014; Leichsenring et al., 2024; Luyten et al., 2024; Malda-Castillo et al., 2019; NICE, 2009; Rameckers et al., 2021).

Differences in mentalizing are typically mapped two-dimensionally, hierarchically (from good to poor or effective to ineffective mentalizing), and as a quality of an individual across contexts: perhaps a legacy of the original ethological and developmental measures indicating binary presence or absence of mentalizing, more often termed 'theory of mind'. Recent reviews of correlations between apparent deficits in mentalizing and diagnoses including Autism, BPD, Schizophrenia and Depression demonstrate mixed evidence when mentalizing is measured two dimensionally and without consideration of relational context. These studies highlight more complex relationships than predicted (Colle et al., 2019; Katznelson, 2014; Nijhof, 2023; Camerlinck et al., 2024). For example, findings that people with BPD diagnoses (when compared with community samples) demonstrate more attuned affective empathy for other people, but less complex or nuanced mentalizing when this is measured more holistically (Dinsdale & Crespi, 2013).

The body of literature on the multidimensional nature of the mentalizing concept (Ballespí et al., 2021; Choi-Kain & Gunderson, 2008; Fonagy & Luyten, 2009; Gori et al., 2021; Lakhani et al., 2021) could help illuminate this mixed evidence for the clinical impacts of mentalizing. The difficulties interpreting these studies speak to the importance of operationalising the increasingly well-evidenced multidimensional nature of mentalizing. A structure of mentalizing consisting of four dimensions, each with two polarities, has been proposed (Fonagy & Luyten, 2009). These are self/other, cognitive/affective, internal/external (referring to cues for mentalizing), and automatic/controlled (also referred to as implicit/explicit or system1/system 2 (Kahneman, 2012; Kivity et al., 2024)). There is evidence for each pair of polarities being underpinned by distinct neural systems (Fonagy & Luyten, 2009). One additional overarching dimension that has been considered in the development of several measurement paradigms (Dziobek et al., 2006; Fonagy et al., 1998; Fonagy et al., 2016) is certainty/uncertainty (with an overlapping concept of hyper/hypo mentalizing). Excessive certainty about mental states indicates inflexibility—a lack of understanding of the often opaque and changeable nature of mental states—and excessive uncertainty indicates avoidance of mentalizing. Despite the evidence in support of this dimensional model (Ballespí et al., 2021; Choi-Kain & Gunderson, 2008; Fonagy & Luyten, 2009; Gori et al., 2021; Lakhani et al., 2021), many measures of mentalizing and related empirical studies do not explicitly consider the multidimensional nature of the concept.

It is also rare to find measures or studies that view mentalizing abilities as relationally dependent. The examples below demonstrate the importance of relationally contextualising mentalizing research. Firstly, the importance of relational context is already recognised in the separation of research around caregiver mentalizing, speaking specifically to mentalizing within the relationship between caregiver and child (Katznelson, 2014). Secondly, the double empathy problem, a concept suggesting that posited mentalizing deficits in Autism are due to communication difficulties between neurodivergent and neurotypical people (noting the dominant neurotypical perspectives of researchers and health professionals) could also speak to the relational nature of mentalizing (Chapple et al., 2021; Ekdahl, 2023; Milton, 2012). Thirdly, theoretical models of BPD that suggest regression to developmentally earlier forms of mentalizing (pre-mentalizing modes) only when the attachment system is activated, also point to important relational aspects of mentalizing (Fonagy & Luyten, 2009). Additionally, within exploration of the links between mentalizing and depression, schizophrenia and bulimia, there are indications that the relationship within which mentalizing is assessed has an impact on findings (Katznelson, 2014). Relational contexts that impact mentalizing include broader contextual differences between those in the relationship, such as cultural differences or similarities. For example, a review of mentalizing research with a cross-cultural perspective highlights differences in emphasis on mentalizing self or others in more individualistic or collectivist cultures respectively (Aival-Naveh et al., 2019).

Relational context of mentalizing is also an important part of assessment context, as demonstrated by the debate around how mentalizing skill will also impact accuracy of self-assessment of that skill (Müller et al., 2023; Rumeo, 2022). Less frequently discussed is the potential impact of the relationship between the researcher and participant on measurement of mentalising. For example, participants need to be able to express themselves in a particular way for mentalizing to be noted by a researcher during an interview or scoring procedure (Fonagy et al., 1998) with no consideration of whether material rated is generated in a research interview where participant and interviewer are strangers, or during therapeutic sessions (Semerari et al., 2003), as well as without consideration of the potential impact of researchers' own mentalizing styles on scoring or even development of measures. For example, perhaps the emphasis on mentalizing as an individual ability, rather than relationally and socially dependent, is due to a more self-focused or individualistic mentalizing profile (adaptive within a particular set of cultural contexts) of many researchers within the history of the field.

Another difficulty in advancing understanding in this area is the wide heterogeneity of terminology, in part due to its parallel emergence across a variety of fields of study (Quesque et al., 2024; Quesque & Rossetti, 2020). There are many overlapping and synonymous concepts, including: mentalizing, reflective functioning, theory of mind, social cognition, metacognition, mind-reading, and mental state attribution (see Table 1) and different fields focus on nuanced differences both in terms of developmental frameworks, conceptualizations in psychopathology and operationalization. The proliferation and interdisciplinary nature of these terms again demonstrates mentalizing as fundamental to psychological development and mental health, and worthy of this breadth of academic attention, while also creating a complex, unstructured web of conceptual perspectives, terminologies and methods of study. This complexity and heterogeneity is added to by the nature of mentalizing as multidimensional as discussed above, with specifiable but interwoven dimensions and polarities (Fonagy & Luyten, 2009).

Table 1 lists key metalizing-related terms taken from existing, non-comprehensive lists of mentalizing measures (Bateman & Fonagy, 2016; Bateman & Fonagy, 2019; Jańczak, 2021; MacLennan, 2021) and their OVID 'explode terms' synonyms. Literature describing the origin

Table 1
Definitions of some of the terms relating to mentalizing (encompassing cognitive[1] and affective[2] mental states of self[3] and others[4])

Term	Mentalizing
Origin	Psychoanalytic theory and practice (Marty, 1990)
Definition (s)	"to conceive of conscious and unconscious mental states in oneself[3] and others[4]… to represent ideas[1] and desires" (Fonagy, 1991) "the capacity to interpret both the self[3] and others[4] in terms of internal mental states such as feelings[2], wishes, goals, desires, and attitudes" (Fonagy et al., 2016)

Term	Reflective Functioning
Origin	Operationalization of the concept of mentalizing from psychoanalytic background (Fonagy et al., 1998) also used in cognitive psychology (Morton & Frith, 1995).
Definition (s)	"the psychological processes underlying the capacity to *mentalize*[1,2,3,4]" (Fonagy et al., 1998)

Term	Theory of mind
Origin	Ethology (Premack & Woodruff, 1978)
Definition (s)	"An individual has a theory of mind if he imputes mental states to himself[3] and others[4]... purpose or intention, as well as knowledge, belief, thinking[1], doubt, guessing, pretending, liking[2]" (Premack & Woodruff, 1978)

Term	Metacognition
Origin	Educational psychology (Flavell, 1979), taken up by psychoanalytical theorists and practitioners (Semerari et al., 2003)
Definition (s)	"'any conscious cognitive[1] or affective[2] experiences that accompany and pertain to any intellectual enterprise' (Flavell, 1979, 906; Meijer et al., 2013) "the knowledge[1] and control individuals have over their own[3] cognition and learning experiences" (Allen & Armour-Thomas, 1993) "Metacognition… refers to a broad set of cognitive[1] and affective[2] skills which allow people to identify mental states, reasoning about them, and ascribing them to themselves[3] and others[4]."(Semerari et al., 2012)

Term	Social cognition
Origin	Social psychology and cognitive psychology (Lakhani et al., 2021; Westen, 1991)
Definition (s)	Social cognition[1]…entails psychological processes involved in making sense of self[3] and others[4] in the social context (Lakhani et al., 2021) "Research and theory in … social cognition [is] interested in the way mental representations of the self[3] and other people[4]… are constructed and encoded, in the cognitive[1] and affective[2] processes brought to bear on these representations, and, at least to some degree, in the way these psychological processes mediate behavior." (Westen, 1991)

Term	Mindreading
Origin	Interdisciplinary term attempting to combine ecological, developmental and computational ideas, (Whiten, 1991)
Definition (s)	"you and I are mindreaders... we have the capacity to imagine or represent states of mind that we[3] or others[4] might hold." (Baron-Cohen, 1997) "explaining and predicting others'[4] behaviour by reference to their inner mental states: their desires, beliefs, expectations and a host of other psychological concepts. Jack is more afraid[2] of Jill If Jack thinks[1] that Jill thinks[1] that Jack is afraid[2] of Jill" (Whiten, 1991)

Term	Mental state attribution
Origin	Used synonymously with mentalizing (Frith et al., 2003)
Definition (s)	"Mentalizing is a cognitive[1] process entailing the attribution of mental states (e.g., beliefs, desires, intentions, feelings[2]) to self[3] and others[4]" (Orr & Gilead, 2022)

and definition of these terms (Table 1) demonstrates that they have come to be defined in ways that are largely overlapping, encompassing cognitive and affective mental state appraisal of both self and others. This is despite these terms often being described as distinct (Aival-Naveh et al., 2019; Choi-Kain & Gunderson, 2008), likely due to their origins in distinct areas of study and original definitions with narrower focus. This conceptual convergence has been noted by researchers from different areas (Choi-Kain & Gunderson, 2008; Colle et al., 2019; Healey et al., 2015; Hutchins et al., 2021; Quesque et al., 2024; Quesque & Rossetti, 2020; Semerari et al., 2003). While it could be argued that using these terms synonymously conflates concepts with subtly different perspectives and meanings, the evidence that this is already happening across areas of study relating to this concept feels important to acknowledge and work with. Unacknowledged use of the same terms to sometimes indicate overlapping, sometimes distinct, sometimes synonymous concepts makes clarity of measurement, development and debate, and subsequently clinical use, much more difficult (Colle et al., 2019; Quesque et al., 2024).

1.2. Review rationale

There are several existing collections of measures of mentalizing, largely for clinical use (Bateman & Fonagy, 2016; Bateman & Fonagy, 2019; Jańczak, 2021; MacLennan, 2021), but this has not been exhaustively reviewed. The conflation and ambiguity around mentalizing terminology creates a seemingly infinite spectrum of search terms for review (Quesque et al., 2024) and difficulties conceptualising finite definitions have previously presented a barrier to systematic review (MacLennan, 2021). The issue of concepts being assumed to be the same because the same terminology is used, or different because terminology differs, has been discussed as an issue across psychological research. Theory of Mind, which overlaps with mentalizing, has been presented as an exemplary construct suffering from this (Flake & Fried, 2020; Warnell & Redcay, 2019).

To address this fallacy and resolve the conflation problems in mentalizing research, improved clarity of the construct validity of existing measures is needed. If measures are differentiated on a theoretical basis, researchers can transparently and explicitly select appropriate measures for their research questions (Flake & Fried, 2020). Evidence-based construct frameworks for describing measures of the overlapping construct of 'Theory of Mind', more focused on the *presence* of mentalizing, have been described, (Barnby et al., 2023; Quesque & Rossetti, 2020). However, no parallel systematic research has yet been carried out to consider which components of the *quality* of mentalizing are being operationalised by which measures (Katznelson, 2014). The dimensional framework described above has been used to develop additional measures of mentalizing (Gagliardini et al., 2018; Gagliardini & Colli, 2019; Gori et al., 2021), which therefore arguably hold greater clarity in their construct validity. However, given the extensive body of literature based on current measures, the ability to compare existing instruments using theoretical constructs would be invaluable.

1.3. Review aims

The broad aim of the following review was to contribute structure and clarity to the wide and heterogeneous use of measures surrounding mentalizing and related concepts (Quesque & Rossetti, 2020). The primary aims of the review were to:

1. Systematically review all text-based (as defined below) English language measures of mentalizing that are available for use with the general adult population.
2. Create a framework to evaluate constructs—the dimensions and the relational context of mentalizing—operationalised by mentalizing measures.

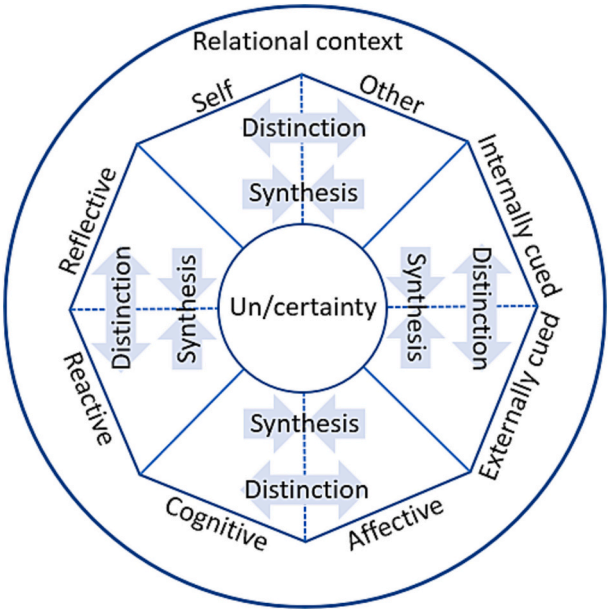

Fig. 1. Proposed dimensional construct framework of mentalizing.

1.4. Definition of text-based measures

Use of a concrete definition based on measurement paradigm, rather than exclusion on the basis of the inconsistently used terminology in the literature (Quesque et al., 2024) was intended to make it easier for future research to build upon the current review. Text-based measures were selected because measures of presence or absence of mentalizing, more commonly termed Theory of Mind, frequently make use of experimental paradigms, for example in false-belief tasks (Aival-Naveh et al., 2019; Barone et al., 2019). Text-based measures were defined as questionnaires, interviews or text coding systems: a series of questions asked directly to the assessee or informant, with multiple choice or open text responses, or methods of coding text produced by other paradigms. Text may be directly created by the assessee or informant, or recorded by an assessor during measurement (including transcription of verbatim narrative data). Experimental measurement paradigms, where the assessee is presented with a task or other stimulus (with the exclusion of questions intended to generate a reflective process) and their response measured in a standardised way, are excluded from the current review.

2. Method

2.1. Development of a construct rating framework for measures of mentalizing

In order to examine construct validity of existing mentalizing measures, a dimensional construct framework of mentalizing was developed, in parallel to similar work on Theory of Mind (Barnby et al., 2023; Quesque & Rossetti, 2020). This was necessary to consider which components of mentalizing are being operationalised by which measures, in order to clearly classify and compare construct validity of instruments. The framework was grounded in the literature described in the introduction, and defined through discussion with experts in the field. The process of development is described below following an outline of the framework.

The final construct rating framework is depicted in Fig. 1. It comprises 18 components: 17 dimensions embedded in a relational context. The dimensions include eight "aspects", four "distinction" skills, four "synthesis" skills, and a "certainty/uncertainty" skill. Each aspect or skill can be measured independently and unidimensionally. Aspects are how well someone is able to appraise: 1. mental states of self, 2. mental states of other(s), 3. cognitive mental states, 4. affective mental states, 5. internally cued mental states, 6. externally cued mental states, 7. reactive (automatic or implicit) mental states and 8. reflective (controlled or explicit) mental states (Bateman & Fonagy, 2016; Bateman & Fonagy, 2019; Fonagy & Luyten, 2009; Luyten et al., 2020).

Skills link pairs of aspects, with distinction skills defined as abilities to distinguish: 1. mental states of self from those of others, 2. internally cued from externally cued mental states, 3. cognition from affect, and 4. reactive from reflective mental states. For example, "I felt embarrassed speaking up (self), but my colleagues did not feel embarrassed for me and said they were glad to hear my views (other)" would be an example of appropriate self-other distinction, or "I had a strong sense of unease (internally cued), but when she smiled I also had the thought that she looked trustworthy (externally cued)" is an example of distinction of internally and externally cued mental states. Ability to distinguish some of these pairs is less well explored in the literature, whereas others have been documented as clinically important (such as self-other distinction/differentiation (Bender & Skodol, 2007; De Meulemeester et al., 2021), highlighted as an important factor in maintenance of symptoms of Borderline Personality Disorder (Fonagy & Luyten, 2009)).

Synthesis skills were defined as abilities to then re-integrate each of these four pair of aspects. For example "I felt embarrassed speaking up (self), but my colleagues said they were glad to hear my views (other). I still felt a bit vulnerable but also felt welcome to speak up again" would be an example of appropriate self-other synthesis, or "I had a strong sense of unease (internal cue), but when she smiled I also had the thought that she looked trustworthy (externally cued) and I felt friendly while at the same time feeling cautious about what to share initially" exemplifies an appropriately balanced synthesis of internally and externally cued mental states. Imbalances in these mentalizing processes have been highlighted as clinically relevant in a number of studies (Gagliardini et al., 2018; Gagliardini & Colli, 2019).

Certainty/uncertainty was also proposed as an additional overarching skill dimension, again with an appropriate balance between the poles indicating better mentalizing. Excessive certainty about mental states across all four aspects would represent poorer mentalizing due to inability to acknowledge that mental states change, and can be inaccessible, within both our own and others' minds. Excessive uncertainty reflects a tendency to avoid mentalizing all together (Dziobek et al., 2006; Fonagy et al., 2016).

This creates a structure that can be used to identify the components of mentalizing examined by different measures. For example, the Reading the Mind in the Eyes Test (RMET) (Baron-Cohen et al., 2001; Fonagy & Bateman, 2006) asks participants to appraise, as quickly as possible (reactive), the emotion (affective) of the person (other) in a drawing of an expression (externally cued). It does not measure any distinction, synthesis or un/certainty dimensions. It measures mentalizing in the context of a relationship with a stranger.

None of the dimensions of the framework are entirely novel and definitions of each aspect can be found in previous literature as referenced above. Defining each of the eight aspects as separate dimensions has been done in clinical handbooks (Bateman & Fonagy, 2016; Bateman & Fonagy, 2019), but not defined as separate dimensions from the skills in balancing these aspects as is suggested here. A variety of arrangements of these dimensions have been suggested previously,

(Bateman & Fonagy, 2016; Colle et al., 2019; Gagliardini & Colli, 2019; Gori & Topino, 2023) but generally without stating explicitly that the eight aspects are themselves unidimensional. Simultaneously, the clinical importance of balance between these pairs is frequently discussed in the literature (Bateman & Fonagy, 2019; Choi-Kain & Gunderson, 2008; Gagliardini & Colli, 2019; Katznelson, 2014) with contextually appropriate balance hypothesised to characterize effective mentalizing. One example is the hypothesised association of a mentalizing imbalance towards self with Narcissistic or Antisocial Personality Disorder diagnoses, and a mentalizing imbalance towards other with Borderline or Dependent Personality Disorder diagnoses (Gagliardini et al., 2018).

Following initial development of the model from the literature, discussion with three experts in the field about balance between pairs drew out two types of skills: distinguishing parts of a pair, and then re-synthesising them effectively. During discussions, including both distinction and synthesis skills for all four pairs in the model was felt to allow for the broadest examination of mentalizing instruments, although some skills are much more established and well-examined (for example self-other distinction/differentiation (Bender & Skodol, 2007; De Meulemeester et al., 2021)). It is important to note that this framework, while developed using existing literature and expert discussion, is currently unvalidated.

2.2. Mentalizing measure construct rating procedure

For each measure found through systematic search (see section 2.3), all content items and the measure's rating or scoring system were considered against the framework (Fig. 1) and coded against each dimension of the model by two independent raters, with disagreements resolved through discussion. For each dimension in the model, raters asked 'is there an explicit attempt to measure this dimension in the items or scoring protocol of this measure?' and 'would the final score produced by the measure be altered in tandem with this aspect or skill?'. Measures were given a traffic light rating against each dimension of the framework, with whether each dimension was operationalised as a necessary part of the measure (green), partially measured or gave an opportunity for the dimension to be measured (amber), or was not operationalised by the measure (red). It was possible for one measure item to be coded as operationalising more than one dimension of the model. Raters recorded specific textual evidence from measure content and procedures to justify their decisions (see Appendices A & B). Any relational contexts arising during application and scoring of measures that could impact measure output were recorded. A shorthand for this was developed as part of the rating process (see Results). Specific textual evidence from the measure content, manual or scoring procedures was recorded by raters to justify this assessment (see Appendices A & B). A worked example of the rating procedure is laid out in Appendix B.

2.3. Searches and data extraction

MEDLINE, PSYCHINFO, HAPI and PSYCHTESTS were searched to identify measures of mentalizing published or described in publications prior to 23/02/2025. Key terms from existing, non-comprehensive lists of mentalizing measures (Bateman & Fonagy, 2016; Bateman & Fonagy, 2019; Jańczak, 2021; MacLennan, 2021), were used to develop the search strategy (see Appendix C), with OVID 'explode terms' functionality used to identify synonyms. Searches were also checked to ensure they returned all measures referenced in existing lists. Finally, terms and the search strategy were reviewed by experts in the field.

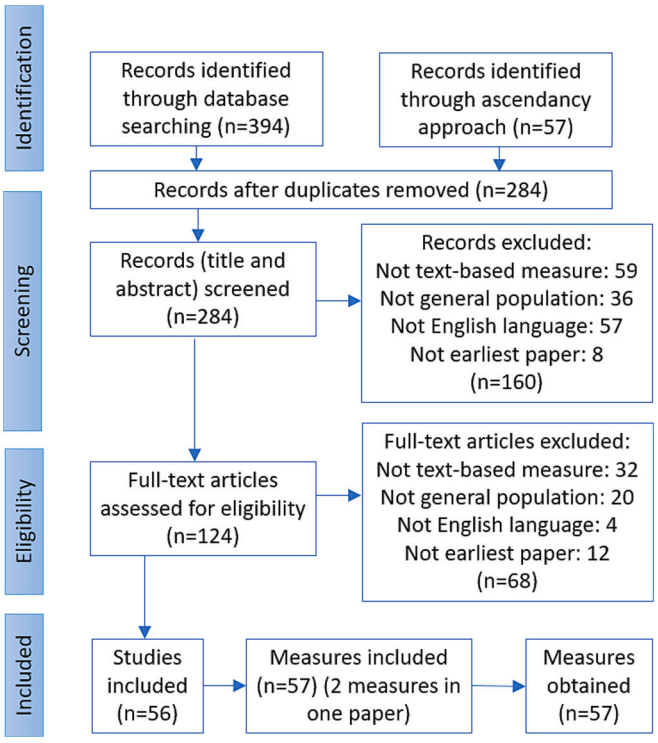

Fig. 2. PRISMA diagram.

Publications were included if they described the development or validation of text-based English-language measures of mentalizing for the general adult population. Publications were excluded if there was an earlier paper within the search results describing the same measure, so only the original published validation of each measure is compared in the current study. Publications were also excluded if the measures they described were not relevant to mentalizing, or if the measures they described were experimental or observational instruments. Publications were also excluded if there was no evidence they could be used with the general adult population (e.g. those developed for sole use with either those under 18 years or specific adult populations, such as parents, particular clinical groups or forensic populations).

An ascendancy approach was then employed with the same criteria applied to references of review papers found by the search strategy. Two reviewers screened all records retrieved by this search strategy for inclusion, with an agreement rate of 85.36%, and disagreements resolved through discussion. Numbers of publications and measures excluded at each stage are outlined in the below Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) diagram (Tricco et al., 2018), (see Fig. 2). See Appendix D for full references of included records and papers.

PRISMA 2020 guidelines for systematic reviews were adhered to (Page et al., 2021). Data and research materials are available at https://osf.io/cwd6z/overview?view_only=f8287943cd4e4923aea54795340fd8c4. The protocol for this review was registered with Open Science Framework (Hawksdale, 2023) before starting data extraction. It includes the Preferred Reporting Items for Systematic Reviews and Meta-analysis Protocols (Moher et al., 2015).

Table 2
Data for extraction

Measure Content
Name of the outcome measurement instrument
The version of the outcome measurement instrument (including variation in protocol)
Construct measured by the measurement instrument (assessed using new framework)
Instrument protocol:
- equipment
- preparatory actions
- data collection (procedure and time taken)
- data processing and storage
- score assignment (procedure and time taken)
Characteristics of the population originally used for development/validation:
- age
- gender
- ethnicity
- clinical features such as diagnoses/symptoms of mental health difficulties
- clinically relevant life experiences (e.g. studies within inpatient populations)

The features listed in Table 2 and Appendix E were then extracted from each of the included papers. These features are adapted from the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) risk of bias guidelines (Mokkink et al., 2020). Where not all details of measure protocol and content were available within the paper, this was accessed either from publicly available sources, other published papers, or by requesting these directly from authors via the email of the corresponding author of the publication.

3. Results

3.1. Overview

56 publications, describing 57 text-based measures of mentalizing (Tables 4, 6, 8 & 10 and Appendix D) for use with the general adult population were found by the search strategy and included for rating against the construct framework. Of the measures reviewed, 38 were self report questionnaires, eight were transcript rating systems, six were interviews, three were clinician report questionnaires, and two were informant report questionnaires (see Tables 5, 7, 9 & 11). All measures reviewed had some form of validity and reliability testing reported to be acceptable in the earliest paper highlighted by the search strategy (see Appendix E). Raters initially agreed on 75% of rating decisions using the construct framework, with a significant, large effect size Spearman's correlation coefficient, $R = .66$, $p < .001$ between ordinal rater decisions. This suggests that the rating framework can be used reliably. Disagreements were resolved through discussion to arrive at the final ratings in Tables 5, 7, 9 & 11.

No reviewed measures overlapped on all operationalised dimensions, (other than shorter or longer forms of the same measure and those that operationalised none of the dimensions), and no reviewed measures operationalised all of the dimensions described in the framework. Every dimension in the framework was operationalised by at least one measure. A majority (>28) of reviewed instruments operationalised Self, Other, Cognitive and Affective aspects of mentalizing, and a majority at least partially operationalised Self, Other, Internally cued, Externally cued, Cognitive, Affective, Reactive, Reflective and Un/Certainty dimensions of mentalizing (partial and full operationalisation ratings have been summed to demonstrate this) (see Table 3). A minority of measures (<29) operationalised Distinction or Synthesis of any of the four pairs of aspects.

When examining relational context ratings across all measures, a majority of measures were applied in the context of mentalizing the self (50), and/or mentalizing hypothetical, unspecified others (34 – including unspecified professionals (2)). 22 measures were applied in the context of specific relationships with others (unspecified known others (7), close others (2), friends (5), attachment figures (2), interviewer (6)). Six measures, all transcript rating systems, could also be applied in a variety of relationship contexts.

To support interpretation of the review, since measure ratings did not overlap, they are reported below in clusters focused broadly on synthesis (including certainty/uncertainty), distinction or aspect dimensions. Measures were included in these categories if they were rated as operationalising more than half of the dimensions in that category, (with partial operationalisation counted as 0.5). The clusters should be used with caution as the results show that all measures are distinct from each other in their detailed dimensional construct validity. Clustering by domain (self/other, cognitive/affective etc.) was considered, but 48 of the 57 measures operationalised dimensions from three or more domains.

3.2. Broadly synthesis focused measures

10 of the included measures were classed as broadly synthesis

Table 3
Number of measures operationalising each dimension (of 57 reviewed).

	Self	Other	Distinction	Synthesis	Internally cued	Externally cued	Distinction	Synthesis	Cognitive	Affective	Distinction	Synthesis	Reactive	Reflective	Distinction	Synthesis	Un/certainty
Y	45	36	18	12	13	17	4	1	32	39	6	8	18	25	4	2	25
Y P	52	42	24	17	31	29	12	5	48	45	12	19	36	37	7	2	34
N	5	15	33	40	26	28	45	52	9	12	45	38	21	20	50	55	23

Majority (>28) Y=Measured, Y P = Measured + partially measured, N =Not measured.

Table 4
Broadly synthesis focused measures: descriptions and protocols

Measure (Included paper for this review)	Authors' description and brief protocol
Metacognition Assessment Scale (MAS-R) (Semerari et al., 2003)	A method and a scale for the evaluation of the metacognitive profiles of psychotherapy patients. *Transcript rating system* for therapy session transcripts, or interviews/narratives about close relationships. A manual and 20 hours' training assists raters. Passages between two interruptions by a therapist/interviewer are rated on 3 scales: understanding own mind, understanding others' minds and mastery, on a 1–6 scale (not engaged–sophisticated)
Reflective Functioning Rating Scale (RFRS) (Meehan et al., 2009)	A multi-item *clinician report questionnaire* for assessing reflective functioning, based on the Reflective Function Scale (RFS, below), that can also be applied to interviews. 50 items, scale 1–5 (not true–very true) based on last 6 months of treatment
Reflective Functioning Scale (RFS) (Fischer-Kern et al., 2010)	A *transcript rating system* giving a global rating of the quality of mentalizing in attachment narratives. Usually applied to the Adult Attachment interview (AAI), an approximately 20 item semi-structured interview about experiences with early attachment figures, family and own children. A manual (Fonagy et al., 1998), scale and three days' training with reliability test assists raters. Illustrations of moderate–high reflective functioning are identified in the text. Ratings of -1 to 9 are given to each passage using a manual. An overall rating is given to each interview based on passage ratings and guidance from the manual and training
The Mentalization Scale (MentS) (Dimitrijević et al., 2018)	A *self report questionnaire* assessing implicit and explicit interpretation of the meaning of own and others' actions using intentional mental states (e.g. feelings, beliefs, desires). 28 items, scale 1–5 (completely incorrect–completely correct)
Metacognition Assessment Scale abbreviated (MAS-A) (Lysaker et al., 2005)	A *transcript rating system* adapted from the MAS-R (above), Usually applied to Indiana Psychiatric Illness Interviews (or community version), a 5 section, semi-structured interview, with 15 prompts, asking for life story and mental illness (often schizophrenia) or times of hardship narratives. A manual with examples and anchor points for scores, and training assists raters. Scored on 4 scales: Self-Reflectivity (0–9), Understanding of Others' Minds (0–7) Decentration (0–3) and Mastery (0–9). Whole transcripts are rated on each scale.
Brief Reflective Functioning Interview (BRFI) (Rutimann & Meehan, 2012)	A briefer alternative to the AAI as a measure for coding reflective functioning. *Semistructured interview* with approximately 10 questions, which asks about relationship with one attachment figure and someone currently important to participant. Scoring as per the RFS manual (see above)
Mentalized Affectivity Scale (Greenberg et al., 2017)	A *self report questionnaire* for multidimensional assessment of emotion regulation, incorporating the reliance of regulation on capacity for mentalizing. 60 items, scale 1–7 (Disagree strongly–Agree strongly)
Multidimensional Mentalizing Questionnaire (MMQ). (Gori et al., 2021)	A *self report questionnaire* for dimensional assessment of mentalizing. Scales include cognitive-affective, self-other, internal-external, explicit-implicit axes and integration of positive and negative mentalizing clusters on different polarities. 33 items, scale 1–5 (Not at all–A great deal)
Psychological Mindedness Scale (PMS) (Shill & Lumley, 2002)	A *self report questionnaire* assessing the ability to relate thoughts, feelings, and actions to learn meanings and causes of experience and behaviour. 45 items, scale 1–4 (strongly agree–strongly disagree)
Interactive Mentalizing Questionnaire (IMQ) (Wu et al., 2022)	A *self report questionnaire* assessing mentalizing in different relationships: self–self, self–other, and other–self. 20 items, scale 1–4 (very true for me–very false for me)

focused. For authors' descriptions and brief protocols for these measures see Table 4.

Table 5 shows the ratings against the construct framework for the broadly synthesis focused measures. Five of the ten measures operationalised mentalizing in the context of a specific relationship, including two measures with an attachment relationship context. Un/certainty is operationalised by all of these measures, and almost all operationalise cognitive-affective and self-other synthesis. Half operationalise or partially operationalise internally-cued-externally-cued synthesis, and only one operationalises reactive-reflective synthesis.

Table 5
Construct framework ratings for broadly synthesis focused mentalizing measures.

Broadly synthesis focused mentalizing measures	Relational context	Construct framework ratings																
		Self	Other	Distinction	Synthesis	Internally cued	Externally cued	Distinction	Synthesis	Cognitive	Affective	Distinction	Synthesis	Reactive	Reflective	Distinction	Synthesis	Un/certainty
Metacognition Assessment Scale Revised [D A TRS]	R-S-S/kO (d)	Y	Y	Y	Y	N	Y	P	P	Y	Y	Y	Y	N	P	Y	N	Y
Reflective Functioning Rating Scale [D A CRQ]	C-S-S/hO	Y	Y	Y	Y	Y	Y	Y	P	Y	Y	N	Y	Y	Y	N	N	Y
Reflective Functioning Scale [A TRS]	R-S-S/aO/iO-S/O	Y	Y	Y	Y	N	N	N	P	P	P	N	Y	P	Y	N	N	Y
The Mentalization Scale [D A SRQ]	S-S/hO	Y	Y	Y	N	Y	Y	N	Y	P	Y	P	Y	Y	Y	P	N	Y
Metacognition Assessment Scale Abbreviated [D TRS]	R-S-S/kO/iO (d)	Y	Y	Y	Y	N	N	N	N	P	Y	Y	Y	N	N	N	N	Y
Brief Reflective Functioning Interview [A SSI]	R-S-S/aO/fO/iO	Y	Y	Y	Y	P	P	P	P	Y	P	N	P	P	Y	N	N	Y
Mentalized Affectivity Scale [A SRQ]	S-S/hO	Y	Y	N	P	P	Y	N	N	P	Y	P	P	Y	Y	N	Y	Y
Multidimensional Mentalizing Questionnaire [SRQ]	S-S/hO	Y	Y	N	Y	N	N	N	N	N	Y	N	P	P	N	N	N	Y
Psychological Mindedness Scale [A SRQ]	S-S/hO/pO	Y	Y	P	P	Y	N	P	N	Y	Y	N	Y	P	Y	P	N	Y
Interactive Mentalizing Questionnaire [A SRQ]	S-S/fO/hO	Y	Y	Y	Y	Y	P	P	N	Y	P	N	P	Y	N	N	N	Y

Y = Measured, P = Partially measured, N = Not measured. [D] Also meets criteria for distinction focus. [A] Also meets criteria for aspect focus. [TRS] Text rating system, [CRQ] clinician report questionnaire, [SRQ] self report questionnaire, [SSI] semi-structured interview

Relational context shorthand: - = mentalizing, / = or, R = Rater, C = Clinician, S = Self/participant, O = other (h = hypothetical & unspecified, k = known, f = friend, a = attachment figure, i = interviewer, p = professional), (d) – can be dependent on type of text rated using measure.

As would be expected, four of these measures also meet the criteria for being broadly distinction focused, and all but two meet the criteria for being broadly aspect focused (The Metacognition Assessment Scale Abbreviated and The Multidimensional Mentalizing Questionnaire). The Metacognition Assessment Scale Abbreviated does not operationalise any of the dimensions linked with internally and externally cued or reactive and reflective mentalizing, with an exclusive and clear focus on self, other, cognitive, and affective dimensions. The Multidimensional Mentalizing Questionnaire does not operationalise any of the dimensions linked with internally and externally cued, or reflective mentalizing, with a focus on self, other, and affective dimensions.

3.3. Broadly distinction focused measures

Seven measures were classed as broadly distinction focused, including a clinician report questionnaire, semi-structured interview and five self report questionnaires. Table 6 gives descriptions and protocols for these measures.

Table 6
Broadly distinction focused measures: descriptions and protocols.

Measure (Included paper for this review)	Authors' description and brief protocol
The Reflective Functioning Questionnaire-Revised (7) (Horváth et al., 2023)	A *self report questionnaire* developed as a shorter version of the Reflective Functioning Questionnaire (8) (below), as one item was not found to be valid in a factor analysis.
The Reflective Functioning Questionnaire (8) (Fonagy et al., 2016)	A *self report questionnaire* easy to administer measure of capacity to interpret the self and others in terms of internal mental states (e.g. feelings, attitudes, desires). 8 items, scale 1–7 (strongly disagree–strongly agree). Scores are recoded to capture high certainty and high uncertainty on two subscales. Overall score is the difference between subscales
Metacognition Assessment Interview (Semerari et al., 2012)	A *semistructured interview* adapted from the MAS-R (above) that evaluates metacognition used during real-life experiences as elicited by the interviewer. Approximately 36 items and prompts ask for a narrative of the most troubling interpersonal experience from the previous six months. The whole interview is scored on 16 facets across 4 subscales: Monitoring, Differentiation, Integration and Decentralisation. Scores are 0–5 (not applicable-poor–sophisticated)
Metacognition Self-Assessment Scale (MSAS) (Pedone et al., 2017)	A *self report questionnaire* developed to evaluate the different functions of metacognition (Monitoring, Integration, Differentiation and Decentration) defined in the Metacognitive Multi-Function Model (Semerari et al., 2003, 2007). 18 items, scale 1–5 (never– almost always)
Theory of Mind Inventory: Self Report-Adult (Hutchins et al., 2021)	A *self report questionnaire* designed to be sensitive to variation in a range of advanced Theory of Mind understandings, and subtlest and difficult-to-detect domains, among those with high cognitive and language abilities. Validated as distinguishing between Autistic and neurotypical participants. 60 items, 20-unit visual continuum anchored by 'definitely not', 'probably not', 'undecided', 'probably', and 'definitely.'
Mentalization Imbalances Scale (Gagliardini et al., 2018)	A *clinician report questionnaire* measuring dimensions of mentalizing: cognitive, affective, others, self, automatic, external. 22 items, scale 0–5 (not descriptive–absolutely descriptive, in the last month). Clinician must have met the selected patient at least four times before assessment.
Empathy Quotient (EQ) (Lawrence et al., 2004)	A *self report questionnaire* designed be sensitive to a lack of empathy as a feature of psychopathology or neurodivergence. 40 items, scale 1–4 (strongly disagree–strongly agree)

Table 7
Construct framework ratings for broadly distinction focused mentalizing measures.

Broadly distinction focused mentalizing measures	Relational context	Construct framework ratings																
		Self	Other	Distinction	Synthesis	Internally cued	Externally cued	Distinction	Synthesis	Cognitive	Affective	Distinction	Synthesis	Reactive	Reflective	Distinction	Synthesis	Uncertainty
The Reflective Functioning Questionnaire (7) [A SRQ]	S-S/hO	Y	Y	Y	N	P	P	N	N	Y	Y	Y	N	Y	Y	Y	N	Y
The Reflective Functioning Questionnaire (8) [A SRQ]	S-S/hO	Y	Y	Y	N	P	P	N	N	Y	Y	Y	N	Y	Y	Y	N	Y
Metacognition Assessment Interview [A SSI]	R-S-S/kO/iO	Y	Y	Y	N	P	Y	P	N	Y	Y	P	P	Y	Y	P	N	Y
The Metacognition Self-Assessment Scale [A SRQ]	S-S/hO	Y	Y	Y	P	P	N	P	N	Y	Y	P	P	P	Y	N	N	Y
Theory of Mind Inventory: Self Report-Adult [A SRQ]	S-S/fO/hO	Y	Y	Y	N	P	Y	Y	N	Y	Y	N	P	Y	Y	N	N	P
The Mentalization Imbalances Scale [A CRQ]	C-S-S/hO	Y	Y	Y	Y	N	Y	N	N	Y	Y	Y	P	Y	Y	N	N	N
Empathy Quotient [A SRQ]	S-S/hO	Y	Y	P	N	N	N	Y	N	Y	Y	Y	N	Y	P	N	N	P

Y = Measured, P = Partially measured, N = Not measured. [A] Also meets criteria for aspect focus. [SRQ] self report questionnaire, [SSI] semi-structured interview, [CRQ] clinician report questionnaire.

Relational context shorthand: - = mentalizing, / = or, R = Rater, C = Clinician, S = Self/participant, O = other (h = hypothetical & unspecified, k = known, f = friend, i = interviewer).

Table 8
Broadly aspect focused measures: descriptions and protocols.

Measure (Included paper for this review)	Authors' description and brief protocol
Modes of Mentalization Scale. (Gagliardini & Colli, 2019)	A *clinician report questionnaire* for assessing prementalizing modalities of thought, with factors: excessive certainty, concrete thinking, good mentalization, teleological thought and intrusive pseudomentalization. 24 items, scale 0–5 (absolutely not descriptive–absolutely descriptive)
The Mentalizing Emotions Questionnaire (Kasper et al., 2024)	A *self report questionnaire* developed to assess an individual's ability to identify, reflect on, and make sense of one's emotional state as well as to communicate one's emotions to oneself and others.16 items, scale 1–7 (never–always)
Observable Social Cognition: A Rating Scale (Healey et al., 2015)	An *informant report questionnaire* assessing domains of social cognition often impaired alongside schizophrenia diagnoses: emotion perception, cognitive rigidity, jumping to conclusions and attributional style. 8 items, scale 1–7 (none–extremely severe) with specific anchor points for each question. Rated over last 7 days based on interactions with person and of observed interactions of person with others
The brief version of the Mentalization Scale (Ments-12) (Stefana et al., 2024)	A *self report questionnaire* and briefer version of the MentS (see Table 4). 12 items, scale 1–5 (completely incorrect–completely correct)
Mentalizing Values Scale (Aival-Naveh et al., 2022)	A *self report questionnaire* measuring the extent to which thinking about internal mental states is valued across different cultures. 12 items, scale 1–5 (completely disagree–completely agree)
Mentalization Questionnaire (MZQ 15) (Hausberg et al., 2012)	A *self report questionnaire* for assessing mentalization intended to differentiate mentalizing of people with psychiatric diagnoses. Factors included avoidance of mentalizing, emotional awareness, psychic equivalence and affect regulation. 15 items, scale 1–5 (no agreement at all–total agreement)
The Certainty About Mental States Questionnaire (Müller et al., 2023)	A *self report questionnaire* assessing the certainty involved in making inferences about the mental states of the self and others. 20 items, scale 1–7 (never–always)
Mentalization Questionnaire (6) (Riedl et al., 2023)	A *self report questionnaire* and briefer version of the MZQ-15 (above). 6 items, scale 1–5 (no agreement at all–total agreement)
Grille de l'élaboration verbale de l'affect (Bouchard et al., 2008)	A *text rating system* used with therapy session transcripts, providing operational descriptions of affect mentalization levels and forms of verbal affect expression of dimensions. Single expressions of affect are categorised as verbal, imagery, motor, or somatic expression and rated from 1–5 (Acting out–Meaning Association). Affect content and valence is recorded. Manualised decision trees and rater training are used assist raters. Overall weighted score is calculated (1–4, with 4 indicating higher quality verbal elaboration of affect).
Metacognition Brief Rating Scale (Pedone & Semerari, 2023)	*Informant report questionnaire* developed from the MSAS (above). 18 items, scale 1–5 (never–almost always)
Interpersonal Reactivity Index (IRI): Perspective Taking subscale (Davis, 1983)	A *self report questionnaire* with subscales that each tap some aspect of the global concept of empathy. 7 items, scale 0–4 (does not describe me well–describes me very well). The full IRI consists of four 7 item subscales
Mental States Measure (Lecours et al., 2000)	A *text rating system* used by therapists to rate mental state elaboration and implicit structural development. Usually used with AAI: an approximately 20 item semi-structured interview about experiences with early attachment figures, family and own children. Transcripts are segmented thematically, then each segment scored 1–6 (concrete thinking–reflective), following 45 hours of training and using a manual. The number of words within each mental state category is calculated and a percentage score generated reflecting mentalizing level
Mind-Reading Belief Scale (Realo et al., 2003)	A *self report questionnaire* indicating general mind-reading belief. 8 items, scale 0–4 (strongly disagree–strongly agree)
Theory of Mind Assessment Scale (Th.o.m.a.s) (Bosco et al., 2009)	A *semi-structured interview* measuring multi-component Theory of Mind, developed with people diagnosed with Schizophrenia. Approximately 37 open-ended questions. For scoring, questions are classified into four scales: I-me (my view of myself), other-self (others' views of themselves), I-other (my view of others), other-me (others' views of me). Each scale has three subscales: awareness (perceive & differentiate mental states), relation (consider causal relations between mental states and behaviours), realization (using strategies to achieve a state). Each question is rated 0–4, ranging from silence or incomprehensible answers to coherent and detailed answers contextualised in personal experience. Raters cannot also be interviewers
Beliefs about Emotions Scale (Rimes & Chalder, 2010)	*Self report questionnaire* for assessing beliefs about the unacceptability of experiencing or expressing negative emotions. 12 items, scale 0–6 (totally disagree–totally agree)
Narrative of Emotions Task (Buck et al., 2014)	A *structured Interview* with 24 questions. For 8 emotions (happy, sad, angry, afraid, suspicious, surprised, guilty and ashamed) participants are asked for definition, narrative account and explanation (e.g. why did that make you feel x?) Each question is scored using 8 scales (Definition of emotion, Presence of narrative, Contextual appropriateness, Causal inferences, Clarity of meaning, Clarity of grammar, Elaboration and Sociality) each rated 0–3, with higher scores indicating better performance. Scores are totalled across all emotions and scales
Mental State Discourse analysis as applied to the Thematic Apperception Test (Lee-Parritz, 2015)	A *text rating system* used with the Thematic Apperception Test (TAT) to evaluate a participant's tendency to use mental state language. The TAT is a set of ambiguous picture cards showing figures in situations. Participants are asked to invent a story around the image, with prompts to refer to what the people think and feel. Transcripts are scored by counting words referencing behavioural, cognitive, affective or affective desire states. Overall score is the ratio of mental state and behavioural language used

(*continued on next page*)

Table 8 (*continued*)

Measure (Included paper for this review)	Authors' description and brief protocol
Social Cognition and Object Relations Scale (SCORS-G) as applied to the Thematic Apperception Test (TAT) (Ridenour et al., 2022)	A *text rating system* used to assess facets of object relations and social cognition. Used with the TAT: ambiguous picture cards showing figures in situations. Participants are asked to invent a story around the image, with prompts to refer to what the people think and feel. The SCORS-G rates narratives from 1–7 (more pathological-healthier, anchor points in manual) on eight dimensions: complexity of representations of people, understanding of social causality, affective quality of representations, emotional investment in relationships, emotional investment in values and moral standards, experience and management of aggressive impulses, self-esteem and identity and coherence of self
Mind Reading Motivation Scale (Carpenter et al., 2016)	A *self report questionnaire* for assessing an individuals' willingness to effortfully engage with other people's perspectives and mental states.13 items, scale 1–7 (disagree completely–agree completely)
Awareness of Independent Learning Inventory (Meijer et al., 2013)	A *self report questionnaire* to measure metacognitive activities of students, e.g., aligning to a learning task, monitoring comprehension, checking learning outcomes, metacognitive curiosity. 45 items, scale 1–7 (not true at all–completely true)
The Moral Metacognition Scale (McMahon & Good, 2016)	A self report questionnaire for assessing metacognition as a domain-specific capacity related to ethical decision making. 20 items, scale 1-6 (Very strongly disagree-Very strongly agree)
Kentucky Inventory of Mindfulness—Describe and Act With Awareness subscales (Baer et al., 2004)	*Self report questionnaire* subscales for assessing the tendency to be mindful, and components of mindfulness: observe, describe, act with awareness and accept without judgement. 39 items, scale 1–5 (almost never–almost always)
Toronto Structured Interview for Alexithymia (Bagby et al., 2006)	A *structured interview* with prompts for assessing difficulties identifying and describing subjective emotional feelings, and an operative thinking style with preoccupation with external events and few fantasies about drives and feelings. 24 items, scored 0–2 (never or rarely present–present most of the time) with model answers used for scoring
Toronto Alexithymia Questionnaire (Bagby et al., 1994)	A *self report questionnaire* for measuring alexithymia, with factors: difficulty identifying feelings, difficulty describing feelings, and externally orientated thinking. 15 items, scale 1–6 (almost always–almost never)
Metacognitions Questionnaire 30 (Cook et al., 2014)	A *self report questionnaire* measuring metacognition from the metacognitive model of psychological disorder, which states that emotional distress is maintained by maladaptive and prolonged patterns of thinking. 30 items, scale 1–4 (do not agree-agree very much)

Table 7 shows the ratings for these measures. Self-other distinction is operationalised at least partially by all seven, and all but one at least partially operationalise cognitive-affective distinction. Four operationalise or partially operationalise internally cued-externally cued distinction and three reactive-reflective distinction. Two operationalise or partially operationalise self/other synthesis, four partially operationalise cognitive-affective synthesis, and only one does not operationalise un/certainty. All measures also meet the criteria for being broadly aspect focused, as expected. All operationalise or partially operationalise self, other, cognitive, affective, reactive and reflective aspects, and five operationalise or partially operationalise both externally cued and internally cued aspects. Two use the context of a specified relationship (interviewer or friend).

3.4. Broadly aspect focused measures

25 measures were classed as aspect focused (in addition to the 15 presented above). These included one clinician report questionnaire, two informant report questionnaires, 15 self report questionnaires, four text rating systems, one semi-structured interview and two structured interviews. Authors' descriptions of included measures broadly classed as examining aspect dimensions, and a brief protocol for each measure, are presented in Table 8.

Table 9 shows the ratings for these measures. The two starred measures in this table capture aspects of mentalizing in a particular context. The Awareness of Independent Learning Inventory measures the rated aspects of mentalizing in a learning context, and the Moral Metacognition Scale in the context of making moral decisions. Of the 25 broadly aspect focused measures, all but one operationalised or partially operationalised mentalizing the self, and all but four mentalizing the other. Only two measures did not operationalise or partially operationalise affective or cognitive aspects of mentalizing. Over half operationalise or partially operationalise externally cued, internally cued, reflective and reactive aspects of mentalizing. Some measures in this group also operationalised distinction and synthesis, with 10 measures operationalising distinction and or synthesis in the self/other domain, three in the internally/externally cued domain, 4 in the cognitive/affective domain, and one in the reactive/reflective domain. Over half of the measures operationalised or partially operationalised the un/certainty dimension. Eight measures operationalised mentalizing in the context of a specific relationship.

3.5. Measures with a more specific or differing focus

The remaining 15 measures were not rated as measuring enough of a breadth of synthesis, distinction or aspect dimensions to fall into any of the categories discussed above, having either a more specific focus, or not operationalising any of the proposed dimensions of mentalising quality. This group includes one text-rating system, one structured interview and 13 self report questionnaires. Authors' descriptions and brief protocols for these measures are presented in Table 10.

Table 11 shows the ratings for these measures. The 10 starred measures in this table operationalise aspects of mentalizing in a particular context. The Porous Theory of Mind Scale measures mentalizing in the context of considering others perceiving one's thoughts, the Metacognition in Self-Control Scale measures mentalizing of one's own self-control conflicts, the Metacognitive subscale of the Cultural Intelligence Scale operationalises use of cultural knowledge in mentalizing,

Table 9
Construct framework ratings for broadly aspect focused mentalizing measures.

Broadly aspect focused measures	Relational context	Construct framework ratings															
		Self	Other	Distinction	Synthesis	Internally cued	Externally cued	Distinction	Synthesis	Cognitive	Affective	Distinction	Synthesis	Reactive	Reflective	Distinction	Synthesis	Uncertainty
Modes of Mentalization Scale [CRQ]	C-S-S/hO	Y	Y	Y	Y	Y	Y	N	N	P	Y	N	N	Y	Y	N	N	Y
The Mentalizing Emotions Questionnaire [SRQ]	S-S/hO	Y	Y	P	P	P	P	N	N	P	Y	N	N	P	Y	N	N	Y
Observable Social Cognition: A Rating Scale [IRQ]	I-S-kO	P	Y	N	N	Y	Y	N	N	P	Y	N	N	Y	Y	Y	N	Y
The Brief Version of the Mentalization Scale [SRQ]	S-S/hO/cO	Y	Y	N	N	Y	P	N	N	Y	Y	N	P	Y	P	N	N	Y
Mentalizing Values Scale [SRQ]	S-S/hO/cO	Y	Y	N	N	Y	Y	Y	N	Y	Y	N	N	P	P	N	N	N
Mentalization Questionnaire (15) [SRQ]	S-S/hO	Y	P	N	N	Y	Y	N	N	P	Y	N	N	Y	P	N	N	Y
Certainty About Mental States Questionnaire [SRQ]	S-S/hO	Y	Y	N	N	P	Y	P	N	Y	Y	N	N	P	N	N	N	Y
Mentalization Questionnaire (6) [SRQ]	S-S/hO	Y	P	N	N	Y	Y	N	N	P	Y	N	N	Y	N	N	N	Y
Grille de l'élaboration verbale de l'affect [TRS]	R-S-S/O (d)	Y	Y	N	N	N	Y	N	N	N	Y	N	N	Y	Y	N	N	N
Metacognition Brief Rating Scale [IRQ]	I-S-S/hO/kO	Y	Y	Y	N	N	N	N	N	Y	Y	P	P	P	Y	N	N	Y
Interpersonal Reactivity Index —Perspective Taking subscale [SRQ]	S-S/hO/fO	Y	Y	N	N	Y	N	N	N	P	Y	N	N	P	P	N	N	Y
Mental States Measure [TRS]	R-S-S/O (d)	Y	P	N	P	P	N	N	N	P	Y	N	N	Y	Y	N	N	N
Mind-Reading Belief Scale [SRQ]	S-S/hO	P	Y	N	N	N	Y	N	N	Y	Y	N	N	Y	N	N	N	P
Theory of Mind Assessment Scale [SSI]	R-S-S/hO/kO	Y	Y	Y	Y	N	N	N	N	Y	Y	P	Y	N	Y	N	N	N
Beliefs about Emotions Scale [SRQ]	S-S/hO	Y	Y	N	Y	P	N	N	N	Y	Y	N	N	P	N	N	N	N
Narrative of Emotions Task [SI]	R-S-S/kO/iO	Y	Y	N	N	P	P	N	N	P	Y	N	N	N	P	N	N	N
Mental State Discourse analysis [TRS]	R-S-hO/iO	N	Y	N	N	N	Y	N	N	Y	Y	N	N	N	P	N	N	N
Social Cognition and Object Relations Scale [TRS]	R-S-S/hO (d)	Y	Y	Y	Y	N	N	N	N	Y	Y	N	P	N	N	N	N	P
Mind Reading Motivation Scale [SRQ]	S-S/hO	P	Y	P	N	N	Y	N	N	Y	N	N	N	P	N	N	N	P
Awareness of Independent Learning Inventory* [SRQ]	S-S/hO	Y	Y	P	N	P	N	N	N	Y	P	N	N	P	P	N	N	Y
The Moral Metacognition Scale* [SRQ]	S-S	Y	N	P	N	P	P	N	N	Y	N	N	N	N	Y	N	N	P
Kentucky Inventory of Mindfulness† [SRQ]	S-S	Y	N	N	N	P	P	N	N	Y	Y	N	N	P	N	N	N	N
Toronto Structured Interview for Alexithymia [SI]	S-S/hO	Y	P	N	N	N	N	N	N	P	Y	N	N	N	Y	N	N	Y
Toronto Alexithymia Questionnaire [SRQ]	S-S	Y	N	N	N	Y	P	P	N	N	Y	N	N	N	P	N	N	N
Metacognitions Questionnaire 30 [SRQ]	S-S	Y	N	N	N	N	N	N	N	Y	P	N	N	P	Y	N	N	P

Y=Measured, P=Partially measured, N=Not measured. *These measures capture aspects of mentalizing in a specific context. †Describe and Act With Awareness subscales. [CRQ] Clinician report questionnaire, [SRQ] self report questionnaire, [IRQ] informant report questionnaire, [TRS] text rating system, [SSI] semi-structured interview, [SI] structured interview.

Relational context shorthand: - = mentalizing, / = or, R = Rater, I = informant, C = Clinician, S = Self/participant, O = other (h = hypothetical & unspecified, k = known, f = friend, i = interviewer, c = close), (d) = dependent of type of text rated using measure.

Table 10
Mentalizing measures with a specific or differing focus: descriptions and protocols.

Measure (Included paper for this review)	Authors' description and brief protocol
The Porous Theory of Mind Scale (van Elk et al., 2020)	A *self report questionnaire* measuring endorsement of the belief that one's thoughts can be directly perceived by others. 4 items, scale 1–5 (I don't agree at all–I very much agree)
Brief-Mentalized Affectivity Scale (Greenberg et al., 2021)	A *self report questionnaire* that is shorter version of the Mentalized Affectivity Scale. 12 items scale 1–7 (Disagree strongly–Agree strongly)
The Metacognition in Self-Control Scale (Bürgler et al., 2022)	A *self report questionnaire* developed to assess trait-level metacognition in self-control. 12 items scale 1–5 (Disagree strongly–Agree strongly)
Metacognitive subscale of the Cultural Intelligence Scale (Ang et al., 2007)	A *self report questionnaire* subscale assessing mental processes that individuals use to acquire and understand cultural knowledge, including knowledge of and control over individual thought processes relating to culture. 4 items (of 20 in the full scale), scale 1–7 (strongly disagree–strongly agree)
Metacognitive Awareness Inventory (Schraw & Dennison, 1994)	A *self report questionnaire* measuring the ability to reflect upon, understand and control one's learning. 12 items with a visual 10cm continuum anchored by True and False
Basic Empathy Scale (Jolliffe & Farrington, 2006)	A *self report questionnaire* designed to measure understanding and sharing in another's emotional state or context, including affect congruence (affective empathy) and the understanding of another's emotions (cognitive empathy). 20 items, scale 1–5 (strongly disagree–strongly agree)
Difficulties in Emotion Regulation Scale—Lack of Emotional Clarity/ Awareness subscales (Gratz & Roemer, 2004)	A *self report questionnaire* subscale for assessing dimensions of emotion dysregulation, with these subscales measuring awareness and understanding of emotions. 20 items (of 41 in the full scale), scale 1–5 (almost never–almost always).
Mindful Attention Awareness Scale (Brown & Ryan, 2003)	A *self report questionnaire* measuring a unique quality of consciousness that is related to a variety of well-being constructs, that differentiates mindfulness practitioners from others, and that is associated with enhanced self awareness. 15 items, scale 1-5 (almost never-almost always)
Metacognitive Activities Inventory (Sandi-Urena, 2009)	A *self report questionnaire* assessing monitoring and evaluation of one's cognitive behaviours in a learning environment, used in chemistry problem solving. 27 items, scale 1–5 (strongly disagree–strongly agree)
Assessment of Cognitive Monitoring Effectiveness (Osborne, 1998)	A *self report assessment* used with routine classroom tests to assess knowledge about, and control of, cognitive processes. Participants indicate answers in tests they believe are incorrect, and a ratio of truly incorrect to falsely believed to be incorrect answers is calculated.
Swanson's 15-item Metacognition Questionnaire (Swanson, 1990)	A *structured interview* with probes to measure knowledge and control over thinking and learning activities during problem solving. 17 items, scored 1–5 (from least to most metacognitive awareness) using model answers
Metacognition Scale (Ford et al., 1998)	A *self report questionnaire* assessing metacognitive activity, e.g., self-monitoring of learning, choice of practice scenarios, and self-evaluation of progress. Used with the Tactical Naval Decision Making System. 52 items, scale 1–5 (strongly disagree–strongly agree)
Mental-Physical Verb Norms (Orr & Gilead, 2022)	A *text rating system* used to assess the use of common verbs reflecting mental or physical activities and occurrences. Dictionary of English verbs scored 0–100 (strictly physical activity–strictly mental activity) that can be applied to any text to produce a score reflecting mental state attribution
Metacognition in Multiple Contexts Inventory. (Allen & Armour-Thomas, 1993)	A *self report questionnaire* measuring the knowledge and control individuals have over their own cognition and learning experiences. 24 items, scale 1–5 (almost never–almost always)
Motivated Strategies for Learning Questionnaire (Pintrich et al., 1993)	A *self report questionnaire* assessing motivation and use of learning strategies by college students. 81 items, scale 1–7 (not at all true of me–very true of me)

and the other starred measures operationalise mentalizing in the context of learning.

Of these 15 measures, over half operationalised or partially operationalised self and cognitive aspects. A small minority of measures operationalised other, affective, internally or externally cued, reactive, reflective, un/certainty, distinction or synthesis dimensions of mentalizing, with most of these ratings being for partial operationalisation. Two measures operationalised mentalizing in the context of a specific relationship.

Three measures were not rated as operationalising any of the proposed dimensions of mentalizing: The Mental-Physical Verb Norms, Metacognition in Multiple Contexts Inventory, and Motivated Strategies for Learning Questionnaire. The Mental-Physical Verb Norms operationalises a very broad measure of the extent of mentalizing present in text by calculating the ratio of mental state verbs to physical activity verbs. This measurement paradigm does not operationalise any of the dimensions within the proposed framework. It could be argued that this measure is of the extent of presence or absence of mentalizing, perhaps more akin to the original definition of theory of mind. The Metacognition in Multiple Contexts Inventory and Motivated Strategies for Learning Questionnaire are both used in the context of learning, and have a non-mentalizing focus on behaviours and strategies used when learning.

4. Discussion

4.1. Initial Implications

This systematic review has summarised the text-based, English language general population measures of mentalizing that are available to date, and the components of mentalizing they operationalise. These

Table 11
Construct framework ratings for measures with a more specific or differing focus.

Measures with a specific or differing focus	Relational context	Construct framework ratings																
		Self	Other	Distinction	Synthesis	Internally cued	Externally cued	Distinction	Synthesis	Cognitive	Affective	Distinction	Synthesis	Reactive	Reflective	Distinction	Synthesis	Uncertainty
The Porous Theory of Mind Scale* [SRQ]	S-S/hO	P	P	Y	N	Y	P	N	N	P	P	N	N	N	N	N	N	N
Brief-Mentalized Affectivity Scale [SRQ]	S-S	Y	N	N	N	N	N	N	N	N	Y	N	Y	P	Y	N	Y	N
The Metacognition in Self-Control Scale* [SRQ]	S-S	Y	N	N	N	N	N	N	N	Y	N	N	N	N	Y	N	N	N
Metacognitive subscale of the Cultural Intelligence Scale* [SRQ]	S-S	Y	N	N	N	P	P	N	N	Y	N	N	N	P	N	N	N	P
Metacognitive Awareness Inventory* [SRQ]	S-S/pO	Y	P	N	N	P	N	N	N	Y	N	N	N	N	P	N	N	N
Basic Empathy Scale [SRQ]	S-hO/fO	P	Y	N	N	N	N	N	N	N	Y	N	N	N	N	N	N	P
Difficulties in Emotion Regulation Scale† [SRQ]	S-S	Y	N	N	N	P	N	N	N	N	Y	N	N	N	Y	N	N	N
Mindful Attention Awareness Scale [SRQ]	S-S	Y	N	N	N	N	N	N	N	Y	Y	N	N	N	N	N	N	N
Metacognitive Activities Inventory* [SRQ]	S-S	Y	N	N	N	N	N	N	N	Y	N	N	N	N	P	N	N	N
Assessment of Cognitive Monitoring Effectiveness* [SRQ]	S-S	P	N	N	N	N	N	N	N	P	N	N	N	P	N	N	N	N
Swanson's 15-item Metacognition Questionnaire* [SI]	S-hO	N	Y	N	N	N	N	N	N	Y	N	N	N	N	N	N	N	N
Metacognition Scale* [SRQ]	S-S	P	N	N	N	N	N	N	N	P	N	N	N	N	N	N	N	N
Mental-Physical Verb Norms [TRS]	(d)	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Metacognition in Multiple Contexts Inventory* [SRQ]	N/A	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N
Motivated Strategies for Learning Questionnaire* [SRQ]	N/A	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N	N

Y=Measured, P=Partially measured, N=Not measured. *These measures capture aspects of mentalizing in a specific context. [SRQ] self report questionnaire, [SI] structured interview, [TRS] Text rating system. †Lack of Emotional Clarity/Awareness subscales.

Relational context shorthand: - = mentalizing, / = or, I = informant, S = Self/participant, O = other (h = hypothetical & unspecified, f = friend, p = professional)

components were not identical for any of the reviewed measures (excluding shorter and longer versions of the same measure, and measures not operationalising any of the proposed dimensions). This suggests that, as hypothesised, careful selection of appropriate measures on the basis of mentalizing contexts and dimensions of interest is key in ensuring clarity of measurement and conclusions in mentalizing research (Flake & Fried, 2020). A measure selection guide based on the results of this review has been created to support researchers in this process, see Supplementary Material.

4.2. Relational contexts of mentalizing measures

The current paucity of attempts to relationally contextualise mentalizing research is striking in the results of this review. A lack of ecological validity in measurement of mentalizing has been discussed in the literature (Dziobek et al., 2011; Nijhof, 2023) with criticism of measures involving reflecting on or imagining mentalizing, rather than prompting mentalizing in live situations (sometimes termed "online mentalizing"). The findings of this review lend support to this criticism of existing measures, with mentalizing largely assessed in the context of hypothetical and unspecified relationships. However, these criticisms have generally focused on procedural rather than relational context, with use of video prompts or interactive scenarios proposed as more ecologically valid. This is despite the persisting relational context of strangers, who may lack salience as meaningful individuals, even in these live situations (Canty et al., 2017; Quek et al., 2019). Difficulties mentalizing in formation of longer-term and close relationships are more often clinically relevant (American Psychiatric Association, 2013; Leichsenring et al., 2024; World Health Organisation, 2022), than immediate difficulties mentalizing strangers (with the possible exception of some mentalizing in psychosis (Lysaker et al., 2014)). This review highlights a lack of continuity between the relational context of these difficulties and that of the measures used to investigate them.

The only two measures that examine mentalizing in the context of attachment relationships—demonstrably central to certain clinically relevant mentalizing profiles (Fonagy & Luyten, 2009; Katznelson, 2014)—are both interviews using the Reflective Functioning Scale text rating system (the RFS; Fischer-Kern et al., 2010; and BRFI; Rutimann & Meehan, 2012). This is generally agreed to be a gold-standard measurement method (Bouchard et al., 2008), but is also resource intensive. The finding here of a plethora of mentalizing measures in the context of relating to unspecified and hypothetical others contrasts with the lack of self report measures applied in the context of attachment relationships, which is feasible. For example, item 1 from the Reflective Functioning Questionnaire (Fonagy et al., 2016) 'People's thoughts are a mystery to

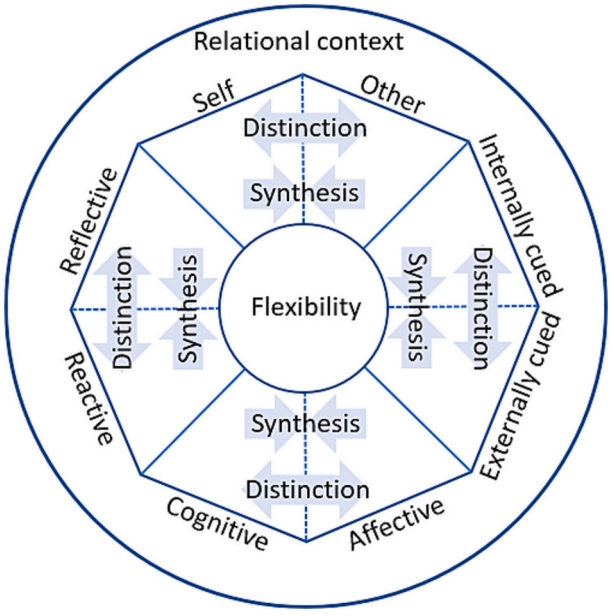

Fig. 3. Revised dimensional construct framework of mentalizing.

me' could be adapted to attempt to capture an attachment context: 'My parents'/caregivers' thoughts were a mystery to me growing up'. A further differentiation may pertain to whether mentalizing in relational contexts is probed with the affective schema of that relationship being activated according to the biobehavioural switch model, i.e. following the activation of the attachment system (Fonagy & Luyten, 2009; Nolte et al., 2013) (perhaps best approximated by reactive mentalizing) or not (reflective mentalizing). This will allow for operationalisations of measures that may compare "baseline" mentalizing capacities and those associated with "hot" social cognition. There are suggestions in the literature of other relatively unexplored clinically relevant relational contexts (Aival-Naveh et al., 2019; Chapple et al., 2021; Taubner et al., 2011). Development of more relationally specific measures could begin to illuminate some of these areas.

Within the framework, all relational contexts of measurement were assessed and reported as far as possible, including relating and mentalizing between raters or interviewers and participants that could potentially impact measurement outputs. Advances in mentalizing or theory of mind research in children and non-human species have been made due to acknowledging researchers themselves as mentalizing agents with a context, with a potential for bias towards assuming their own mentalizing to be normative or superlative (Bräuer et al., 2020; Call & Tomasello, 2008; de Waal & Ferrari, 2010; Gopnik & Wellman, 1992). The potential impacts of researchers' own mentalizing on measurement were not discussed in the reviewed papers, which perhaps speaks to the positioning of much mental health and cognitive psychology research within a positivist and individualistic (as opposed to social constructionist) tradition (Kutney, 2006; Pilgrim, 2014; Williams, 2015). This theoretical and philosophical perspective could also explain the research gap in specific relational contexts of mentalizing found here, as mentalizing deficits (as with many other mental health difficulties) have been initially conceptualised primarily as an individual rather than interpersonal trait. The prevalent measurement of mentalizing the self found by this review may also speak to this. This has been highlighted previously as a limitation by cross-cultural and feminist perspectives on the mentalizing literature (Aival-Naveh et al., 2022; Linehan, 1993). Critically applying mentalizing theory to researchers and research methods (for example considering relationships within which mentalizing takes place in research, or whether certain mentalizing styles might be more prevalent in certain research fields) could ensure mentalizing theory and measurement are based on robust assumptions.

Differences in relational contexts of measurement found here could also be used to consider apparent contradictions in the existing literature. Broad examples were discussed in the introduction. A specific illustrative example is the 'empathy paradox' described in studies of BPD. Individuals with BPD diagnoses have been found to exhibit increased sensitivity to others' mental states, which has been argued to contradict the theoretical centrality of ineffective mentalizing to BPD (Dinsdale & Crespi, 2013; Jańczak, 2018). This apparent contradiction can be reconciled by considering mentalizing as dynamic and relationally dependent, as described in the mentalization-based model of BPD (Fonagy & Luyten, 2009). This model accounts for the changes in mentalizing following the relationally dependent activation of the attachment system, and heightened sensitivity of the attachment system for many individuals with this diagnosis. Generation of hypotheses like this that account for relational context can create new understandings of the clinical relevance of mentalizing from previous inconsistencies in the literature (Katznelson, 2014; Nijhof, 2023).

4.3. Dimensions of mentalizing

That it was feasible to cluster measures by synthesis, distinction and aspects, but not by domain (self-other, cognitive-affective etc.), demonstrates that measures reviewed tended to capture a breadth of mentalizing dimensions. This suggests that mentalizing is usually implicitly conceptualised as broad and multidimensional. Many measures were solely aspect focused, but synthesis and distinction measures overlapped and nearly all had an aspect focus. The hierarchical structure of these clusters reveals how mentalizing might be seen and measured as an emergent property, that builds in complexity, of interactions between aspects.

That all proposed dimensions were used by at least one measure indicates that they are all already being viewed as significant within mentalizing research. The domains of reflective and reactive, and internally and externally cued mentalizing, while well supported in the literature (Fonagy & Luyten, 2009; Kivity et al., 2024; Luyten et al., 2012; Luyten et al., 2020), are often not the focus of definitions of mentalizing and related concepts (see Table 1) – likely also due to limitations in which dimensions lend themselves to be operationalization by text-based instruments. The prevalence of self, other, cognitive and affective mental states in the review findings confirm that these are better established and more well-used concepts than internal and external cueing or reactive and reflective mentalizing. More exploration of the less prevalent domains could be beneficial.

The complete diversity of combinations of mentalizing dimensions operationalised by measures, and the loose criteria needed in order to cluster the results, demonstrate that measures also operationalise components of mentalizing specifically and selectively. This was expected, and the need to specify the different constructs underlying measures was part of the rationale for the review. This specification could now be used to compare previous studies utilising different measures. For example, without considering dimensions of mentalizing, the non-significant association between depressive symptoms and RFS score (Taubner et al., 2011) and significant association between depressive symptoms and Th.o.m.a.s. score (Colle et al., 2019) could be interpreted as contradictory. The apparent contradiction here could instead be used alongside the framework to generate hypotheses about the mentalizing profile of depression. The Th.o.m.a.s. (which found mentalizing differences in depression) is rated by the proposed framework as operationalising cognitive and affective aspects and their distinction more completely than the RFS (which did not). A lack of cognitive-affective distinction may therefore be a particular part of mentalizing profiles in depression, perhaps connected to negative valance biases found across cognition in this group (Kircanski et al., 2012).

The range of dimensions being measured in the review reveals active curiosity in mentalizing research towards a wide range of dimensional profiles of mentalizing.

4.4. Mentalizing measures construct framework

All of the proposed dimensions are used by at least one measure, suggesting none of these are redundant (see Table 3). This was not necessarily expected, as the framework split the 'balance' dimensions underpinning effective mentalizing into distinction and synthesis across all domains in a novel way (Fig. 1). For example, self-other distinction (often referred to as differentiation) is a well-established concept (Bender & Skodol, 2007; De Meulemeester et al., 2021), and originally had been proposed as the only balance dimension for this domain. However self-other synthesis was also operationalised frequently by the reviewed measures, supporting its inclusion in the framework. That all measures reviewed tended to capture a breadth of mentalizing dimensions might suggest partial overlap or frequent coincidence of some of the dimensions, with many measure items operationalising more than one dimension (see Appendices B & E). Possible redundancy of overlapping dimensions should be investigated to improve the framework, but given the lack of total overlap between measures, this may be explained by the interwoven nature of the dimensions of mentalizing.

Distinction and synthesis dimensions were operationalised in a minority of measures reviewed here, which would perhaps be expected as these are novel ideas for dimensions. However, the original four-dimensional model of mentalizing suggests that measuring balance between these pairs is crucial (Fonagy & Luyten, 2009; Gagliardini et al.,

2018). That the majority of measures operationalise dimensions at the aspect level, which is more concrete and well defined, may speak to the conflation of aspects and balance in the four-dimensional model, again potentially supporting use of the proposed framework.

4.5. Limitations

This review has used a novel framework for rating and comparing measures of mentalizing. While the literature and evidence base for dimensions of mentalizing were used to develop the framework, and it shows promising initial inter-rater reliability and feasibility, there has not yet been any direct empirical evidence in support of this framework. There was also no consideration of the quality of measures in terms of comprehensively reviewed validity and reliability, or in terms of the type of measurement used. For example, when rating measures, existing analyses of structural validity of these measures were not considered. Concerns regarding whether self report is appropriate for assessment of mentalizing abilities, given the likely impairment of self-assessment if mentalizing is impaired, have been raised previously (Müller et al., 2023; Rumeo, 2022). There was also no consideration of the differences between measures in validated definition of scores and cut-offs for clinically relevant mentalizing styles or deficits. Some measures have clear and well-validated clinical indicators, which is also an important consideration when selecting a measure.

During rating there were difficulties clearly establishing the dimension of certainty-uncertainty as operationalised within measures. Rater discussions suggested this construct may not be unidimensional, as has been highlighted by the authors of measures utilising this dimension, (Fonagy et al., 2016; Fossati et al., 2016; Müller et al., 2023). For future use of the framework, a 'flexibility' dimension (see Fig. 3), that captures understanding of changeability of mental states across perspectives and/or time, is proposed as a clearer single dimension. This represents the concept of over-certainty described in the literature. The concept of under-certainty as a failure of mentalizing is arguably captured by failures of mentalizing across the eight aspects.

During rater discussions it was noted that the majority of disagreements were resolved by changing from a 'Not measured' to a 'Measured or Partially measured' rating, as explicit evidence was always provided to support 'Measured or Partially measured' ratings. Not measured ratings were often given on the basis of a lack of evidence, and this limitation in support of 'Not measured' ratings should be considered during use of these ratings. Raters were not trained in use of the majority of these measures, which is another limitation of the review.

4.6. Research recommendations

Following the initial finding that use of this framework is feasible with acceptable levels of inter-rater reliability, replication of this finding and use of the framework in other contexts is recommended. For example, ratings of the measures reviewed here could be generated by raters trained in use of the measures that require this. Mentalizing measures that were excluded from this review could also be rated, to examine the dimensions or mentalizing operationalised by measures that use experimental paradigms (including those assessing implicit mentalizing, i.e. without awareness), and those that are designed for use with specific populations, such as children and young people or clinical groups. Exploration of possible redundancy due to overlap between dimensions would help to ensure that the framework is theoretically robust.

It is hoped that this review might contribute to enabling measure selection specific to mentalizing constructs used in future studies (see Supplementary Material for a guide to measure selection on this basis). The relative sparsity of measures rated here as operationalising reactive, reflective, internally cued and externally cued dimensions of mentalizing, and of mentalizing measurement in specific relational contexts, could also inform beneficial areas for future measure development, with less resource-intensive measures of mentalizing in attachment contexts being a particular area for development. This may support ecological validity of these types of research.

The framework and findings here around construct validity of particular measures could be used to inform reviews or experimental studies of mentalizing in particular clinical groups. Improved clarity of the dimensions and relational context of mentalizing measures could support clearer hypothesis formation and testing and clarify the dimensions and context of mentalizing involved in hypotheses of previous studies. This could be particularly fruitful in connecting clinical presentations and mentalizing profiles. Using this framework to support clearer exploration of the clinically relevant dimensions of mentalizing could help elucidate how mentalizing might operate as a transdiagnostic mechanism, and support appropriately targeted treatment development and evaluation strategies.

5. Conclusion

This study has demonstrated that the constructs underlying mentalizing measures can be categorised and compared, and vary markedly between existing measures. The divergence of construct validity in mentalizing measurement may account for contradictory findings in the mentalizing literature. Explicit and transparent construct definition, including relational context and dimensions of interest, is recommended when undertaking new mentalizing research or reviewing the literature.

Funding statement

PF is in receipt of a National Institute for Health Research (NIHR) Senior Investigator Award (NF-SI-0514-10157), and was in part supported by the NIHR Collaboration for Leadership in Applied Health Research and Care (CLAHRC) North Thames at Barts Health NHS Trust.

Data statement

PRISMA 2020 guidelines for systematic reviews were adhered to (Page et al., 2021). Data and research materials are available at https://osf.io/cwd6z/overview?view_only=f8287943cd4e4923aea54795340fd8c4. The protocol for this review was registered with Open Science Framework (Hawksdale, 2023) before starting data extraction. It includes the Preferred Reporting Items for Systematic Reviews and Meta-analysis Protocols (Moher et al., 2015).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Evidence for mentalizing measure review ratings

Measure	1. Reflective Functioning Rating Scale
Self	S/he recognizes the role his/her own mental states might have had on a behavior (i.e. “because I think perhaps I was insecure I decided to…”).
	8. When asked to reflect on his/her motivations and/or the motivations of others, s/he often perceives these questions as an attack (i.e. “How the hell should I know why he did that?”).
Other	S/he appropriately normalizes the experience(s) of others. (i.e. "It’s understandable that she…")
	S/he recognizes that different people may view a given situation differently.
	His/her attributions of other people’s behavior are often marked by “splitting”, where others are portrayed as either all good or all bad.
	S/he is often accurately attuned to the therapist’s emotional reactions to what is occurring in session.
	S/he often expresses certainty about what other people are thinking or feeling.
Distinction	S/he often fails to recognize that his/her perception of an event is not definitive, and that there may in fact be other explanations.
	S/he often assumes knowledge on the part of the therapist which s/he should know the therapist has no access to.
	S/he recognizes that the therapist may not necessarily share his/her perspective.
	4. S/he recognizes that one can never be sure of what motivates another person’s behavior.
	36. S/he recognizes that the therapist may not necessarily share his/her perspective.
	S/he makes spontaneous efforts to clarify confusing aspects of his/her narrative in session.
Synthesis	S/he recognizes that individuals within relationships have a mutual/bidirectional effect upon one another. (i.e. The therapist’s impact on the patient, as well as patient’s impact on the therapist.)
	S/he has difficulty reflecting on how others might perceive him/her.
	S/he recognizes the powerful influence that families have on each others’ thoughts and feelings.
	10. S/he often over-estimates how much he/she has influenced the behaviors of others.
	28. S/he is able to reflect on how his/her behavior may have influenced the behavior of another person.
Internal cues	2. S/he acknowledges that one can be feeling different inside from what emotion is being overtly displayed. (i.e. I felt angry, but I was smiling while talking to you..”).
	5. S/he does not recognize that strong emotions may cloud one’s perception of another person’s motivations.
External cues	S/he often employs literal, concrete thinking.
	22. S/he often normalizes painful experiences in terms of reference to cultural trends or sociological explanations. (i.e. "Well that’s just how it was in the 70’s...")
Distinction	2. S/he acknowledges that one can be feeling different inside from what emotion is being overtly displayed. (i.e. I felt angry, but I was smiling while talking to you..”).
	S/he recognizes that mental states are susceptible to disguise. (i.e. “I felt sad, but I would never show that to him…”)
	S/he recognizes that his/her feelings about a situation may be unrelated to the actual (observable) situation itself.
	S/he recognizes that one can never be sure of what motivates another person’s behavior.
Synthesis	S/he recognizes that his/her perception of an event might have been distorted by what he/she was feeling or thinking at the time.
Cognitive	S/he seems to make accurate attributions of what may have led to his/her thoughts and feelings.
	45. S/he tends to become overtly defensive when asked to reflect on his/her behavior. (ex. “How the hell should I know why I did that?!”).
Affective	S/he recognizes what would be a commonly expected reaction in a given situation (i.e. “He got angry at me , which is reasonable given the fact that…”).
	S/he seems to make accurate attributions of what may have led to his/her thoughts and feelings.
	S/he often side-steps questions about what may have caused a particular feeling.
	S/he is often accurately attuned to the therapist’s emotional reactions to what is occurring in session.
Distinction	N
Synthesis	S/he does not recognize that strong emotions may cloud one’s perception of another person’s motivations.
	S/he recognizes that people may alter their thinking in order to reduce negative feelings. (i.e. "I tend to forget events that make me upset.")
	S/he does not recognize that some mental states can be defensive or serve as a buffer against uncomfortable or unwanted thoughts/feelings.
	14. S/he can grapple with complex thoughts and feelings, even when dealing with emotionally charged subject matter (such as 49. S/he is often accurately attuned to the therapist’s emotional reactions to what is occurring in session. the death of a loved one).
	39. S/he recognizes that his/her perception of an event might have been distorted by what he/she was feeling or thinking at the time.
Reactive	S/he often gives the impression of thinking spontaneously and vividly about people’s thoughts and feelings.
	45. S/he tends to become overtly defensive when asked to reflect on his/her behavior. (ex. “How the hell should I know why I did that?!”).
Reflective	S/he tends to look back on his/her childhood with insights gained through development. (i.e. "Looking back at it now as an adult I can see that…")
	When asked to reflect on his/her motivations and/or the motivations of others, s/he often perceives these questions as an attack (i.e. “How the hell should I know why he did that?”).
	S/he tends to become overtly defensive when asked to reflect on his/her behavior. (ex. “How the hell should I know why I did that?!”).
Distinction	N
Synthesis	N
Uncertainty	S/he recognizes that one can never be sure of what motivates another person’s behavior.
	S/he recognizes that people’s views are often shaped by the stage of life one is in. (i.e. “Now that I’m middle aged my concerns are different from when I was 20.)
	S/he recognizes the limitations of one’s own insights (i.e. “I think I might have felt…”).
	50. S/he often expresses certainty about what other people are thinking or feeling.
	S/he acknowledges that feelings and attitudes are not set in stone, but rather may change or evolve based on personal experiences..
	S/he often expresses certainty about what other people are thinking or feeling.

Measure	**2. The Mentalization Scale**
Self	I can easily describe what I feel
	When someone annoys me I try to understand why I react in that way.
Other	I can recognise other people’s feelings
Distinction	I can sympathise with other people’s feelings
	I do not like to waste time trying to understand in detail other people’s behavior
	While people talk about their feelings and needs my thoughts often drift away.
Synthesis	N would need a higher order question, e.g. about understanding others thoughts/feelings about self
Internal cues	I can easily describe what I feel
	When I get upset I am not sure whether I am sad, afraid, or angry.
External cues	I find it important to understand reasons for my behaviour (?M although might not be mental state reasons)
	When I make conclusions about other people’s personality traits I carefully observe what they say and do.
Distinction	N
Synthesis	I can make good predictions of other people’ s behaviour when I know their beliefs and feelings
Cognitive	I do not like to think about my problems
	I do not want to find out something about myself that I will not like

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Affective	I can easily describe what I feel
	I can recognize other people's feelings
Distinction	I can sympathise with other people's feelings
Synthesis	I can make good predictions of other people' s behaviour when I know their beliefs and feelings
	To understand someone's behavior, we need to know her/his thoughts, wishes, and feelings.
	I do not like to think about my problems.
Reactive	often I cannot explain, even to myself, why I did something
	Sometimes I can understand someone's feelings before s/he tells me anything.
Reflective	To understand someone's behaviour, we need to know her/his thoughts, wishes, and feelings
	When someone annoys me I try to understand why I react in that way.
Distinction	when someone annoys me I try to understand why I react in that way
Synthesis	would need a question more explicitly asking about the weighing up and use of both reactionary and reflective mental states
Uncertainty	I am often confused about my exact feelings
	When I get upset I am not sure whether I am sad, afraid, or angry.

Measure	**3. Metacognition Assessment Scale (MAS-R)**
Self	UM1 COGNITIVE IDENTIFICATION the person is able to distinguish and differentiate his/her own cognitive operations (e.g. remembering, imagining, having fantasies, dreaming, desiring, deciding, foreseeing and thinking). UM2 EMOTIONAL IDENTIFICATION the person is able to define, distinguish and name his/her own emotional states.
Other	UOM1 COGNITIVE IDENTIFICATION the person is able to define and distinguish the others' cognitive operations (e.g. remembering, imagining, having fantasies, dreaming, desiring, deciding, foreseeing and thinking). UOM2 EMOTIONAL IDENTIFICATION the person is able to define and distinguish the others' emotional states.
Distinction	D The person is able to describe the other's mental state forming hypothesis which are independent from his/her own perspective and from his/her own involvement in the relationship.
Synthesis	Understanding of other's mind: relating variables explanation, example: "He thought I wanted to stop him proposing his project because in the past I've disagreed with him. That's why he got worried and came to ask me what my intentions were. I explained to him that I didn't want to stop him and was on his side, and this reassured him".
Internal cues	N
External cues	Emotional identification: This refers to the ability to describe one's own emotions with language more or less capable of transmitting their variety, nuances and intensity. For example, one can pass from a linguistically limited description of a single emotion (e.g. "I'm ashamed"), through a richer use of language to describe the same emotion (e.g. "I'm ashamed, I'm very embarrassed, I'm almost trembling, I blush for the slightest reason and I walk with my eyes to the ground"), to the ability to describe a wide range of emotions (e.g. "I feel guilty", "I'm anxious", "I'm angry", etc.) UM4 the person recognises his/her thought as subjective, his/her opinions and forecasts as hypotheses, considering the possibility they change as contexts change and time passes (including the ability to take a critical distance from own beliefs). Thoughts are not considered reality per se and ideas or wishes cannot influence directly events or change reality.
Distinction	Emotional identification: This refers to the ability to describe one's own emotions with language more or less capable of transmitting their variety, nuances and intensity. For example, one can pass from a linguistically limited description of a single emotion (e.g. "I'm ashamed"), through a richer use of language to describe the same emotion (e.g. "I'm ashamed, I'm very embarrassed, I'm almost trembling, I blush for the slightest reason and I walk with my eyes to the ground"), to the ability to describe a wide range of emotions (e.g. "I feel guilty", "I'm anxious", "I'm angry", etc.) UM4 the person recognises his/her thought as subjective, his/her opinions and forecasts as hypotheses, considering the possibility they change as contexts change and time passes (including the ability to take a critical distance from own beliefs). Thoughts are not considered reality per se and ideas or wishes cannot influence directly events or change reality.
Synthesis	""He's ashamed, he's very embarrassed, he's almost trembling, he blushes for the slightest reason and walks with his eyes to the ground"
Cognitive	UM1 COGNITIVE IDENTIFICATION the person is able to distinguish and differentiate his/her own cognitive operations (e.g. remembering, imagining, having fantasies, dreaming, desiring, deciding, foreseeing and thinking).
Affective	UM2 EMOTIONAL IDENTIFICATION the person is able to define, distinguish and name his/her own emotional states.
Distinction	Mastery, third level strategies: a) a rational criticism of the beliefs underlying a problematic state, like "I know I tend to be afraid of fainting when I get on a bus but I realise that it isn't true and depends on my usual sensation of being constrained, which makes me feel suffocated. I can overcome it because I've grasped that I won't really faint."
Synthesis	UM6 the person is able to describe in a coherent narrative the cognitive and emotional aspects of his/her own states of mind and how they were changing during time, grasping links and causal relations that promoted changes.
Reactive	N
Reflective	M1 The person discusses his own behaviour and psychological processes and states not as simple matter-of-fact data but as tasks to be done and problems to be solved, defining the terms of the problem in a plausible way and adopting an active problem-solving stance
Distinction	"I had a terrible dream in which I fainted. When I woke up, I still had this fear of falling down senseless. Then I realised I was awake and that it had just been a bad dream." Mastery, third level strategies: a) a rational criticism of the beliefs underlying a problematic state, like "I know I tend to be afraid of fainting when I get on a bus but I realise that it isn't true and depends on my usual sensation of being constrained, which makes me feel suffocated. I can overcome it because I've grasped that I won't really faint."
Synthesis	N
Uncertainty	UM4 the person recognises his/her thought as subjective, his/her opinions and forecasts as hypotheses, considering the possibility they change as contexts change and time passes (including the ability to take a critical distance from own beliefs). Thoughts are not considered reality per se and ideas or wishes cannot influence directly events or change reality.

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
	differentiation explanation in manual: With this we are aware that our representations are subjective and not a perfect reflection of reality.

Measure	**4. Metacognition Assessment Interview**
Self	1. Considering this situation, how were you feeling? 4. which were your thoughts 1. Section for Monitoring: composed of 9 questions designed to identify the different components of a mind state. The purpose of these questions is to assess the subject's ability to monitor the emotions and thoughts that make up their own mental states.
Other	27. You told me that___ (name the protagonist of the story) had a significant role in this story. I would like you to take his/her point of view. In your opinion, how did the other person emotionally percept the situation? 30. What could he / she have thought? Re interviewer: Integration descriptor 4: The subject rebuilds and describes the relevant aspects in the narration (without excesses or narrative errors) of their mental functioning without getting lost in irrelevant details for the communicative purpose, conveying the interlocutor a sense of order and clarity Which kind of emotions did he/she experience?
Distinction	Decentralism descriptor 3&4: - The subject understands that same events can have a different meaning and a different significance for others. - The subject evaluates and understands others' psychological processes, regardless of their relationship with him/her. As opposed to: the subject considers himself/herself in the middle of the other's intentions, thoughts and expectations. The other's psychological processes seem to revolve around him/her. Decentralisation section: composed of 9 questions aiming to stimulate the patient to describe the mental states of a person, who has played a prominent role in the story evoked spontaneously in the first section. The aim is to assess the ability to understand that others are acting on the basis of goals, beliefs, values, criteria and hierarchies of importance that may be different from our own and independent of the relationship they have with us.
Synthesis	N
Internal cues	From monitoring scoring section: "Another possibility of poor monitoring is that the patient is unable to describe clearly the emotions without getting back to somatic descriptions or related behavior."
External cues	From monitoring scoring section: "Another possibility of poor monitoring is that the patient is unable to describe clearly the emotions without getting back to somatic descriptions or related behavior." Decentralism descriptor 1: The subject describes the other person's psychology in a clear, plausible manner, without recourse to stereotypes and clichés. As opposed to: {the other is described in behavioral terms or solely based on the thoughts and emotions he/she clearly expressed.} There is no effort to understand the thought processes and the connections between thoughts and emotions that were the basis of his behavior, or other processes are inferred in a stereotyped way on the basis of the role that the other had in the episode, or the subject makes inferences from a description of the mental state in an implausible manner or not fully argued on the basis of the reported events.
Distinction	Generally, not mental states? Differentiation descriptor 4: The subject can distinguish between different kinds of representations such as: dream, fantasize, imagine, think, assume, expect, believe, remember, distinguishing them according to the alleged relationship with the outside world. As opposed to: He/She shows the feeling that external events are unreal or that reality is inconsistent, or he/she is unsure whether certain memories relate to real events or fantasies, or, conversely, perceives images and patterns with extreme intensity as if they were actual events.
Synthesis	N
Cognitive	"Can you tell me, what - from a psychological point of view -was the worst event or interpersonal situation, that you had to face in the last 6 months?" Possibly a relational episode, that means an episode that involved another person. Which were your thoughts?
Affective	"""Can you tell me, what - from a psychological point of view -was the worst event or interpersonal situation, that you had to face in the last 6 months?"" Possibly a relational episode, that means an episode that involved another person. Which emotions did you experience?
Distinction	Differentiation descriptor 4: The subject can distinguish between different kinds of representations such as: dream, fantasize, imagine, think, assume, expect, believe, remember, distinguishing them according to the alleged relationship with the outside world.
Synthesis	Integration descriptor 1: The subject shows a coherent and comprehensible connection between thoughts, emotions, events and behaviors.
Reactive	Why, do you think did you react in the first way? And why in the second way? How did you react in that moment? Which were your emotions in that circumstances? What about your thoughts?
Reflective	Differentiation descriptor 3: The subject reflects upon and evaluates events. (as opposed to the tendency to act impulsively) So, you told me you were feeling... (referred emotion) When did your state of mind change?
Distinction	10. So, you told me you were feeling... (referred emotion), When did your state of mind change? 11. How did it change? 12. What do you think made it change? 19. Considering this now, is there anything that helped you change your point of view compared to 6 months ago? Q25: So, sometimes you react ___ (first example, the interviewer sums up the individual's habitual state, using the closest terminology to the one of the individual), while other times you react ___ (episode number 2, the interviewer sums up the new situation he obtained after his question). How do you explain this difference? example: "Actually, at work I could never act instinctively, the risks would be too high .. I can say that the difference depends if I consider the other responsible for some wrong action .. in my brother's case, his behavior was clearly the result of his being so shallow and uncaring, whereas in my boss' case, the situation was less clear and it wouldn't have made no sense to externalize anger instinctively, without having first understood the real dynamics .. "
Synthesis	N

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Uncertainty	13.How deep did you believe, in that moment, that… (referred thought)? 15. Do you think, it was possible to have a different interpretation of the occurring facts? Integration descriptor 2: The subject describes the transitions between different states of mind and explains the reasons

Measure	**5. Brief Reflective Functioning interview**
Self	Do you have any thoughts how your relationship came to be that way? 8. Why do you think you decided to talk about this parent today?
Other	How do you think your parent came to be that way (that kind of person)? 2. Why do you think your father/mother became the way he/she is now?
Distinction	codes for errors e.g. self-serving/distorted 1B
Synthesis	codes statements about impact of self & others mental states on each other
Internal cues	Asks about thoughts
External cues	codes for behaviours related to mental states e.g. B5 5. Evaluating mental states from point of view of its impact on behaviour of self and/or other
Distinction	2. Feelings concerning a situation may be unrelated to observable aspects of it
Synthesis	codes for integration of different cues, e.g. B2, B5, B4 3. Recognition of diverse perspectives 4. Taking account of one's mental state in interpreting others' behaviour 5. Evaluating mental states from point of view of its impact on behaviour of self and/or other
Cognitive	Do you have any thoughts how your relationship came to be that way? Can you give me a specific memory of a situation that you experienced together with your father/mother between the ages of 5 and 12? (If a general memory is recalled, e.g. "We always….", then ask for one memory that is special or specific; something that happened once.)
Affective	3. Can you tell me something about your relationship with your father/mother – what is it like?
Distinction	N
Synthesis	c3. Revising thoughts and feelings in light of adult understanding
Reactive	B7? or no because not timed or explicit question about immediate response 7. A freshness of recall and thinking about mental states
Reflective	Can you tell me why you chose to talk about this parent (today)? How do you think your parent came to be that way (that kind of person)? Do you have any thoughts how your relationship came to be that way? 2. Why do you think your father/mother became the way he/she is now?
Distinction	N
Synthesis	N
Uncertainty	

Measure	**6. Mentalized Affectivity Scale**
Self	I try to understand the complexity of my emotions. 22. I often know the reasons why I feel the emotions I do.
Other	Item included: I am good at understanding other people's complex emotions.
Distinction	N
Synthesis	64. Thinking about other people's emotional experiences helps me to think about my own.
Internal cues	20. It is hard for me to talk about my complex emotions. 28. It is easy for me to notice when I am feeling different emotions at the same time.
External cues	I am open to what others say about me to help me know what I am feeling.
Distinction	N
Synthesis	N
Cognitive	I can still think rationally even if my emotions are complex.
Affective	I can still think rationally even if my emotions are complex. 33. I can easily label "basic emotions" (fear, anger, sadness joy and surprise) that I feel.
Distinction	18. I can still think rationally even if my emotions are complex.
Synthesis	1. I often think about how the emotions that I feel stem from earlier life experiences (e.g. family dynamics during childhood). 45. If I feel something, I prefer not to discuss it with others.
Reactive	I can quickly identify my emotions without having to think too much about it. 35. I am good at controlling my emotions.
Reflective	I often figure out where my emotions stem from. 1. I often think about how the emotions that I feel stem from earlier life experiences (e.g. family dynamics during childhood).
Distinction	N
Synthesis	50. I often look back at my life history to help inform my current emotional state and situation.
Uncertainty	It takes me a while to know how I am really feeling. 24. I am often confused about the emotions that I feel.

Measure	**7. Psychological Mindedness Scale**
Self	11. Usually, if I feel an emotion, I can identify it. 5. Often I don't know what I'm feeling.
Other	20. If a good friend of mine suddenly started to insult me, my first reaction might be to try to understand why he was so angry. 32. I really enjoy trying to figure other people out. 43. I think that no matter how hard you try, you'll never really understand what makes people tick. 2. I am always curious about the reasons people behave as they do.
Distinction	43. I think that no matter how hard you try, you'll never really understand what makes people tick.
Synthesis	18. I get annoyed when people give me advice about changing the way I do things.

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Internal cues	26. Understanding the reasons you have deep down for acting in certain ways is important.
External cues	N
Distinction	17. People sometimes say that I act as if I'm having a certain emotion (anger, for example) when I am unaware of it. 26. Understanding the reasons you have deep down for acting in certain ways is important.
Synthesis	N
Cognitive	20. If a good friend of mine suddenly started to insult me, my first reaction might be to try to understand why he was so angry
Affective	5.Often I don't know what I'm feeling. 11. Usually, if I feel an emotion, I can identify it. 20. If a good friend of mine suddenly started to insult me, my first reaction might be to try to understand why he was so angry. 23. Often, even though I know that I'm having an emotion, I don't know what it is. 29. I am sensitive to the changes in my own feelings. 35. I'm usually in touch with my feelings. 41. I frequently don't want to delve too deeply into what I'm feeling.
Distinction	N
Synthesis	21. I think that when a person has crazy thoughts, it is often because he is very anxious and upset
Reactive	38. If I suddenly lost my temper with someone, without knowing exactly why, my first impulse would be to forget about it. 17. People sometimes say that I act as if I'm having a certain emotion (anger, for example) when I am unaware of it
Reflective	8. I often find myself thinking about what made me act in a certain way.
Distinction	I find that once I develop a habit, it is hard to change, even if I know there is another way of doing things that might be better.
Synthesis	N
Uncertainty	5. Often I don't know what I'm feeling. 11. Usually, if I feel an emotion, I can identify it.

Measure	**8. The Metacognition Self-Assessment Scale**
Self	1. I can distinguish and differentiate my own mental abilities (e.g. remembering, imagining, having fantasies, dreaming, desiring, deciding, foreseeing and thinking). (and whole 1st section) 2 I can define, distinguish and name my own emotions.
Other	1. I'm aware that I am not necessarily at the centre of the other's thoughts, feelings and emotions and that other's behaviours arise from reasons and goals that can be independent from my own perspective and from my own involvement in the relationship. (and two middle sections) I can identify and understand the emotions of people I know.
Distinction	1. I'm aware that I am not necessarily at the centre of the other's thoughts, feelings and emotions and that other's behaviours arise from reasons and goals that can be independent from my own perspective and from my own involvement in the relationship. 2. I am aware that others may perceive facts and events in a different way from me and interpret them differently.
Synthesis	4. When problems are related to the relationship with the other people, I try to solve them on the basis of what I believe to be their mental functioning.
Internal cues	I can describe the thread that binds my thoughts and my emotions even when they differ from one moment to the next.
External cues	N
Distinction	2. I am aware that others may perceive facts and events in a different way from me and interpret them differently. 1. I'm aware that I am not necessarily at the centre of the other's thoughts, feelings and emotions and that other's behaviours arise from reasons and goals that can be independent from my own perspective and from my own involvement in the relationship.
Synthesis	N
Cognitive	1. I can distinguish and differentiate my own mental abilities (e.g. remembering, imagining, having fantasies, dreaming, desiring, deciding, foreseeing and thinking).
Affective	2. I can define, distinguish and name my own emotions.
Distinction	1. I can distinguish and differentiate my own mental abilities (e.g. remembering, imagining, having fantasies, dreaming, desiring, deciding, foreseeing and thinking). 1. I can understand and distinguish the different mental activities as when they are, for example, remembering, imagining, having fantasies, dreaming, desiring, deciding, foreseeing and thinking. 3. I am aware of what are the thoughts or emotions that lead my actions.
Synthesis	7. I can describe the thread that binds my thoughts and my emotions even when they differ from one moment to the next. 3. I can describe the thread that binds thoughts and emotions of people I know, even when they differ from one moment to the next.
Reactive	1. I can deal with the problem voluntarily imposing or inhibiting a behaviour on myself.
Reflective	3. I can deal with the problems trying to challenge or enrich my views and my beliefs on problems themselves. 2. I can deal with the problems voluntarily trying to follow my own mental order.
Distinction	N
Synthesis	N
Uncertainty	4. I am aware that what I think about myself is an idea and not necessarily true. I realize that my opinions may not be accurate and may change. 2. I am aware that others may perceive facts and events in a different way from me and interpret them differently.

Measure	**9. The Reflective Functioning Questionnaire-Revised- 7**
Self	I don't always know why I do what I do
Other	People's thoughts are a mystery to me
Distinction	1. _ People's thoughts are a mystery to me
Synthesis	N
Internal cues	Strong feelings often cloud my thinking
External cues	When I get angry I say things without really knowing why I am saying them, If I feel insecure I can behave in ways that put others' backs up
Distinction	N
Synthesis	N
Cognitive	People's thoughts are a mystery to me
Affective	Strong feelings often cloud my thinking
Distinction	Strong feelings often cloud my thinking
Synthesis	N
Reactive	Sometimes I do things without really knowing why
Reflective	When I get angry I say things that I later regret
Distinction	When I get angry I say things that I later regret

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Synthesis	N
Uncertainty	People's thoughts are a mystery to me - plus two scales 3. _ When I get angry I say things without really knowing why I am saying them

Measure	**10. The Reflective Functioning Questionnaire (8)**
Self	I don't always know why I do what I do
Other	People's thoughts are a mystery to me
Distinction	1. _ People's thoughts are a mystery to me
Synthesis	N
Internal cues	Strong feelings often cloud my thinking
External cues	When I get angry I say things without really knowing why I am saying them, If I feel insecure I can behave in ways that put others' backs up
Distinction	N
Synthesis	N
Cognitive	People's thoughts are a mystery to me
Affective	Strong feelings often cloud my thinking
Distinction	Strong feelings often cloud my thinking
Synthesis	N
Reactive	Sometimes I do things without really knowing why
Reflective	When I get angry I say things that I later regret
Distinction	When I get angry I say things that I later regret
Synthesis	N
Uncertainty	People's thoughts are a mystery to me - plus two scales

Measure	**11. Theory of Mind Inventory: Self Report-Adult**
Self	I can accurately identify and reflect on my own emotions (that is, I know what I am feeling and why I am feeling it) 4. I have rich memories for my own past experiences (when recalling something, I remember where I was, when something happened, who else was there, what I was thinking or feeling)
Other	If I heard a waitress say to a coworker "Hey, the ham sandwich left me a big tip!", I would understand that the waitress was referring to a customer who had ordered a ham sandwich 21. I can put myself in other people's shoes and understand how they feel 11. I understand that people can hold two or more contradictory beliefs, ideas or values at the same time (e.g., believing in fuel efficient cars but wanting a large fuel inefficient car)
Distinction	If I had a friend who was sad about something, I would feel sad myself I can put myself in other people's shoes and understand how they feel 34. I can predict how my words and actions will be interpreted by others 49. If I were watching a movie with a friend, I could predict what my friend was thinking and feeling in response to the movie
Synthesis	N
Internal cues	I feel sad or bad when I think about others who experience misfortune 42. I can not only remember but also re-experience or relive an earlier experience (e.g., when thinking of a birthday party, I can re-experience the sights, smells, or sounds) 29. I can accurately identify and reflect on my own emotions (that is, I know what I am feeling and why I am feeling it) 39. I can read people to tell if they like me or not
External cues	19. People do certain things when they are not interested in talking to us (e.g., they might look away for a long time or start fidgeting). I can recognize when a listener is not interested in what I am saying 35. I can read other people's body language to tell how they feel (e.g., feeling happy, sad, mad, or scared)
Distinction	I understand how people use facial expressions to hide how they really feel 37. I understand how people can move their bodies to mislead others (e.g., someone who 'walks tall' to hide a lack of confidence)
Synthesis	N
Cognitive	I understand why people often cling to mistaken beliefs despite overwhelming evidence to the contrary 46. I understand my own actions (that is, I know what I am doing and why I am doing it) 47. I understand my own desires (that is, I know what I want and why I want it)
Affective	I understand that it is possible to experience two conflicting emotions at the same time 50. I am good at explaining my feelings
Distinction	N
Synthesis	58. When I recount stories, I talk about people's thoughts and feelings to explain their actions
Reactive	I understand how people make snap judgments about others based on very little information 53. I am good at picking up on other people's moods without having to think about it
Reflective	4. I have rich memories for my own past experiences (when recalling something, I remember where I was, when something happened, who else was there, what I was thinking or feeling)
Distinction	N
Synthesis	N
Uncertainty	29. I can accurately identify and reflect on my own emotions (that is, I know what I am feeling and why I am feeling it)

Measure	**12. Interactive Mentalizing Questionnaire**
Self	I'm good at keeping my thoughts to myself. I have accurate insight into why I think the way I do
Other	I'm confident that I can tell what others are thinking. I cannot lie, because people will know my intentions. When I watch a movie, I can always guess what the character will do next 4. I believe that I am good at telling what another person is thinking

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Distinction	How confident are you that others can guess what you are thinking?
	8. Sometimes, I think people have direct insight into what I am thinking
Synthesis	Sometimes, I think people have direct insight into what I am thinking
	Under the right conditions, I'm good at lying make people feel better.
Internal cues	I have accurate insight into why I think the way I do
External cues	I have accurate insight into why I act the way I do
	17. I can tell if others are teasing me.
Distinction	I cannot lie, because people will know my intentions.
Synthesis	N
Cognitive	I have accurate insight into why I think the way I do
	5. I'm confident that I can tell what others are thinking.
Affective	Under the right conditions, I'm good at lying make people feel better.
Distinction	N
Synthesis	9. Under the right conditions, I'm good at lying make people feel better.
Reactive	7. When I watch a movie, I can always guess what the character will do next
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	I'm confident that I can tell what others are thinking.
	How confident are you that others can guess what you are thinking?
	24. I have high confidence in knowing who I am.

Measure	**13. the Mentalization Imbalances Scale**
Self	6. P. feels that his/her emotions are out of his/her control.
	P. often seems to lack words to describe his/her own feelings and emotions.
Other	4. P. can't assume other people's perspective when reflecting on behaviors.
Distinction	5. P. can easily be influenced by other people's emotions.
	P. can't consider points of view that are different from his/her own.
Synthesis	5. P. can easily be influenced by other people's emotions.
Internal cues	N - conflated with imbalance towards self?
External cues	1. P. is excessively focused on the facial expressions and/or nonverbal cues when communicating with others (including the therapist).
Distinction	N
Synthesis	N
Cognitive	7. P. understands people more on a cognitive level than on an affective one.
Affective	7. P. understands people more on a cognitive level than on an affective one.
Distinction	7. P. understands people more on a cognitive level than on an affective one.
Synthesis	3. P. often seems to lack words to describe his/her own feelings and emotions.
Reactive	22. P. seems to preverbally intuit people's feelings or thoughts.
	When solicited (e.g., with questions or confrontations),P. fails to reflect on his/her own behaviors.
	P. fails to reflect on the first impression he or she has of a person or a situation.
Reflective	19. P. fails to reflect on the first impression he or she has of a person or a situation.
	P. feels that his/her emotions are out of his/her control
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**14. Modes of Mentalization Scale.**
Self	can't consider a point of view that differs from his/her own.
	M24. P. seems to have all the answers regarding his/her own and/or other people's behavior.
Other	can't consider a point of view that differs from his/her own.
	M9. P. is excessively sure of the motivations and/or thoughts and/or emotions of others.
Distinction	can't consider a point of view that differs from his/her own.
	M16. P. can't consider point of view that differs from his/her own.
Synthesis	M1. P. seems to use his/her mental capacities to manipulate other people.
Internal cues	M13. P. tends to rely in an excessive way to his/her intuitive capacity.
External cues	M4. P. seems to recognize the interest of significant others only if it is supported by concrete actions.
Distinction	N
Synthesis	N
Cognitive	seems to focus more on what people do rather than on what they think or feel.
	M9. P. is excessively sure of the motivations and/or thoughts and/or emotions of others.
Affective	seems to focus more on what people do rather than on what they think or feel.
	M23. When solicited with specific questions, P. interprets behaviors in term of mental states.
Distinction	N
Synthesis	N
Reactive	M11. P. spontaneously interprets behaviors in term of mental states.
Reflective	M14. P. is curious about the comprehension of his/her own or other people's functioning.
Distinction	N
Synthesis	N
Uncertainty	Factor 1. Excessive certainty items, e.g. tends to express an excessive certainty about other people's thoughts or feelings.
	M9. P. is excessively sure of the motivations and/or thoughts and/or emotions of others.

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Measure	**15. Reflective Functioning Scale**
Self	Asks about thoughts & feelings of self 4.3.4 Envisioning changes of mental states between past and present, and present and future A key indicator of reflective-functioning is manifested when speakers consider developmental changes in their own mental states.
Other	Asks about thoughts & feelings of other Mental states in relation to the interviewer
Distinction	asks separately and codes for errors e.g. self-serving/distorted 1B 4.2.3 Recognition of diverse perspectives The speaker explicitly acknowledges that different people may perceive a given behaviour or situation differently. For example: "My mother had the habit of lifting her hand and slapping us, um, if we were naughty, or when she thought that we were naughty." (5) My father thought it was fine for Mr X (the teacher) to behave like that, that it would teach us self-discipline and that kind of stuff, my mother thought that it was appalling to treat children that way. I think he just didn't know what was going on. How all the children felt" (5) 4.4.1 Acknowledging the separateness of minds Speakers demonstrating this understanding make explicit their appreciation that the interviewer may not necessarily share their mental state.
Synthesis	codes statements about impact of self & others mental states on each other 4.2.4 Taking into account one's own mental state in interpreting others' behaviour Speakers who make explicit reference to how an interpretation of an event might have been distorted by what they were feeling or thinking at the time should be credited with this understanding, e.g., "I saw him as behaving in an uncaring way but actually I was very angry with him at the time because of the way he treated my mother ... So perhaps, that's why." (6) 4.4.3 Emotional attunement Others are far more attuned to the interviewer's likely emotional reaction, and indicate their awareness in their narrative. 4.3.1 Taking an intergenerational perspective, making links across generations
Internal cues	Asks about thoughts and feelings
External cues	n
Distinction	n
Synthesis	codes for integration of different cues, e.g. B2, B5, B4
Cognitive	asks about thoughts a3 4.1.3 Recognition of the limitations on insight Another sign of awareness of the nature of mental states is the explicit qualifying of insight concerning oneself or others, i.e. awareness of one's limitations in being able to understand self and others: "I had a lot of respect for my mother, although sometimes she infuriated me because she was a very, er, anxious person, and, um, she would get very uptight about things and sometimes get a little bit hysterical, I think she was very insecure in her relationship with my father, but I don't know if that was true" (6) "I had a real go at her, but I think I was just frightened that she didn't care about me." (6) When rating this item, one would mostly be talking about thoughts (cognitive)
Affective	were you ever frightened or worried as a child? through out the interview
Distinction	asks about them separately but not aspect of coding
Synthesis	Revising thoughts and feelings in light of adult understanding
	4.1.2 Mental states as susceptible to disguise Related to the issue of opaqueness is the possibility of deliberate disguising of internal states. Recognition of this possibility may be implicit or explicitly stated. A common example would be instances of awareness that the individual may experience different emotions to the ones they display, and may refer to the other or to the self (e.g. "I am so angry at her ... but I would never show that to her." (4)) "My mother always kept everything controlled and calm, but I think underneath she often felt very angry"
Reactive	B7? or no because not timed or explicit question about immediate response Mental states in relation to the interviewer
Reflective	Asked to consider and voice during interview through out the interview
Distinction	N
Synthesis	many transcripts do this but not explicitly asked for or coded except maybe B7
Uncertainty	Overcertainty 3B or undercertainty 1A coded 4.1. Awareness of the nature of mental states people may demonstrate their uncertainty about their own and others mental state through out the interview

Measure	**16. Observable Social Cognition: A Rating Scale**
Self	Only in terms of own beliefs about others - The individual will consider evidence that contradicts misinterpretations she or he has made, although they might maintain the false belief anyway.
Other	Recognising other people's emotions, particularly negative emotions (sadness, fear and anger) based on facial expression, body language and/or vocal tone and rate? Seeing things from the perspective of others (i.e., putting themselves in other people's shoes): The individual may have difficulty understanding why someone feels upset or angry in a particular situation, or they may not feel moved by a sad film or story.
Distinction	N
Synthesis	N
Internal cues	Making decisions quickly (jumping to conclusions): The individual may assume that if they try to call someone and there is no answer, the person is intentionally ignoring them out of anger. This is an example of not considering other possible reasons for why the person might not have answered the call.
External cues	1. Recognising other people's emotions, particularly negative emotions (sadness, fear and anger) based on facial expression, body language and/or vocal tone and rate? Understanding subtle social cues, hints and indirect requests (an example of an indirect request is if your son/daughter wants a toy, but rather than say so directly, comments on how pretty it is. Recognising other people's emotions: The individual may have difficulty telling if another person is upset or angry. They may seem "clueless" about how other people are feeling.
Distinction	N
Synthesis	N
Cognitive	Most of the time I don't feel like talking about my thoughts and feelings with others. Can change or correct their interpretation of social interactions when wrong?: The individual may initially believe that others are talking about them in a negative way

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
	when they are actually discussing a personal matter unrelated to the individual. The ability to reassess and change this interpretation when presented with evidence to the contrary is rated.
Affective	Recognising other people's emotions, particularly negative emotions (sadness, fear and anger) based on facial expression, body language and/or vocal tone and rate? Seeing things from the perspective of others (i.e., putting themselves in other people's shoes): The individual may have difficulty understanding why someone feels upset or angry in a particular situation, or they may not feel moved by a sad film or story.
Distinction	N
Synthesis	N
Reactive	Making decisions quickly (jumping to conclusions): The individual may assume that if they try to call someone and there is no answer, the person is intentionally ignoring them out of anger. This is an example of not considering other possible reasons for why the person might not have answered the call.
Reflective	Does not rush to decisions; is thoughtful and deliberate. The person seeks out other information and takes time to carefully weigh the pros and cons before making a decision.
	Can change or correct their interpretation of social interactions when wrong?: The individual may initially believe that others are talking about them in a negative way when they are actually discussing a personal matter unrelated to the individual. The ability to reassess and change this interpretation when presented with evidence to the contrary is rated.
Distinction	Sometimes uses only immediate information to make decisions. The person makes decisions using additional information some of the time and can weigh the pros and cons if motivated. (and other anchor points on item 3)
Synthesis	N
Uncertainty	Does not understand that more than one interpretation of an event is possible and is unable to generate any alternative guesses. (item 4)

Measure	**17. Metacognition Brief Rating Scale**
Self	A1-7 RELATIVE TO HIMSELF/HERSELF, ..
Other	B1-3 RESPECT TO OTHERS
	D4 When problems are related to relationships with other people, he/she tries to solve them on the basis of what he/she believes to be their mental functioning.
Distinction	C2 He/She is aware that others may perceive facts and events differently from him/her and interpret them differently.
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	A1. He/She is able to define and distinguish his/her own mental activities such as:... thinking.
	B1 He/She is able to understand and distinguish the different mental activities of the people he/she knows; like when, for example: they... think.
Affective	A2 He/She is able to define and distinguish his/her emotions.
	B2 He/She is able to identify and understand the emotions of the people he/she knows
Distinction	MON 1 He/She is able to define and distinguish his/her own mental activities such as: remembering, imagining, fantasizing, dreaming, desiring, deciding, foreseeing and thinking.
	(MEMORY AS COGNITIVE)
Synthesis	MON 3 He/She is aware of what thoughts or emotions are driving him/ her to perform certain actions.
Reactive	A7 He/She is able to describe the thread that binds his/her thoughts and emotions even when they change from moment to moment.
	B3 He//She is able to describe the thread that binds the thoughts and emotions of the people he/she knows even when they change from moment to moment
Reflective	INT 1 He/She is able to perceive and clearly describe his/her thoughts, emotions and the relationships he/she is involved in.
Distinction	
Synthesis	
Uncertainty	A4 He/She is aware that what he/she thinks of himself, of others and of things, are ideas and representations that are not necessarily true. He/She realizes that his views can be temporary and can change.
	D3 He/She addresses problems by trying to question or enrich his/her assessments and beliefs about the problems themselves.

Measure	**18. The brief version of the mentalization scale (ments-12)**
Self	10. I am often confused about my exact feelings 1 2 3 4 5
Other	3. I can make good predictions of other people's behavior when I know their beliefs and feelings.
	7. To understand someone's behavior, we need to know her/his thoughts, wishes, and feelings
Distinction	
Synthesis	
Internal cues	5. Sometimes I can understand someone's feelings before s/he tells me anything.
	3. I can make good predictions of other people's behavior when I know their beliefs and feelings. 1 2 3 4 5
External cues	5. Sometimes I can understand someone's feelings before s/he tells me anything.
Distinction	N
Synthesis	
Cognitive	7. To understand someone's behavior, we need to know her/his thoughts, wishes, and feelings
	5, 6, 11 - understand
	4. Often I cannot explain, even to myself, why I did something. 1 2 3 4 5
Affective	2. When I get upset I am not sure whether I am sad, afraid, or angry.
Distinction	
Synthesis	3. I can make good predictions of other people's behavior when I know their beliefs and feelings. 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
Reactive	5. Sometimes I can understand someone's feelings before s/he tells me anything.
Reflective	12. I have always been interested in why people behave in certain ways
Distinction	
Synthesis	
Uncertainty	2. When I get upset I am not sure whether I am sad, afraid, or angry.
	4. Often I cannot explain, even to myself, why I did something
	9. I find it difficult to admit to myself that I am sad, hurt, or afraid.
	10. I am often confused about my exact feelings

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale

Measure	**19. Theory of Mind Assessment Scale**
Self	Do you happen to experience emotions that make you feel good? Scale A, I–Me. It investigates the interviewee's knowledge of her own mental states. The viewpoint of the questions is centered on the interviewee (I) reflecting on her own mental states (Me), (e.g., "Do you ever experience emotions that make you feel good?"). This scale investigates first-person ToM in an egocentric perspective.
Other	Do the other persons happen to experience emotions that make them feel good? Scale C, I–Other. It investigates the interviewee's knowledge of the mental states of other persons. The viewpoint of the questions is centered on the interviewee (I) reflecting on the others' mental states (Other) (e.g., "Do you notice it when the others feel good?"). This scale is similar to scale B in that they both investigate third-person ToM; however, while the perspective there is centered on the other, here it is centered on the interviewee. In other words, here the subject is asked to take an egocentric perspective.
Distinction	Do you notice when the others feel good? Scale B, Other–Self. It investigates the knowledge that, according to the interviewee, the other persons have of their own mental states, independently of the subject's perspective. The viewpoint of the questions is centered on the other persons (Other) reflecting on their own mental states (Self), (e.g., "Do the others try to fulfill their wishes?"). This scale investigates third-person ToM in an allocentric perspective.
Synthesis	When you notice that another person feels good, does that make any difference to you? Scale D, Other–Me. It investigates the knowledge that, from the interviewee's point of view, the others have of her mental states. The viewpoint of the questions is centered on the other persons (Other) reflecting on the mental states of the interviewee (Me) (e.g., "Do the others notice it when you feel good?"). This scale can be compared with a second-order ToM task, in that the abstract form of the questions is: "What do you think that the others think that you think?".
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	When you feel bad, do you feel you understand why? Can you give an example of how you act or think, or of things that happen to you when you feel good?
Affective	Do you happen to experience emotions that make you feel good? Do you notice when the others feel good? When does that happen? Can you give an example?
Distinction	When you feel bad, do you feel you understand why?
Synthesis	When you feel bad, do you feel you understand why? [22] When you notice that another person feels good, does that make any difference to you? What differences does it make? Can you give an example, of how you act or think, or of the things that happen to you? [26] Do you ever want to influence the mood of the others?
Reactive	N
Reflective	[24] When you notice that another person feels bad, does that make any difference to you? What differences does it make?
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**20. Metacognition Assessment Scale abbreviated (MAS-A)**
Self	S3 Patients can name and distinguish between the different cognitive operations which comprise mental activity (e.g., remembering, imagining, wishing, deciding, and anticipating). Scale S-self reflectivity This scale measures self-reflectivity or one's ability to from representations of oneself which are increasingly complex and integrated. It stretches from the mere recognition of mental events within one's mind to the telling of a complex personal narrative that portrays the narrator as composed of a complex array of interacting intentions, thoughts and feelings.
Other	O3 Patients can recognize and distinguish between another person's different cognitive operations (e.g., remembering, imagining, wishing, deciding, and anticipating). O4 Patients are able to distinguish many different emotional states experienced by another person. Scale O: Understanding of the Other's Mind This scale measures one's ability to form increasing complex and integrated representations of another person. It refers to an awareness of a particular person or well-recognized group, such as a family, and not a nebulous "they" or "other people." It stretches from a mere awareness that others experience mental events to the construction of complex accounts of the lives of others.
Distinction	D2 Patients can recognize that others can perceive and/or interpret events in a validly different way than how the patient perceives and/or interprets events. Scale D: Decentration This scale measures one's ability to recognize that other people lead lives which may intersect with the participant, but that the lives of others concern much more than the interests of the participant. Thus, the participant is consequently not the center of the lives of others. As such this scale reflects the ability of participants to situate their ideas of themselves in the larger social world.
Synthesis	D3 – This score suggests participant can understand that the events that occur in regular life are often the result of complex emotional, cognitive, social and environmental factors, which vary according to the individual people involved. At this level, participants are able to form a representation of themselves and others as existing in a larger world where unique individuals have unique relationships with one another which involve no central organizing theme. Participants who can describe a conflict they had with their brother and their brother's partner in terms of each person's history and perspective, understanding that a unique relationship exists between their brother and his partner would receive a rating of D3. M8 – This level reflects a more complex form of mastery. At this level, participants are not only able to utilize the knowledge they have of themselves as in M7 but also call upon their knowledge of other people as well. An example of this level of mastery would be to know that an interpersonal conflict was both the result of {from M7:

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
	one's need to be appreciated leading one to misperceive another person as offering an insult when the other person was making an innocuous comment.} and the other person's difficulty with being receptive to feedback due to a recent death in their family.
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	S3 Patients can name and distinguish between the different cognitive operations which comprise mental activity (e.g., remembering, imagining, wishing, deciding, and anticipating). Scale S It stretches from the mere recognition of mental events within one's mind to the telling of a complex personal narrative that portrays the narrator as composed of a complex array of interacting intentions, thoughts and feelings.
Affective	S4: Patients can name and distinguish between significantly different valanced emotions. Scale S It stretches from the mere recognition of mental events within one's mind to the telling of a complex personal narrative that portrays the narrator as composed of a complex array of interacting intentions, thoughts and feelings.
Distinction	S3 Patients can name and distinguish between the different cognitive operations which comprise mental activity (e.g., remembering, imagining, wishing, deciding, and anticipating). S4: Patients can name and distinguish between significantly different valanced emotions. O6 Patients can give a complete description of another person's mental states in a specific moment, or narrative episode, distinguishing between and integrating different mental activities including thoughts, intentions and feelings.
Synthesis	S7: Patients can form representations of themselves within at least one specific situation, or narrative episode, in which they can describe how different mental activities such as thoughts and feelings influence one another. S9: Patients are able to recognize psychological patterns across their life, synthesizing multiple narrative episodes into a coherent and complex narrative which integrates different modes of cognitive and/or emotional functioning. O6 Patients can give a complete description of another person's mental states in a specific moment, or narrative episode, distinguishing between and integrating different mental activities including thoughts, intentions and feelings. O5 – This score reflects the ability to see how the thoughts and feelings of another person may influence one another, and to make reasonable guesses about the intentions of other people. At this level participants are able to make hypotheses about the cognitive and emotional functioning of other people based on verbal or nonverbal cues. An utterance suggestive of this capacity might include "I think he was so quiet because he was embarrassed by what he had done."
Reactive	N
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	S5: Patients can recognize that the ideas they have about themselves and the world are subjective, have changed, or are changeable and/or are fallible.

Measure	**21. Mentalizing Values Scale**
Self	5. She believes that she needs to be in touch with her own emotions, even negative ones, such as anger, sadness, or jealousy.
Other	1. She believes that people need to pay attention to others' emotions—being sensitive to other people's feelings is important to her. 11. She believes that people need to know others' thoughts; trying to "read the mind" of others enables a better understanding of other people
Distinction	
Synthesis	
Internal cues	4. She believes that what people feel deep inside their hearts is important, no less than what meets the eye in people's external behaviors.
External cues	4. She believes that what people feel deep inside their hearts is important, no less than what meets the eye in people's external behaviors.
Distinction	4. She believes that what people feel deep inside their hearts is important, no less than what meets the eye in people's external behaviors.
Synthesis	
Cognitive	7. She believes that people should know themselves; Therefore, they need to know their true feelings and be honest with themselves about their own thoughts. 8. It is important for her to recognize others' thoughts; she is always interested in knowing what is going through people's minds. 9. She believes people should know the thoughts that are going through their own minds; by delving into their thoughts, people can know themselves better.
Affective	1. She believes that people need to pay attention to others' emotions—being sensitive to other people's feelings is important to her.
Distinction	
Synthesis	
Reactive	9. She believes people should know the thoughts that are going through their own minds; by delving into their thoughts, people can know themselves better 7. She believes that people should know themselves; therefore, they need to know their true feelings and be honest with themselves about their own thoughts
Reflective	12. It is important for her to understand herself; she believes that self-awareness helps to maintain her mental health. 9. She believes people should know the thoughts that are going through their own minds; by delving into their thoughts, people can know themselves better.
Distinction	
Synthesis	
Uncertainty	

Measure	**22. Mentalization Questionnaire (15)**
Self	Frequently it's difficult for me to perceive my feelings at their full intensity.
Other	only in relation to perception of self - If someone yawns in my presence, that's a reliable sign that he is bored in my company. Often I feel threatened by the idea that someone could criticize or offend me. 11. I only believe that someone really likes me a lot if I have enough realistic proof for it (e.g., a date, a gift or a hug).
Distinction	N
Synthesis	N
Internal cues	Often I don't even know what is happening inside of me. I tend to ignore feelings of physical tension or of discomfort until they compel my full attention.
External cues	I only believe that someone really likes me a lot if I have enough realistic proof for it (e.g., a date, a gift or a hug).

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Distinction	N
Synthesis	N
Cognitive	Most of the time I don't feel like talking about my thoughts and feelings with others. 11. I only believe that someone really likes me a lot if I have enough realistic proof for it (e.g., a date, a gift or a hug).
Affective	Often I can't control my feelings. 6. Frequently it's difficult for me to perceive my feelings at their full intensity.
Distinction	N
Synthesis	N
Reactive	Often I can't control my feelings.
Reflective	Explanations from others are of little assistance in understanding my feelings. 5. Sometimes I only become aware of my feelings in retrospect.
Distinction	N
Synthesis	N
Uncertainty	Often I don't even know what is happening inside of me.

Measure	**23. The Certainty About Mental States Questionnaire**
Self	I know my innermost wishes and desires.
Other	I can tell whether another person is at peace with themselves.
Distinction	N
Synthesis	N
Internal cues	I know my innermost wishes and desires.
External cues	I know how a person feels when I look at their face.
Distinction	I know when other people are hiding their thoughts. I can tell when other people don't give their honest opinions.
Synthesis	N
Cognitive	I know my innermost wishes and desires. I know when other people are hiding their thoughts. (5) I know what I am trying to achieve with my behavior.
Affective	I understand my feelings. (4) I understand why certain things make me happy.
Distinction	N
Synthesis	N
Reactive	(14) When I'm in a bad mood, I know the reason why.
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	I know my innermost wishes and desires. I know how a person feels when I look at their face. All questions

Measure	**24. Mentalization Questionnaire (6)**
Self	Frequently it's difficult for me to perceive my feelings at their full intensity.
Other	only in relation to perception of self -4. I only believe that someone really likes me a lot if I have enough realistic proof for it (e.g., a date, a gift or a hug). Often I feel threatened by the idea that someone could criticize or offend me.
Distinction	N
Synthesis	N
Internal cues	Often I don't even know what is happening inside of me.
External cues	4. I only believe that someone really likes me a lot if I have enough realistic proof for it (e.g., a date, a gift or a hug).
Distinction	N
Synthesis	N
Cognitive	I only believe that someone really likes me a lot if I have enough realistic proof for it (e.g., a date, a gift or a hug).
Affective	Often I can't control my feelings.
Distinction	N
Synthesis	N
Reactive	Often I can't control my feelings.
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	Often I don't even know what is happening inside of me.

Measure	**25. Social Cognition and Object Relations Scale as applied to the thematic apperception test**
Self	May indicate own mental states when telling story about TAT Complexity of representation of people: 1 = is egocentric, or sometimes confuses thoughts, feelings, or attributes of the self and others; 3 = tends to describe people's personalities and internal states in minimally elaborated, relatively simplistic ways, or splits representations into good and bad; 5 = representations of the self and others are stereotypical or conventional, is able to integrate both good and bad characteristics of self and others, has awareness of impact on others; 7 = is psychologically minded, insight into self and others, differentiated and shows considerable complexity we recommend asking yourself "What is the extent to which understanding of (1) self and (2) others are depicted as well as the extent to which (3) understanding of self and others are integrated and differentiated." The more these three things apply, the higher the ratings.

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
	Identity and Coherence of Self (ICS) Ratings in the 6–7 range represent a person who has a sense of purpose and thinks about her needs, wants, goals, future, etc.
Other	others depicted in TAT and asked to tell a story about them. Prompt ?includes asking about mental states Complexity of representation of people: 1 = is egocentric, or sometimes confuses thoughts, feelings, or attributes of the self and others; 3 = tends to describe people's personalities and internal states in minimally elaborated, relatively simplistic ways, or splits representations into good and bad; 5 = representations of the self and others are stereotypical or conventional, is able to integrate both good and bad characteristics of self and others, has awareness of impact on others; 7 = is psychologically minded, insight into self and others, differentiated and shows considerable complexity we recommend asking yourself "What is the extent to which understanding of (1) self and (2) others are depicted as well as the extent to which (3) understanding of self and others are integrated and differentiated." The more these three things apply, the higher the ratings.
Distinction	Complexity of representations of people scale "measures the extent to which the subject clearly differentiates the perspectives of self and others" Complexity of representation of people: 1 = is egocentric, or sometimes confuses thoughts, feelings, or attributes of the self and others; 3 = tends to describe people's personalities and internal states in minimally elaborated, relatively simplistic ways, or splits representations into good and bad; 5 = representations of the self and others are stereotypical or conventional, is able to integrate both good and bad characteristics of self and others, has awareness of impact on others; 7 = is psychologically minded, insight into self and others, differentiated and shows considerable complexity we recommend asking yourself "What is the extent to which understanding of (1) self and (2) others are depicted as well as the extent to which (3) understanding of self and others are integrated and differentiated." The more these three things apply, the higher the ratings.
Synthesis	Level 5 of CRP scale: "complex representations indicating... understanding of personality as a system of processes interacting with each other and the environment" Complexity of representation of people: 1 = is egocentric, or sometimes confuses thoughts, feelings, or attributes of the self and others; 3 = tends to describe people's personalities and internal states in minimally elaborated, relatively simplistic ways, or splits representations into good and bad; 5 = representations of the self and others are stereotypical or conventional, is able to integrate both good and bad characteristics of self and others, has awareness of impact on others; 7 = is psychologically minded, insight into self and others, differentiated and shows considerable complexity we recommend asking yourself "What is the extent to which understanding of (1) self and (2) others are depicted as well as the extent to which (3) understanding of self and others are integrated and differentiated." The more these three things apply, the higher the ratings.
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	Prompt ?includes asking about thoughts Emotional Investment in Values and Morals For scores of 6 and 7, EIM becomes the central theme. There is both a genuine compassion for others as well as thought surrounding one's belief system
Affective	Affect Tone of relationship paradigms scale (specifically malevolent or benevolent tone of affect of others) Emotional Investment in Values and Morals For scores of 6 and 7, EIM becomes the central theme. There is both a genuine compassion for others as well as thought surrounding one's belief system
Distinction	N
Synthesis	Understanding of Social Causality Scale: e.g. level 3 "complex, accurate situational causality and rudimentary understanding of the role of thoughts and feelings in mediating action"
Reactive	N
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	Identity and Coherence of Self (ICS) Ratings in the 6–7 range represent a person who has a sense of purpose and thinks about her needs, wants, goals, future, etc.

Measure	**26. Interpersonal Reactivity Index—Perspective Taking subscale**
Self	13 When I see someone get hurt, I tend to remain calm.
Other	3 I sometimes find it difficult to see things from the "other guy's" point of view. 11 I sometimes try to understand my friends better by imagining how things look from their perspective.
Distinction	N
Synthesis	N
Internal cues	10 I sometimes feel helpless when I am in the middle of a very emotional situation.
External cues	N
Distinction	N
Synthesis	N
Cognitive	1 I daydream and fantasise, with some regularity, about things that might happen to me
Affective	25 When I'm upset at someone, I usually try to "put myself in his shoes" for a while. 28 Before criticizing somebody, I try to imagine how I would feel if I were in their place. 2 I often have tender, concerned feelings for people less fortunate than me.
Distinction	N
Synthesis	N
Reactive	24 I tend to lose control during emergencies.
Reflective	8 I try to look at everybody's side of a disagreement before I make a decision.
Distinction	N
Synthesis	N
Uncertainty	5. Often I don't know what I'm feeling.

Measure	**27. Mental States Measure**
Self	2-Reflective Mental State (REF) 1) The subject is involved in subjective self-perception (observation or understanding). Mental states are acknowledged as such and the subject tries to identify and give personal meaning to his own experience. He shows some capacity to perceive and refer in some verbal form to his inner, private reactions (emotions and affects, cognitions, memories, actions). The more candid or genuine, the better, but not required as such. The subject activates his internal perceptual surface, to perceive, grasp, explore and understand his own subjective inner world.

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Other	2) Empathic Reflective. The subject manifests a further Reflective capacity, in the form of basic empathy, the result of trial identification. Further exploration of either his self experience, in the present or past, or of the experience of others (in AAI protocols, mostly parents) is demonstrated. However, no elaboration is made of this material, beyond a clear basic empathy. See M020.12.
Distinction	N
Synthesis	Reflective Mental State: 3) Integrative-Elaborative Reflective : Racker (1968), and Kernberg (1987, p. 261). The subject moves beyond the previous two forms of Reflective activity, and demonstrates a further personal elaboration of the previously given material. The subject's inner situation is related to the mental situation of the significant other. This may concern present or past situations. Concordant identification involves a symmetrical identification of the subject with the mental situation of the significant other (he must have felt ashamed. Complementary identification involves an elaboration of the counterpart mental posture, while relating to the significant other (he was ashamed and I became protective). See also M020.6.
Internal cues	1-Objective-Rational Mental States High-level Objective-Rational (OBR) versus low level Objective-Rational (CONC) A.-Objective-Rational mode (high-level) (OBR) The subject activates his internal perceptual surface to take on an observing-distancing stance, treating his own endopsychic activity or content as a thing which might be observed from the outside.
External cues	N
Distinction	N
Synthesis	N
Cognitive	2-Reflective Mental State (REF) 1) The subject is involved in subjective self-perception (observation or understanding). Mental states are acknowledged as such and the subject tries to identify and give personal meaning to his own experience. He shows some capacity to perceive and refer in some verbal form to his inner, private reactions (emotions and affects, cognitions, memories, actions). The more candid or genuine, the better, but not required as such.
Affective	2-Reflective Mental State (REF) 1) The subject is involved in subjective self-perception (observation or understanding). Mental states are acknowledged as such and the subject tries to identify and give personal meaning to his own experience. He shows some capacity to perceive and refer in some verbal form to his inner, private reactions (emotions and affects, cognitions, memories, actions). The more candid or genuine, the better, but not required as such. Two main forms of Reactive mental states can be distinguished 1) Unaware immediate emotional experience. Little or no self-perception of what is being experienced and expressed is active. The subject does not demonstrate a clear awareness of the fact that through his words, he is experiencing some emotionally laden element. This is in contrast to a Reflective, self-perceived recognition. 2) The emotionally laden element (a drive derivative) can be identified by the rater. Usually by a careful examination of emotionally charged words. Anger and violence : explosion, fire, combat, fight, hostile acts (verbally expressed), persons ready to attack, threatening or potentially threatening persons, more abstract concepts of violence. Fear : victims of aggression, person or animal suffering, being threatened, terrorized, hurt, internal destruction, blood, death, dying, etc. Sexuality : oral, oral-aggressive, anal, genital sex.
Distinction	N
Synthesis	N
Reactive	2) This self-perception is at least partly or implicitly related to his own immediate presently activated situation (in a therapeutic setting), or some awareness of the inner implications of what is asked of him (in a structured interview setting). 3-Reactive Mental State (REAC) 1) Unaware immediate emotional experience. Little or no self-perception of what is being experienced and expressed is active. The subject does not demonstrate a clear awareness of the fact that through his words, he is experiencing some emotionally laden element. This is in contrast to a Reflective, self-perceived recognition. In a Reactive mental state the processes of mental perception, monitoring and elaboration are blocked, inactive or have become distorted.
Reflective	3) Integrative-Elaborative Reflective : Racker (1968), and Kernberg (1987, p. 261). The subject moves beyond the previous two forms of Reflective activity, and demonstrates a further personal elaboration of the previously given material. The subject's inner situation is related to the mental situation of the significant other. This may concern present or past situations. Concordant identification involves a symmetrical identification of the subject with the mental situation of the significant other (he must have felt ashamed. Complementary identification involves an elaboration of the counterpart mental posture, while relating to the significant other (he was ashamed and I became protective). See also M020.6. 1-Objective-Rational Mental States High-level Objective-Rational (OBR) versus low level Objective-Rational (CONC) A.-Objective-Rational mode (high-level) (OBR) The subject activates his internal perceptual surface to take on an observing-distancing stance, treating his own endopsychic activity or content as a thing which might be observed from the outside.
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**28. Multidimensional Mentalizing Questionnaire**
Self	Reflexivity Emotional dyscontrol 8. I am able to reflect on my behaviours
Other	Relational Attunement Relational Discomfort I'm able to get the deepest aspects of people around me 5. I can tune in other people's mental states
Distinction	N
Synthesis	Ego strength Distrust 9. Relationships with other people prevent me from being myself
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	N
Affective	23. It happens to me to have conflicting emotions
Distinction	N
Synthesis	All 6 factors 3. I sometimes experience mood swings I can't control
Reactive	26. When I feel an intense emotion, I can control it

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	19. For me things are either white or black

Measure	**29. Mind-Reading Belief Scale**
Self	1.Usually, I know beforehand what my conversation partner is going to say
	6.I do not think I am good at knowing human nature/ judging people
Other	All items refer to mental state or personality of other
	1. Usually, I know beforehand what my conversation partner is going to say
Distinction	N
Synthesis	N
Internal cues	N
External cues	2.I can read people's intentions in their faces
	7.It is hard to judge if somebody is lying or not by their appearance
	8.It is not possible to say what a person actually feels by their covert behavior
Distinction	N
Synthesis	N
Cognitive	2.I can read people's intentions in their faces
	5.It is hard for me to tell a person's thoughts by their looks
	3. It is possible to deduce from a person's attitude what they are going to do next
Affective	8.It is not possible to say what a person actually feels by their covert behavior
Distinction	N
Synthesis	N
Reactive	4.A stranger's character is revealed to me at first sight
	1.Usually, I know beforehand what my conversation partner is going to say
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	BUT will be confounded in score as high and low certainty indicate poor understanding of mental states
	8. It is not possible to say what a person actually feels by their covert behavior
	6. I do not think I am good at knowing human nature/ judging people

Measure	**30. Beliefs about Emotions Scale**
Self	1. It is a sign of weakness if I have miserable thoughts
	10. It is stupid to have miserable thoughts.
Other	3. If I lose control of my emotions in front of others, they will think less of me
	12. Others expect me to always be in control of my emotions.
Distinction	N
Synthesis	3. If I lose control of my emotions in front of others, they will think less of me
Internal cues	7. I should not let myself give in to negative feelings.
External cues	N
Distinction	N
Synthesis	N
Cognitive	3. If I lose control of my emotions in front of others, they will think less of me
	1. It is a sign of weakness if I have miserable thoughts
Affective	1. It is a sign of weakness if I have miserable thoughts
	3. If I lose control of my emotions in front of others, they will think less of me
Distinction	N
Synthesis	N
Reactive	3. If I lose control of my emotions in front of others, they will think less of me
Reflective	4. I should be able to control my emotions.
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**31. Narrative of Emotions Task**
Self	a time when YOU felt…
	provide a narrative account "Tell me about a time when you felt happy"; explain "Why did that make you feel happy?"
Other	sociality part of scoring
	sociality (the extent to which others were involved),
	plausibility of causal inferences (why the given event elicited the target emotion): The causal inferences scale explores one's explanation for why certain emotions were elicited during given events. Given the tendency of people with schizophrenia to personalize and externalize attributions for negative events, this could be a relevant domain in determining how rich and nuanced a person's attributions for emotional events are. Therefore, this causal inferences domain represents an attributional index. Importantly, the NET assesses how accurate and thorough participants are in providing attributions for others' behavior. According to this logic, an inability to generate accurate and thorough causal attributions for others results in either bias or confusion.
Distinction	N (unlikely for 1-1.5 minutes to give opportunity)
Synthesis	unlikely for 1-1.5 minutes to give opportunity
Internal cues	free response
External cues	free response - greater emphasis on facial expressions
Distinction	unlikely for 1-1.5 minutes to give opportunity

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Synthesis	unlikely for 1-1.5 minutes to give opportunity
Cognitive	free response - why did that make you feel x gives opportunity for cognition
	Items assess the correctness of definition provided for each emotion: cognitive mentalizing?
Affective	eight emotions asked about in questions
	plausibility of causal inferences (why the given event elicited the target emotion)
Distinction	unlikely for 1-1.5 minutes to give opportunity
Synthesis	unlikely for 1-1.5 minutes to give opportunity
Reactive	N
Reflective	free response - could raise initial and subsequent feelings/thoughts
	whether an account is given in narrative form; elaboration (the richness of narrative account, as evaluated by the amount of information provided about time, place, or other relevant details)
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**32. Mind Reading Motivation Scale**
Self	13. If the way I define something works for me, I don't need to know what other people think about it
Other	2. If someone's actions do not relate to me directly, I generally do not concern myself with why they do what they do. r (.38)
	3. When I am conversing with more than one person, I like to think about how one person is interpreting what another person says in the conversation
Distinction	2. If someone's actions do not relate to me directly, I generally do not concern myself with why they do what they do.
Synthesis	N
Internal cues	N
External cues	2. If someone's actions do not relate to me directly, I generally do not concern myself with why they do what they do
	8. When I see two strangers arguing, I often catch myself speculating on what their conflict is.
Distinction	N
Synthesis	N
Cognitive	1. When I meet new people, I like wondering how they got to where they are in life.
	3. When I am conversing with more than one person, I like to think about how one person is interpreting what another person says in the conversation
	5. I rarely find myself wondering what other people are thinking.
Affective	N
Distinction	N
Synthesis	N
Reactive	8. When I see two strangers arguing, I often catch myself speculating on what their conflict is
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	use of wondering and speculating - but I don't think it's operationalised.
	9. People who disagree with me about important issues are generally just misinformed

Measure	**33. The Porous Theory of Mind Scale**
Self	Only in relation to others ability to read thoughts and feelings
	The human mind is like a glass cube; others can read your thoughts as well.
	The human mind is like a treasure box; you have the key, but everyone who is capable to open the lock has access to your thoughts and feelings.
Other	Only in relation to others ability to read thoughts and feelings
	The human mind is like a glass cube; others can read your thoughts as well.
	The human mind is like a treasure box; you have the key, but everyone who is capable to open the lock has access to your thoughts and feelings.
Distinction	Only in relation to others ability to read thoughts and feelings
	The human mind is like a glass cube; others can read your thoughts as well.
	The human mind is like a treasure box; you have the key, but everyone who is capable to open the lock has access to your thoughts and feelings.
Synthesis	N
Internal cues	The human mind is like a treasure box; you have the key, but everyone who is capable to open the lock has access to your thoughts and feelings.
External cues	The human mind works like an antenna; signals from the outside may be received and sent.
Distinction	N
Synthesis	N
Cognitive	The human mind is like a treasure box; you have the key, but everyone who is capable to open the lock has access to your thoughts and feelings.
	The human mind is like an Internet router, receiving and sending thoughts and feelings.
Affective	The human mind is like a treasure box; you have the key, but everyone who is capable to open the lock has access to your thoughts and feelings.
	The human mind is like an Internet router, receiving and sending thoughts and feelings.
Distinction	N
Synthesis	N
Reactive	N
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**34. Grille de l'élaboration verbale de l'affect**
Self	Verbal modality
	A) Internal subjective experience, affect
	2) Determine who "owns" the affect (the participant or someone else). If it "belongs" to another, attribute a supplementary score "a".
Other	2) Determine who "owns" the affect (the participant or someone else). If it "belongs" to another, attribute a supplementary score "a".

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Distinction	N
Synthesis	N
Internal cues	Verbal modality A) Internal subjective experience, affect
External cues	Verbal modality: B) Direct expression of affect or qualification Logic tree of somatic modality
Distinction	Verbal modality A) Internal subjective experience, affect « talked about ». Explicit and controlled? ® Yes ® Ver4 No ® Ver1 B) Direct expression of affect or qualification Explicit and controlled? ® Yes ® Ver2 No ® Ver1
Synthesis	N
Cognitive	N
Affective	Identify which is the affect isolated by the AU Verbal modality A) Internal subjective experience, affect
Distinction	N
Synthesis	N
Reactive	Verbal modality B) Direct expression of affect or qualification B) A verbal mental content is used to express an affect without trying to describe or label it. Division of the logic tree: Is the AU explicit/controlled?
Reflective	Verbal modality A) Internal subjective experience, affect « talked about ». Explicit and controlled? Yes Ver4 / No Ver1 Division of the logic tree: Is the AU explicit/controlled?
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**35. Brief-Mentalized Affectivity Scale**
Self	I am good at distinguishing between different emotions that I feel. I try to understand the complexity of my emotions.
Other	N
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	N
Affective	I am good at distinguishing between different emotions that I feel. I try to understand the complexity of my emotions.
Distinction	N
Synthesis	I often look back at my life history to help inform my current emotional state and situation
Reactive	I rarely think about the reasons behind why I am feeling a certain way. It is hard for me to manage my emotions.
Reflective	I try to understand the complexity of my emotions. I often look back at my life history to help inform my current emotional state and situation
Distinction	N
Synthesis	I often look back at my life history to help inform my current emotional state and situation
Uncertainty	N

Measure	**36. Mental State Discourse analysis as applied to the Thematic Apperception Test**
Self	N
Other	presented with images of others and asked about their mental states in TAT The whole analysis is targeting the ability to mentalize with others
Distinction	a limitation of MSD as applied to the TAT is that is does not enable a distinction between mentalization of self and others
Synthesis	N
Internal cues	N
External cues	Ask about depicted characters - external appearance used as visual prompt
Distinction	N
Synthesis	N
Cognitive	Cognitive states code Cognitive states involved processing information mentally through reasoning, believing, remembering, guessing and imagining. These were coded whenever a participant referred to a cognitive event, including thoughts, beliefs, memories, dreams, and something they know, imagine, or think. e.g. Cognitive states: Think, believe, remember, dream, imagine, know, forget

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Affective	affective states code Affect desire states were coded when participants expressed a want, desire, wish, need, or missing someone. e.g. Affective desire states: want, desire, wish, need, miss Affect states were coded whenever participants referred to any kind of basic emotion, including fear, happy, sad, surprised, hate, love, and like. e.g. Affective States: Any kind of specific emotion, including, afraid, hate, love, like, good, bad, cry, worry
Distinction	N - person isn't asked about distinction although coder does distinguish
Synthesis	N
Reactive	N
Reflective	The whole analysis is targeting the ability to explicitly mentalize others
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**37. Mental-Physical Verb Norms**
Self	N
Other	N
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	N
Affective	N
Distinction	N
Synthesis	N
Reactive	N
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**38. The Moral Metacognition Scale**
Self	All items about self - refer to I 8. After engaging in the ethical decision making process I ask myself if I successfully followed an ethical guideline. 10. I know my strengths and weaknesses when it comes to making an ethical decision.
Other	N
Distinction	2. I am a better ethical decision maker when faced with an ethical dilemma that directly impacts me.
Synthesis	N
Internal cues	I know what is ethical and unethical.
External cues	During the ethical decision making process I periodically check to make sure the ethical guideline I am using is effective in making an ethical decision.
Distinction	N
Synthesis	N
Cognitive	I know what is ethical and unethical. 3. I try to apply ethical guidelines that I found helpful when faced with ethical dilemmas in the past.
Affective	N
Distinction	N
Synthesis	N
Reactive	N
Reflective	I spend time reflecting on my decision after I have made it. I find myself pausing regularly to confirm that I am considering all aspects of an ethical dilemma.
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**39. Mindful Attention Awareness Scale**
Self	1. I could be experiencing some emotion and not be conscious of it until some time later.
Other	N
Distinction	N
Synthesis	N
Internal cues	1. I could be experiencing some emotion and not be conscious of it until some time later.
External cues	5. I tend not to notice feelings of physical tension or discomfort until they really grab my attention.
Distinction	N
Synthesis	N
Cognitive	2. I break or spill things because of carelessness, not paying attention, or thinking of something else. 13. I find myself preoccupied with the future or the past
Affective	1. I could be experiencing some emotion and not be conscious of it until some time later.
Distinction	N
Synthesis	N

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Reactive	3. I find it difficult to stay focused on what's happening in the present.
	7. It seems I am "running on automatic" without much awareness of what I'm doing.
Reflective	1. I could be experiencing some emotion and not be conscious of it until some time later
Distinction	1. I could be experiencing some emotion and not be conscious of it until some time later
Synthesis	N
Uncertainty	N

Measure	**40. Basic Empathy Scale**
Self	I get caught up in other people's feelings easily.
	I don't become sad when I see other people crying.
	I often become sad when watching sad things on TV or in films.
	I get frightened when I watch characters in a good scary movie.
Other	ALL ITEMS
Distinction	My friends' emotions don't affect me much.
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	N
Affective	ALL ITEMS
Distinction	N
Synthesis	N
Reactive	I can often understand how people are feeling even before they tell me.
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	I find it hard to know when my friends are frightened.
	I have trouble figuring out when my friends are happy.

Measure	**41. Empathy Quotient**
Self	In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking.
	I get upset if I see people suffering on news programmes.
Other	I can usually appreciate the other person's viewpoint, even if I don't agree with it.
	I am quick to spot when someone in a group is feeling awkward or uncomfortable.
	34. I can tune into how someone else feels rapidly and intuitively.
	1. I can easily tell if someone else wants to enter a conversation.
Distinction	In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking.
	23. When I talk to people, I tend to talk about their experiences rather than my own.
Synthesis	N
Internal cues	N
External cues	N
Distinction	I can sense if I am intruding, even if the other person doesn't tell me.
	36. I can tell if someone is masking their true emotion.
Synthesis	N
Cognitive	In a conversation, I tend to focus on my own thoughts rather than on what my listener might be thinking.
Affective	I tend to get emotionally involved with a friend's problems.
	33. I usually stay emotionally detached when watching a film.
	34. I can tune into how someone else feels rapidly and intuitively.
Distinction	25. I am able to make decisions without being influenced by people's feelings.
Synthesis	N
Reactive	34. I can tune into how someone else feels rapidly and intuitively.
Reflective	31. Other people often say that I am insensitive, though I don't always see why.
Distinction	N
Synthesis	N
Uncertainty	8. I often find it difficult to judge if something is rude or polite.

Measure	**42. Kentucky Inventory of Mindfulness—Describe and Act With Awareness subscales**
Self	All items
Other	N
Distinction	N
Synthesis	N
Internal cues	18. I have trouble thinking of the right words to express how I feel about things.
External cues	22. When I have a sensation in my body, it's difficult for me to describe it because I can't find the right words.
Distinction	N
Synthesis	N
Cognitive	14. It's hard for me to find the words to describe what I'm thinking.
	35. When I'm working on something, part of my mind is occupied with other topics, such as what I'll be doing later, or things I'd rather be doing.
	34. My natural tendency is to put my experiences into words.

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Affective	2. I'm good at finding the words to describe my feelings. 18. I have trouble thinking of the right words to express how I feel about things.
Distinction	N
Synthesis	N
Reactive	34. My natural tendency is to put my experiences into words
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**43. Toronto Alexithymia Questionnaire**
Self	All items
Other	N
Distinction	N
Synthesis	N
Internal cues	13. I don't know what's going on inside me. 6. When I am upset. I don't know if I am sad, frightened, or angry. I can feel close to someone, even in moments of silence.
External cues	7. I am often puzzled by sensations in my body.
Distinction	I can feel close to someone, even in moments of silence.
Synthesis	N
Cognitive	N
Affective	1. I am often confused about what emotion I am feeling.
Distinction	N
Synthesis	N
Reactive	N
Reflective	8. I prefer to just let things happen rather than to understand why they turned out that way.
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**44. Difficulties in Emotion Regulation Scale—Lack of Emotional Clarity/Awareness subscales**
Self	12) When I'm upset, I acknowledge my emotions. (r) 34) When I'm upset, I take time to figure out what I'm really feeling
Other	N
Distinction	N
Synthesis	N
Internal cues	6) I am attentive to my feelings
External cues	N
Distinction	N
Synthesis	N
Cognitive	N
Affective	39) When I'm upset, I take time to figure out what I'm really feeling.
Distinction	N
Synthesis	N
Reactive	N
Reflective	39) When I'm upset, I take time to figure out what I'm really feeling.
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**45. Metacognition in Multiple Contexts Inventory.**
Self	N
Other	N
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	N
Affective	N
Distinction	N
Synthesis	N
Reactive	N
Reflective	N

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**46. Metacognitions Questionnaire 30**
Self	All items refer to self
Other	N
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	I am constantly aware of my thinking
Affective	I could make myself sick with worrying
Distinction	N
Synthesis	N
Reactive	I am constantly aware of my thinking
Reflective	I should be in control of my thoughts all of the time I think a lot about my thoughts
Distinction	N
Synthesis	N
Uncertainty	My memory can mislead me at times I should be in control of my thoughts all of the time

Measure	**47. Metacognitive Activities Inventory (Interactive MultiMedia Exercises software package - not reviewed as experimental measure paradigm)**
Self	I reflect upon things I know that are relevant to a problem. I plan how to solve a problem before I actually start solving it (even if it is a brief mental plan).
Other	N
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	I reflect upon things I know that are relevant to a problem. I plan how to solve a problem before I actually start solving it (even if it is a brief mental plan).
Affective	N
Distinction	N
Synthesis	N
Reactive	N
Reflective	I reflect upon things I know that are relevant to a problem.
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**48. Swanson's 15-item metacognition questionnaire**
Self	N
Other	Ryan is 5 years old and knows all about dinosaurs. Ryan's father does not know a lot about dinosaurs. If both Ryan and his father read a book about dinosaurs, who would remember the most? Why? Ann was lost in a forest and she came to a town in which there were two kinds of people, "truthtellers" and "liars." Truthtellers always tell the truth and liars always lie. The first person Ann talks to gives her directions to get home. The second person she talks to gives her different directions. Does Ann have a problem to solve? Why?
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	Do you think he's the smartest person in the class? Why? Who do you think was more correct in solving the mystery? Why? How do children figure out things, like how to do something?
Affective	N
Distinction	N
Synthesis	N
Reactive	N
Reflective	N

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**49. Awareness of Independent Learning Inventory**
Self	3 When I'm reading something I don't pay much attention to whether it comes alive for me. 4 I don't think it's important to feel personally involved in what you are studying. Sometimes while working together with others on an assignment I get a sudden feeling that I'm learning a great deal from them. 34 I am interested in why I sometimes get very little out of my cooperation with others.
Other	23 I can't tell whether a text to be studied will appeal to students. 5 I ignore feedback from tutors on my method of work. 26 I think that feedback on my personal learning aims is unnecessary. 28 I see no reason to talk with others about the usefulness of working together on our studies.
Distinction	26 I think that feedback on my personal learning aims is unnecessary. 28 I see no reason to talk with others about the usefulness of working together on our studies.
Synthesis	N
Internal cues	18 I sometimes get a sudden feeling that my method of work doesn't suit the assignment. 19 Sometimes while working on an assignment I get a sudden feeling that I am learning something valuable from it.
External cues	N
Distinction	N
Synthesis	N
Cognitive	34 I am interested in why I sometimes get very little out of my cooperation with others. 42 When I've studied obligatory material I ask myself whether it aroused my interest. 43 When I have to study information I try to find out what I will find interesting about it.
Affective	35 I am not interested in why I have an aversion to some of the texts I have to study. 4 I don't think it's important to feel personally involved in what you are studying.
Distinction	N
Synthesis	N
Reactive	14 Sometimes while working together with others on an assignment I get a sudden feeling that I'm learning a great deal from them.
Reflective	8 When I've finished an assignment I don't check for myself whether I've worked at it systematically enough. 41 When I've finished studying information I check for myself whether I've gone into enough depth. 2 I think it's necessary to make a conscious effort to work systematically when you are studying.
Distinction	N
Synthesis	N
Uncertainty	23 I can't tell whether a text to be studied will appeal to students.

Measure	**50. Motivated Strategies for Learning Questionnaire**
Self	N
Other	N
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	N
Affective	N
Distinction	N
Synthesis	N
Reactive	N
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**51. Assessment of Cognitive Monitoring Effectiveness**
Self	I'm going to ask you to do is pay attention for the feeling that you think an answer is wrong, and when you get it, put a big circle around that answer. Don't do anything if you think an answer is right, just if you think it might be wrong.
Other	N
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	feeling that you think an answer is wrong
Affective	N
Distinction	N
Synthesis	N
Reactive	How many of you have ever been taking a test or doing a homework problem and had the feeling that an answer you just gave is wrong ?
Reflective	N

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**52. Metacognition Scale**
Self	1. I revised my tactics for performing the task to deal with the different complexity levels of the scenarios.
	4. As I practiced the scenarios, I evaluated how well I was learning the skills of the simulation.
Other	N
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	1. I revised my tactics for performing the task to deal with the different complexity levels of the scenarios.
	4. As I practiced the scenarios, I evaluated how well I was learning the skills of the simulation.
Affective	N
Distinction	N
Synthesis	N
Reactive	N
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**53. Metacognitive Awareness Inventory**
Self	Most questions own mind focus
	I understand my intellectual strengths and weaknesses
Other	specific context - teacher
	16. I know what the teacher expects me to learn
	25. I ask others for help when I don't understand something
Distinction	N
Synthesis	N
Internal cues	44. I reevaluate my assumptions when I get confused
External cues	N
Distinction	N
Synthesis	N
Cognitive	6. I think about what I really need to learn before I begin a task
	13. I consciously focus my attention on important information
	32. I am a good judge of how well I understand something
	I understand my intellectual strengths and weaknesses
	I find myself analyzing the usefulness of strategies while I study
Affective	N
Distinction	N
Synthesis	N
Reactive	N
Reflective	specific context
	11. I ask myself if I have considered all the option s when solving a problem
	13. I consciously focus my attention on important information
	20. I have control over how well I learn
	38. I ask myself if I have considered all options after I solve a problem
Distinction	N
Synthesis	N
Uncertainty	N

Measure	**54. Metacognitive subscale of the Cultural Intelligence Scale; CQS - subscale - valid?**
Self	I am conscious of the cultural knowledge I apply to cross-cultural interactions.
	I am conscious of the cultural knowledge I use when interacting with people with different cultural backgrounds.
Other	N
Distinction	N
Synthesis	N
Internal cues	In specific context - conscious of internal states' impact on behaviour?
	I am conscious of the cultural knowledge I use when interacting with people with different cultural backgrounds.
	I adjust my cultural knowledge as I interact with people from a culture that is unfamiliar to me.
External cues	In specific context - external cues from others used to adjust own thoughts
	I adjust my cultural knowledge as I interact with people from a culture that is unfamiliar to me.
	I check the accuracy of my cultural knowledge as I interact with people from different cultures.
Distinction	N
Synthesis	N

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Cognitive	I adjust my cultural knowledge as I interact with people from a culture that is unfamiliar to me.
	I am conscious of the cultural knowledge I apply to cross-cultural interactions.
	I am conscious of the cultural knowledge I use when interacting with people with different cultural backgrounds.
Affective	N
Distinction	N
Synthesis	N
Reactive	I am conscious of the cultural knowledge I use when interacting with people with different cultural backgrounds.
Reflective	N
Distinction	N
Synthesis	N
Uncertainty	In specific context
	I check the accuracy of my cultural knowledge as I interact with people from different cultures.

Measure	**Toronto Structured Interview for Alexithymia**
Self	Are you sometimes puzzled about what emotion you are feeling?
Other	When you meet or see a stranger do you imagine what the person might be like?
	Do others believe that you are usually in touch with your feelings?
	Do you become absorbed in thinking about the characters in novels or movies?
Distinction	N
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	Is it usually easy for you to find words to describe your feelings to others?
	Do you use your imagination?
	Is it rare for you to fantasize?
	Do you become absorbed in thinking about the characters in novels or movies?
	Do you think about past emotional experiences to help you cope with more recent emotional problems?
	When you think about some past events do you relive and imagine them in your mind?
Affective	Are you sometimes puzzled about what emotion you are feeling?
	When you are upset, do you know whether you are feeling angry, sad, or anxious?
	Do you find to hard to describe your feelings about other people?
	Do you learn much about yourself on the basis of your feelings?
	Can you imagine some things so they hold or absorb your attention in the way a good movie or story does?
Distinction	N
Synthesis	N
Reactive	N
Reflective	Do you tend to just let things happen rather than trying to understand why they turn out a certain way?
	Do you think about past emotional experiences to help you cope with more recent emotional problems?
	Examples in interview scoring manual demonstrate reflective mentalizing
Distinction	N
Synthesis	N
Uncertainty	Do you experience emotional feelings that you cannot fully identify?

Measure	**The Mentalizing Emotions Questionnaire**
Self	I am interested in my emotions.
Other	I am interested in the emotions of others.
Distinction	I find it exciting to think about where others' emotions come from.
Synthesis	I think it is enriching to recognize emotions in others.
	I try to see situations through the other person's eyes.
Internal cues	I try to understand the different reasons for my feelings.
External cues	I try to understand the different reasons for my feelings.
Distinction	N
Synthesis	N
Cognitive	I think it is enriching to recognize emotions in others.
	I try to see situations through the other person's eyes.
Affective	I am interested in my emotions.
Distinction	N
Synthesis	I try to understand the different reasons for my feelings.
Reactive	I can perceive conflicting feelings in others.
Reflective	With some distance, I can understand my feelings in a new way.
Distinction	N
Synthesis	N
Uncertainty	I try to understand the different reasons for my feelings.
	I can talk to others about how my feelings change.

Measure	**The Metacognition in Self-Control Scale (MISCS)**
Self	I find myself analyzing the usefulness of strategies during self-control conflicts
Other	N
Distinction	N

(*continued on next page*)

(*continued*)

Measure	1. Reflective Functioning Rating Scale
Synthesis	N
Internal cues	N
External cues	N
Distinction	N
Synthesis	N
Cognitive	I find myself analyzing the usefulness of strategies during self-control conflicts
Affective	N
Distinction	N
Synthesis	N
Reactive	N
Reflective	I find myself analyzing the usefulness of strategies during self-control conflicts
Distinction	N
Synthesis	N
Uncertainty	N

Appendix B. Worked example of rating procedure

Table B1
The Mentalization Scale (MentS) (Dimitrijević et al., 2018) items.

1. I find it important to understand reasons for my behaviour
2. When I make conclusions about other people's personality traits I carefully observe what they say and do
3. I can recognize other people's thoughts and feelings
4. I often think about other people and their behaviour
5. usually I can recognize what makes other people feel uneasy
6. I can sympathize with other people's feelings
7. When someone annoys me I try to understand why I react in that way
8. When I get upset I am not sure whether I am sad, afraid or angry
9. I do not like to waste time trying to understand in detail other people's behaviour
10. I can make good predictions of other people's behaviour when I know their beliefs and feelings
11. Often I cannot explain, even to myself, why I did something
12. Sometimes I can understand someone's feelings before s/he tells me anything
13. I find it important to understand what happens in my relationships with people close to me
14. I do not want to find out something about myself that I will not like
15. To understand someone's behaviour, we need to know her/his thoughts, wishes and feelings
16. I often talk about emotions with people that I am close to
17. I like reading books and newspaper articles about psychological subjects
18. I find it difficult to admit to myself that I am sad, hurt or afraid
19. I do not like to think about my problems
20. I can describe significant traits of people who are close to me with precision and in detail
21. I am often confused about my exact feelings
22. It is difficult for me to find adequate words to express my feelings
23. People tell me that I understand them and give sound advice
24. I have always been interested in why people behave in certain ways
25. I can easily describe what I feel
26. While people talk about their feeling and needs my thoughts often drift away
27. Since we all depend on life circumstances, it is meaningless to think of other people's intentions or wishes
28. One of the most important things that children should learn is to express their feelings and wishes

Scoring: Likert scale 1-5 Completely incorrect/mostly incorrect/Both correct and incorrect/mostly correct/completely correct.

The relational context for this measure was recorded as self-mentalizing-self (S-S), evidenced by items 7, 8, 18, 21, 22, & 25, and self-mentalizing-hypothetical unspecified others (S-hO), evidenced by items 3, 5, 6,10, 12, 15 & 27 (see Tables B1 & B2). Shorthand for these relational contexts is S-S/hO (see Table B2).

The dimensions were rated and evidenced as shown in Table B2. Three dimensions were rated as not operationalised. These were synthesis of self and other mentalizing, distinction of internal and externally cued mentalizing, and synthesis of reflective and reactive mentalizing (see Table B2). There was no evidence found of explicit attempts to measure these dimensions, and raters could not see how the final score of the measure would alter in tandem with changes in these skills.

Three dimensions were rated as partially operationalised: cognitive, cognitive-affective distinction, and reactive-reflective distinction. Items 19 and 2 provided some evidence of an explicit attempt to operationalise appraising cognitive mental states. However, item 19 conflates cognitive and affective states, meaning the score of the item (and therefore of the overall measure) might alter in tandem with affective mentalizing despite no change in cognitive mentalizing, and item 19 does not necessarily involve mentalizing, as thinking about problems might not involve appraising mental states.

All other dimensions were rated as operationalized, with evidence for this laid out in Table B2. Exemplar descriptions of rater consideration of an aspect rating, distinction rating and synthesis rating follow. The aspect 'Reactive mentalizing' was evidenced by items 12 and 7 (see Table B2), which both explicitly attempt to operationalize reactive appraisal of mental states. Although item 7 could be argued to conflate reactive and reflective mentalizing similarly to in the partial operationalization of cognitive mentalizing above, item 12 would alter only in tandem with reactive mentalizing – in this instance, being aware of an implicit appraisal of others mental states. An explicit attempt to operationalize self other distinction was

evidenced by item 6 (see Table B2). Sympathy with others demonstrates an awareness of others feelings as separate from one's own, so this item's score would likely alter in tandem with skill in distinguishing self and other's affect. An explicit attempt to operationalize cognitive affective synthesis was evidenced by items 10 and 15. The score of these items would likely alter in tandem with integration of mentalizing cognitive (beliefs, thoughts, wishes) *and* affective (feelings) mental states in order to predict behaviour.

Table B2
A worked example: rating of The Mentalization Scale against the construct framework.

Measure:	The Mentalization Scale (MentS) (Dimitrijević et al., 2018)	
Authors' description and brief protocol:	A *self report questionnaire* assessing implicit and explicit interpretation of the meaning of own and others' actions using intentional mental states (e.g. feelings, beliefs, desires). 28 items, scale 1–5 (completely incorrect–completely correct)	
Relational context:	S-S/hO*	**Evidence** Self mentalizing self: Items 7, 8, 18, 21, 22, & 25 reference self's mental states. Self mentalizing hypothetical unspecified Other: Items 3, 5, 6,10, 12, 15 & 27 reference mental states of 'other people/someone'.
Dimension	**Rating**	**Evidence (examples)**
Self	Y	25. I can easily describe what I feel
Other	Y	3. I can recognise other people's thoughts and feelings
Distinction	Y	6. I can sympathise with other people's feelings
Synthesis	N	N: no questions ask about integrating appraised mental states of self and other in reappraising mental states of self and/or other
Internal Cues	Y	25. I can easily describe what I feel 8. When I get upset I am not sure whether I am sad, afraid, or angry.
External Cues	Y	2. When I make conclusions about other people's personality traits I carefully observe what they say and do.
Distinction	N	N: no questions ask about making use of both external or internal cues and integrating these in appraising mental states.
Synthesis	Y	10. I can make good predictions of other people' s behaviour when I know their beliefs and feelings
Cognitive	P	19. I do not like to think about my problems 2. I can recognize other people's thoughts and feelings
Affective	Y	25. I can easily describe what I feel 2. I can recognize other people's thoughts and feelings
Distinction	P	6. I can sympathise with other people's feelings
Synthesis	Y	10. I can make good predictions of other people' s behaviour when I know their beliefs and feelings 15. To understand someone's behavior, we need to know her/his thoughts, wishes, and feelings.
Reactive	Y	12. Sometimes I can understand someone's feelings before s/he tells me anything. 7. When someone annoys me I try to understand why I react in that way.
Reflective	Y	15. To understand someone's behaviour, we need to know her/his thoughts, wishes, and feelings 7. When someone annoys me I try to understand why I react in that way.
Distinction	P	7. when someone annoys me I try to understand why I react in that way
Synthesis	N	N: no questions explicitly ask about integrating appraised reactive and reflective mental states to reappraise mental states
Un/Certainty	Y	21. I am often confused about my exact feelings 8. When I get upset I am not sure whether I am sad, afraid, or angry.

Y=Measured, P=Partially measured, N=Not measured.
*Relational context shorthand: - = mentalizing, / = or, S = Self, O = other, h = hypothetical & unspecified

Appendix C. Search strategy for the scoping review

OVID syntax

MEDLINE & PSYCHINFO:

((mentalization or mentalizing or mentalized or "mental state attribution" or metacognition or "social cognition" or mind-mindedness or (reading and mind) or "reflective functioning" or "theory of mind") and (assessment or measure or questionnaire or index or test or scale or instrument)).ti. and (("mental states" or metacognition or mentalization or mind-mindedness or mentalising or mind-reading or mentalizing or mentalized or "theory of mind") and (assessment or questionnaire or measure or measurements) and (evaluation or validation or validity or construct or psychometric or reliable or validate or constructs)).ab. not (parental or parent or child or adolescent or adolescents or youth or school).ab,ti.

HAPI & PSYCHTESTS:

(mentalization or mentalizing or mentalized or "mental state attribution" or metacognition or "social cognition" or mind-mindedness or (reading and mind) or "reflective functioning" or "theory of mind").ti. not (parental or parent or child or adolescent or adolescents or youth or school).ab,ti.

Appendix D. Scoping review included records and papers

Included measures	Included papers
Metacognition Assessment Scale (MAS-R)	Semerari, A., Carcione, A., Dimaggio, G., Falcone, M., Nicolò, G., Procacci, M., & Alleva, G. (2003). How to evaluate metacognitive functioning in psychotherapy? The metacognition assessment scale and its applications. *Clinical Psychology & Psychotherapy*, *10*(4), 238–261. https://doi.org/10.1002/cpp.362
Reflective Functioning Rating Scale (RFRS)	Meehan, K. B., Levy, K. N., Reynoso, J. S., Hill, L. L., & Clarkin, J. F. (2009). Measuring Reflective Function With a Multidimensional Rating Scale: Comparison With Scoring Reflective Function On the AAI. *Journal of the American Psychoanalytic Association*, *57*(1), 208–213. https://doi.org/10.1177/00030651090570011008
Reflective Functioning Scale (RFS)	Fischer-Kern, M., Buchheim, A., Hörz, S., Schuster, P., Doering, S., Kapusta, N. D., Taubner, S., Tmej, A., Rentrop, M., Buchheim, P., & Fonagy, P. (2010). The Relationship Between Personality Organization, Reflective Functioning, and Psychiatric Classification in Borderline Personality Disorder. *Psychoanalytic Psychology*, *27*(4), 395–409. https://doi.org/10.1037/a0020862
The Mentalization Scale (MentS)	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
Metacognition Assessment Scale abbreviated (MAS-A)	Lysaker, P. H., Carcione, A., Dimaggio, G., Johannesen, J. K., Nicolò, G., Procacci, M., & Semerari, A. (2005). Metacognition amidst narratives of self and illness in schizophrenia: Associations with neurocognition, symptoms, insight and quality of life. *Acta Psychiatrica Scandinavica*, *112*(1), 64–71. https://doi.org/10.1111/j.1600-0447.2005.00514.x
Brief Reflective Functioning interview (BRFI)	Rutimann, D. D., & Meehan, K. B. (2012). Validity of a Brief Interview for Assessing Reflective Function. *Journal of the American Psychoanalytic Association*, *60*(3), 577–589. https://doi.org/10.1177/0003065112445616
Mentalized Affectivity Scale	Greenberg, D. M., Kolasi, J., Hegsted, C. P., Berkowitz, Y., & Jurist, E. L. (2017). Mentalized affectivity: A new model and assessment of emotion regulation. *PLoS ONE*, *12*(10), Article e0185264. https://doi.org/10.1371/journal.pone.0185264
Multidimensional Mentalizing Questionnaire (MMQ). (Gori et al., 2021)	Gori, A., Arcioni, A., Topino, E., Craparo, G., & Lauro Grotto, R. (2021). Development of a New Measure for Assessing Mentalizing: The Multidimensional Mentalizing Questionnaire (MMQ). *Journal of Personalized Medicine*, *11*(4), 305. https://doi.org/10.3390/jpm11040305
Psychological Mindedness Scale (PMS)	Shill, M. A., & Lumley, M. A. (2002). The Psychological Mindedness Scale: Factor structure, convergent validity and gender in a non-psychiatric sample. *Psychology and Psychotherapy: Theory, Research and Practice*, *75*(2), 131–150. https://doi.org/10.1348/147608302169607
Interactive Mentalizing Questionnaire (IMQ)	Wu, H., Fung, B. J., & Mobbs, D. (2022). Mentalizing During Social Interaction: The Development and Validation of the Interactive Mentalizing Questionnaire. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.791835
The Reflective Functioning Questionnaire-Revised (7) (Horváth et al., 2023)	Horváth, Z., Demetrovics, O., Paksi, B., Unoka, Z., & Demetrovics, Z. (2023). The Reflective Functioning Questionnaire–Revised–7 (RFQ-R-7): A new measurement model assessing hypomentalization. *PLoS ONE*, *18*(2), Article e0282000. https://doi.org/10.1371/journal.pone.0282000
The Reflective Functioning Questionnaire (8) (Fonagy et al., 2016)	Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., Howard, S., Ghinai, R., Fearon, P., & Lowyck, B. (2016). Development and Validation of a Self-Report Measure of Mentalizing: The Reflective Functioning Questionnaire. *PLoS ONE*, *11*(7), Article e0158678. https://doi.org/10.1371/journal.pone.0158678
Metacognition Assessment Interview	Semerari, A., Cucchi, M., Dimaggio, G., Cavadini, D., Carcione, A., Battelli, V., Nicolò, G., Pedone, R., Siccardi, T., D'Angerio, S., Ronchi, P., Maffei, C., & Smeraldi, E. (2012). The development of the Metacognition Assessment Interview: Instrument description, factor structure and reliability in a non-clinical sample. *Psychiatry Research*, *200*(2), 890–895.
Metacognition Self-Assessment Scale (MSAS)	Pedone, R., Semerari, A., Riccardi, I., Procacci, M., Nicolo, G., & Carcione, A. (2017). Development of a self-report measure of metacognition: The metacognition self-assessment scale (MSAS). Instrument description and factor structure. *Clinical Neuropsychiatry*, *14*(3), 185–194.
Theory of Mind Inventory: Self Report-Adult	Hutchins, T. L., Lewis, L., Prelock, P. A., & Brien, A. (2021). The Development and Preliminary Psychometric Evaluation of the Theory of Mind Inventory: Self Report-Adult (ToMI:SR-Adult). *Journal of Autism and Developmental Disorders*, *51*(6), 1839–1852. https://doi.org/10.1007/s10803-020-04654-6
Mentalization Imbalances Scale	Gagliardini, G., Gullo, S., Caverzasi, E., Boldrini, A., Blasi, S., & Colli, A. (2018). Assessing mentalization in psychotherapy: First validation of the Mentalization Imbalances Scale. *Research in Psychotherapy : Psychopathology, Process, and Outcome*, *21*(3), 339. https://doi.org/10.4081/ripppo.2018.339
Modes of Mentalization Scale.	Gagliardini, G., & Colli, A. (2019). Assessing Mentalization: Development and Preliminary Validation of the Modes of Mentalization Scale. *Psychoanalytic Psychology, 36*(3), 249–258. https://doi.org/10.1037/pap0000222
Observable Social Cognition: A Rating Scale	Healey, K. M., Combs, D. R., Gibson, C. M., Keefe, R. S. E., Roberts, D. L., & Penn, D. L. (2015). Observable Social Cognition – A Rating Scale: An interview-based assessment for schizophrenia. *Cognitive Neuropsychiatry*, *20*(3), 198–221. https://doi.org/10.1080/13546805.2014.999915
The brief version of the Mentalization Scale (Ments-12)	Stefana, A., Jolić Marjanović, Z., & Dimitrijević, A. (2024). The Brief Version of the Mentalization Scale (MentS-12): Evidence-Based Assessment of Mentalizing Capacity. *Journal of Personality Assessment*, 1–10. https://doi.org/10.1080/00223891.2024.2326884
Mentalizing Values Scale	Aival-Naveh, E., Rothschild-Yakar, L., Park, J., & Kurman, J. (2022). The Value of Thinking About Feelings Across Cultures: A Preliminary Investigation of the Mentalizing Values Scale. *Journal of Cross-Cultural Psychology*, *53*(3–4), 362–379. https://doi.org/10.1177/00220221221077355
Mentalization Questionnaire (MZQ 15)	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
The Certainty About Mental States Questionnaire	Müller, S., Wendt, L. P., & Zimmermann, Hawksdale (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
Mentalization Questionnaire (6)	Riedl, D., Kampling, H., Nolte, T., Lampe, A., Beutel, M. E., Brähler, E., & Kruse, J. (2023). Measuring Impairments of Mentalization with the 15-Item Mentalization Questionnaire (MZQ) and Introducing the MZQ-

(*continued on next page*)

(*continued*)

Included measures	Included papers
	6 Short Scale: Reliability, Validity and Norm Values Based on a Representative Sample of the German Population. *Diagnostics*, *13*(1), Article 1. https://doi.org/10.3390/diagnostics13010135
Grille de l'élaboration verbale de l'affect	Bouchard, M.-A., Target, M., Lecours, S., Fonagy, P., Tremblay, L.-M., Schachter, A., & Stein, H. (2008). Mentalization in Adult Attachment Narratives: Reflective Functioning, Mental States, and Affect Elaboration Compared. *Psychoanalytic Psychology*, *25*(1), 47–66. https://doi.org/10.1037/0736-9735.25.1.47
Metacognition Brief Rating Scale	Pedone, R., & Semerari, A. (2023). Preliminary Development and Psychometric Evaluation of the Metacognition Brief Rating Scale: An Informant form of the Metacognition Self-Assessment Scale. *Clinical Neuropsychiatry*, *20*(6), 511–522. https://doi.org/10.36131/cnfioritieditore20230606
Interpersonal Reactivity Index (IRI): Perspective Taking subscale	Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113
Mental States Measure	Lecours, S., Bouchard, M.-A., St-Amand, P., & Perry, J. C. (2000). Assessing verbal elaboration of affect in psychotherapy: A preliminary report and single case study. *Psychotherapy Research, 10*(1), 47–56. https://doi.org/10.1080/713663593
Mind-Reading Belief Scale	Realo, A., Allik, J., Nõlvak, A., Valk, R., Ruus, T., Schmidt, M., & Eilola, T. (2003). Mind-reading ability: Beliefs and performance. *Journal of Research in Personality*, *37*(5), 420–445. https://doi.org/10.1016/S0092-6566(03)00021-7
Theory of Mind Assessment Scale (Th.o.m.a.s)	Bosco, F. M., Colle, L., Fazio, S. D., Bono, A., Ruberti, S., & Tirassa, M. (2009). Th.o.m.a.s.: An exploratory assessment of Theory of Mind in schizophrenic subjects. *Consciousness and Cognition*, *18*(1), 306–319. https://doi.org/10.1016/j.concog.2008.06.006
Beliefs about Emotions Scale	Rimes, K. A., & Chalder, T. (2010). The Beliefs about Emotions Scale: Validity, reliability and sensitivity to change. *Journal of Psychosomatic Research*, *68*(3), 285–292. https://doi.org/10.1016/j.jpsychores.2009.09.014
Narrative of Emotions Task	Buck, B., Ludwig, K., Meyer, P. S., & Penn, D. L. (2014). The use of narrative sampling in the assessment of social cognition: The Narrative of Emotions Task (NET). *Psychiatry Research*, *217*(3), 233–239. https://doi.org/10.1016/j.psychres.2014.03.014
Mental State Discourse analysis as applied to the Thematic Apperception Test	Lee-Parritz, O. (2015). *Assessing Mentalization: Toward Construct Validation of Mental State Discourse Analysis as Applied to the Thematic Apperception Test* [Unpublished doctoral dissertation]. William James College. https://www.proquest.com/docview/1906687787/abstract/71A590CEA7644B5BPQ/1
Social Cognition and Object Relations Scale (SCORS-G) as applied to the Thematic Apperception Test (TAT)	Ridenour, J. M., Lewis, K. C., Siefert, C. J., & Stein, M. B. (2022). Longitudinal stability of Social Cognition and Object Relations Scale—Global Rating Method dimensional ratings, score ranges and narrative 'blandness' in a clinical sample. *Clinical Psychology & Psychotherapy*, *29*(4), 1447–1456. https://doi.org/10.1002/cpp.2729
Mind Reading Motivation Scale	Carpenter, J. M., Green, M. C., & Vacharkulksemsuk, T. (2016). Beyond perspective-taking: Mind-reading motivation. *Motivation and Emotion*, *40*(3), 358–374. https://doi.org/10.1007/s11031-016-9544-z
Mental-Physical Verb Norms	Orr, R. I., & Gilead, M. (2022). Development and validation of the Mental-Physical Verb Norms (MPVN): A text analysis measure of mental state attribution. *Behavior Research Methods, 55*, 2501–2521. https://doi.org/10.3758/s13428-022-01911-7
Brief-Mentalized Affectivity Scale	Greenberg, D. M., Rudenstine, S., Alaluf, R., & Jurist, E. L. (2021). Development and validation of the Brief-Mentalized Affectivity Scale: Evidence from cross-sectional online data and an urban community-based mental health clinic. *Journal of Clinical Psychology*, *77*(11), 2638–2652. https://doi.org/10.1002/jclp.23203
The Moral Metacognition Scale	McMahon, J. M., & Good, D. J. (2016). The Moral Metacognition Scale: Development and Validation. *Ethics & Behavior*, *26*(5), 357–394. https://doi.org/10.1080/10508422.2015.1028548
Mindful Attention Awareness Scale	Brown, K. W., & Ryan, R. M. (2003). The Benefits of Being Present: Mindfulness and Its Role in Psychological Well-Being. *Journal of Personality and Social Psychology*, *84*(4), 822–848. https://doi.org/10.1037/0022-3514.84.4.822
Basic Empathy Scale	Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, *29*(4), 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010
Empathy Quotient (EQ)	Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine*, *34*(5), 911–920. https://doi.org/10.1017/S0033291703001624
Kentucky Inventory of Mindfulness—Describe and Act With Awareness subscales	Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of Mindfulness by Self-Report: The Kentucky Inventory of Mindfulness Skills. *Assessment*, *11*(3), 191–206. https://doi.org/10.1177/1073191104268029
Toronto Alexithymia Questionnaire	Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1
Difficulties in Emotion Regulation Scale—Lack of Emotional Clarity/Awareness subscales	Gratz, K. L., & Roemer, L. (2004). Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, *26*(1), 41–54.
Metacognition in Multiple Contexts Inventory.	Allen, B. A., & Armour-Thomas, E. (1993). Construct Validation of Metacognition. *Journal of Psychology*, *127*(2), 203–211.
Metacognitions Questionnaire 30	Cook, S. A., Salmon, P., Dunn, G., & Fisher, P. (2014). Measuring Metacognition in Cancer: Validation of the Metacognitions Questionnaire 30 (MCQ-30). *PLoS ONE*, *9*(9), Article e107302. https://doi.org/10.1371/journal.pone.0107302
Metacognitive Activities Inventory	Sandi-Urena, G. S. (2009). *Design and validation of a multimethod assessment of metacognition and study of the effectiveness of Metacognitive Interventions* [Unpublished doctoral dissertation]. Clemson University. https://www.proquest.com/docview/230661741/abstract/6785A8802DDC434DPQ/1
Swanson's 15-item metacognition questionnaire	Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology*, *82*(2), 306–314. https://doi.org/10.1037/0022-0663.82.2.306
Awareness of Independent Learning Inventory	Meijer, J., Sleegers, P., Elshout-Mohr, M., van Daalen-Kapteijns, M., Meeus, W., & Tempelaar, D. (2013). The development of a questionnaire on metacognition for students in higher education. *Educational Research*, *55*(1), 31–52. https://doi.org/10.1080/00131881.2013.767024
Motivated Strategies for Learning Questionnaire	Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement*, *53*(3), 801–813. https://doi.org/10.1177/0013164493053003024
Assessment of Cognitive Monitoring Effectiveness	Osborne, J. W. (1998). *Measuring metacognition: Validation of the assessment of cognition monitoring effectiveness* [Unpublished doctoral dissertation]. State University of New York at Buffalo. https://www.proquest.com/docview/304471755/abstract/2E46A27B3DC4C1CPQ/1

(*continued on next page*)

(*continued*)

Included measures	Included papers
Metacognition Scale	Ford, J. K., Weissbein, D. A., Smith, E. M., & Gully, S. M. (1998). Relationships of Goal Orientation, Metacognitive Activity, and Practice Strategies With Learning Outcomes and Transfer. *Journal of Applied Psychology*, *83*(2), 218–233. https://doi.org/10.1037/0021-9010.83.2.218
Metacognitive Awareness Inventory	Schraw, G., & Dennison, R. S. (1994). Assessing Metacognitive Awareness. *Contemporary Educational Psychology*, *19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033
Metacognitive subscale of the Cultural Intelligence Scale	Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural Intelligence: Its Measurement and Effects on Cultural Judgment and Decision Making, Cultural Adaptation and Task Performance. *Management and Organization Review*, *3*(3), 335–371. https://doi.org/10.1111/j.1740-8784.2007.00082.x
Toronto Structured Interview for Alexithymia	Bagby, R. M., Taylor, G. J., Parker, J. D. A., & Dickens, S. E. (2006). The Development of the Toronto Structured Interview for Alexithymia: Item Selection, Factor Structure, Reliability and Concurrent Validity. *Psychotherapy and Psychosomatics*, *75*(1), 25–39. https://doi.org/10.1159/000089224
The Porous Theory of Mind Scale	van Elk, M., Maij, D., & Rutjens, B. (2020). Development and Validation of a Porous Theory of Mind Scale. *Journal of Cognition and Culture*, *20*(1–2), 41–65. https://doi.org/10.1163/15685373-12340073
The Metacognition in Self-Control Scale	Bürgler, S., Kleinke, K., & Hennecke, M. (2022). The Metacognition in Self-Control Scale (MISCS). *Personality and Individual Differences*, *199*, 111841. https://doi.org/10.1016/j.paid.2022.111841
The Mentalizing Emotions Questionnaire	Kasper, L. A., Hauschild, S., Berning, A., Holl, J., & Taubner, S. (2024). Development and validation of the Mentalizing Emotions Questionnaire: A self-report measure for mentalizing emotions of the self and other. *PLoS ONE*, *19*(5), e0300984. https://doi.org/10.1371/journal.pone.0300984

Appendix E. Measure validity and reliability reported in reviewed papers

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
N	N=49
Age	18-51 M=30.88
Gender	87.8% female
Ethnicity/Nationality	
Diagnoses	All participants had BPD diagnosis
Content Validity	
Structural Validity	Principal component factor analysis with varimax rotation - 9 factors accounted for 77.26% of the variance. Scree testing, interpretability, and internal consistency all suggested a 3 factor solution.
Internal consistency	For subscales: α = .92, .94, .95
Reliability	
Criterion Validity	RFS was found to be significantly related to Factor 1 (Defensive/Distorted), r = –.37, p < .04, and Factor 2 (Awareness of Mental States), r = .54, p < .001; but it was not related to Factor 3 (Developmental), r = .25, p < .16
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	
Other characteristics	All participants in therapy at time of study

Measure & Paper	**2. The Mentalization Scale (Dimitrijević et al., 2018) Available but not yet validated in English - this study reports validation with Serbian-speaking participants, with a Bulgarian and German validation under way.**
N	N=288 (study 1) N=124 (study 2)
Age	Study 1: 21-61 M=40.1 Study 2: 18-60
Gender	study 1: 44.4% female, study 2: 67.7% female
Ethnicity/Nationality	
Diagnoses	Clinical group in study 2 had BPD diagnosis
Content Validity	
Structural Validity	Principal component analysis with an oblimin rotation - 7 factors accounted for 55.08% of the variance. Parallel analysis suggested a 3 factors, accounting for 37.51% of the variance.
Internal consistency	MentS whole-scale internal consistency was good in the community and acceptable in the clinical sample (α = .84 and .75, respectively). Factors showed acceptable reliabilities (α = .74–.79), except for the motivation subscale in the clinical sample (α = .60).
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	MentS scores had, as hypothesised, positive associations with the Big Five, empathy, trait and ability emotional intelligence, openness, extraversion, and conscientiousness, and negative associations with attachment avoidance and anxiety, and neuroticism. The BPD group scored significantly lower than non-clinical matched pairs on MentS total and the self subscale.
Blinding/ Design	
Other characteristics	

Measure & Paper	**3. Metacognition Assessment Scale (Semerari et al., 2003)**
N	N=11
Age	two case examples discussed: 22 and 26
Gender	two case examples discussed: both female
Ethnicity/Nationality	
Diagnoses	two case examples had diagnoses of MDD and BPD, and MDD and NPD.
Content Validity	
Structural Validity	

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Internal consistency	
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	In the case studies the patient diagnosed with BPD had a low score on the mastery subscale as predicted. The patient diagnosed with NPD had a different profile to the patient with BPD, and also exhibited mentalizing deficits as predicted.
Blinding/ Design	
Other characteristics	All participants in therapy at time of study

Measure & Paper	**4. Metacognition Assessment Interview (Semerari et al., 2012)**
N	N=175
Age	women M=30.69, men M=30.68
Gender	115 women, 60 men
Ethnicity/Nationality	
Diagnoses	No psychiatric diagnoses
Content Validity	
Structural Validity	Exploratory and confirmatory factor analyses reported. 2 factor model recommended, with a range of statistical fit indices indicating good to adequate fit.
Internal consistency	$\alpha = 0.91$ (full scale) factors 0.85-0.90
Reliability	Intraclass Correlation Coefficient for every facet rated by different judges between 0.41 and 0.76, $p<0.0001$
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	Interviewers and raters were blinded to clinical characteristics and others' ratings.
Other characteristics	

Measure & Paper	**5. Brief Reflective Functioning interview (Rutimann & Meehan, 2012)**
N	N=27
Age	18-47 M=26.08
Gender	44.4% female, 53.6% male
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	
Internal consistency	$\alpha = .97$
Reliability	
Criterion Validity	A paired samples t test showed no significant difference ($t = 1.19$, ns) between the mean RFS scores on the AAI (4.00, SD = 1.84) and the BRFI (4.30, SD = 1.20)
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	
Other characteristics	

Measure & Paper	**6. Mentalized Affectivity Scale (Greenberg et al., 2017)**
N	N=2840
Age	18-65 M=31.58
Gender	55% female, 42% male, 1% transgender, 1% other
Ethnicity/Nationality	72% white caucasian, 2% African-American or black, 4% Chinese, 4% latino, 5% mixed
Diagnoses	620 participants indicated they had psychiatric diagnoses
Content Validity	
Structural Validity	Principal-components analysis with varimax rotation suggested 3 components, accounting for 43% of the variance
Internal consistency	For subscales:$\alpha = .93, .90, .88$
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	EQ total scores were correlated with the three factors, Identifying, Processing, and Expressing. The same happened with Cognitive Empathy, with the highest correlation with Processing. Emotional Reactivity was most highly correlated with Identifying, and significantly correlated with Expressing. Social Skills was most highly correlated with Processing, and significantly correlated with Expressing.
Blinding/ Design	
Other characteristics	

Measure & Paper	**7. Psychological Mindedness Scale (Shill & Lumley, 2002)**
N	N=397
Age	17-40 M=19.31
Gender	63.1% female
Ethnicity/Nationality	African-American (N = 102), Arabic/ Middle Eastern (N = 28), Asian (N = 28), Caucasian (N = 190), Hispanic (N = 5), Native American (N = 1), Multiracial (N = 20) and ' Other' (N = 13)
Diagnoses	
Content Validity	
Structural Validity	previous study reported here: principal-components analysis with varimax rotation suggested 5 components, accounting for 38% of the variance. This study: principal components analysis suggested 2 components accounting for 14.7% of the variance

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Internal consistency	
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	There was a significant negative correlation between the PMS and the Toronto Alexithymia Scale: r = ± .309, p = .01.
Blinding/ Design	
Other characteristics	

Measure & Paper	**8. The Metacognition Self-Assessment Scale** (Pedone et al., 2017)
N	N=6659
Age	18 to 75 years M = 38.61
Gender	45.8% male and 54.2% female
Ethnicity/Nationality	
Diagnoses	people with psychiatric diagnoses excluded
Content Validity	It is directly derived from two already validated instruments based on the same model, the Metacognition Assessment Scale and the Metacognition Assessment Interview.
Structural Validity	A confirmatory factor analysis suggested a four dimensional structure, with a range of statistical fit indices indicating good to adequate fit.
Internal consistency	α = 0.72-0.87 for all MSAS subscales and total score
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	
Other characteristics	

Measure & Paper	**9. The Reflective Functioning Questionnaire-Revised- 7** (Horváth et al., 2023) **RFQ-R-7, This study used the Hungarian version.**
N	Young adults N =3890, adults (N =1385)
Age	Young adults M = 27.06 years , adults M = 41.77 years
Gender	Young adults females, adults females: 53.20%
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	Confirmatory factor analysis suggested a one-factor model, which had optimal levels of model fit based on the CFI and the TLI, and insufficient model fit as per the RMSEA in both samples
Internal consistency	Omega (ω) 7 items young adults 0.924 7 items adults 0.908
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	
Other characteristics	

Measure & Paper	**10. The Reflective Functioning Questionnaire (8)** (Fonagy et al., 2016)
N	Study 1: clinical group N = 108 and non-clinical group N = 295. Study 2: clinical group N = 129 and non-clinical group N = 281 Study 3: N = 136 mother-infant dyads
Age	study 1: clinical group M = 36.08, non-clinical group M = 32.68, study 2: clinical group 18-60 M = 29, non-clinical group men M = 31.18 women M = 28.99 study 3: M = 34.24
Gender	study 1: Clinical group 81% female, non-clinical group 61% female, study 2: Clinical group 75% female, non-clinical group 50.2% female study 3: 53.7% of infants female, 100% of parents female
Ethnicity/Nationality	Study 1:White ethnic origin 92% study 2: 91.4% of men and 92.9% of women had Belgian nationality study 3: 88.1% Caucasian
Diagnoses	Study 1: clinical group had Borderline Personality Disorder (BPD) and Eating Disorder (ED) diagnoses. Study 2: clinical group had personality disorder diagnoses
Content Validity	Scale development: 101 items were rated by 14 international experts, who indicated whether they thought each statement reflected mentalizing skills or not.
Structural Validity	study 1: A multi-group CFA tested the invariance of the model across both groups. A range of statistical fit indices indicated a good fit for either a 2 factor or unconstrained model. study 2: A multi-group CFA tested the invariance of the model across both groups. A range of statistical fit indices indicated a good fit for a 2 factor model.
Internal consistency	Study 1: Estimates of internal consistency for RFQ_U and RFQ_C were α = 0.77 and 0.65 in the clinical sample, and 0.63 and 0.67 in the non-clinical sample.
Reliability	The test–retest reliability with 50 of the participants over a period of 3 weeks was excellent, with rs = 0.84 and 0.75 for RFQ_U and RFQ_C, respectively, all ps<0.001.
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	The RFQ differentiated between clinical and non-clinical participants in studies 1 & 2. In study 1: the RFQ_C was, as expected, positively related to other measures of internally based mentalizing, such as the BES and mindfulness as assessed with the MAAS, but was, against expectations, not related to perspective-taking. The RFQ_U was, against expectations, not related to empathy, but was negatively related to mindfulness and perspective-taking. The RFQ_C was slightly positively correlated with the RMET as a measure of externally based mentalizing, but this association was significant only in the non-clinical sample (r = 0.18,p<0.05). The RFQ_U was slightly negatively

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
	related to externally based mentalizing as assessed with the RMET, but only in the non-clinical sample (r = -0.17, p<0.05).Study 2: The RFQ_U was associated with increased levels of self-harm, indices of maladaptive personality functioning as assessed with the IPO, higher levels of depression, greater social impairment, and lower quality of life as assessed with the DID. The RFQ_U was also associated with difficulties with anger regulation, typical of patients with personality pathology, as expressed in higher levels of both state and trait anger, turning of anger towards the self and others, and particularly with problems with anger control, a key feature of patients with borderline levels of functioning. The RFQ_U was also associated with high levels of symptomatic distress, interpersonal problems, and low well-being. Correlations with core indices of psychopathology were typically lower than for the RFQ_U. The RFQ_C was negatively associated with indices of maladaptive personality functioning as measured with the IPO and the primitive defense mechanisms and identity diffusion subscale in particular. The RFQ_C subscale was also negatively related with trait anger and positively related with anger control. Study 3: the RFQ_C was related to infant attachment insecurity.
Blinding/ Design	
Other characteristics	Study 3: parents and infants

Measure & Paper	**11. Theory of Mind Inventory: Self Report-Adult** (Hutchins et al., 2021)
N	N=220, 109 non-ASD group, 111 ASD group
Age	ASD group males ages 18–72, M = 31.68, females ages 18–72, M = 34.84, The non-ASD group males ages 21–85, M = 46.37 females ages 18–78, M = 33.09
Gender	ASD group 56 males 55 females, non-ASD group 54 males, 55 females
Ethnicity/Nationality	ASD group indicated they were from nine countries/regions (primarily United States, United Kingdom, Europe, and Australia). The non-ASD group indicated they were from five countries/regions.
Diagnoses	clinical group had ASD or suspected ASD
Content Validity	Qualitative data were collected with the aim of understanding how participants interpreted the wording and content of individual items. The developers of the ToMI:SR-Adult who collaborated on item development and revision were experts in ASD, theory of mind, and the assessment of social cognition in ASD.
Structural Validity	
Internal consistency	α = 0.98
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	significant between-group differences between autistic and neurotypical groups for all 61 comparisons between item scores and composite scores at p < 0.001, with medium to very large effect sizes. A significant female advantage for neurotypical participants only was expected and found, which is consistent with findings of other social cognition measures.
Blinding/ Design	
Other characteristics	

Measure & Paper	**12. Interactive Mentalizing Questionnaire** (Wu et al., 2022)
N	Sample 1 N=332, Sample 2 N=417, Sample 2 N=450, Sample 4 N=299
Age	Sample 1 mean age=35.36, Sample 2 mean age=31.56, Sample 2 mean age=32.64, Sample 4 mean age=33.17, age range=18-65 for all samples
Gender	Sample 1 %female=37.95, Sample 2 %female=31.65, Sample 3 %female=37. 78, Sample 4 %female=42.81
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	study 1: The PCA analysis identified three factors that explained 51.08% of the variance of responses, studies 2&3: CFA and a range of statistical fit indices indicated satisfactory results with respect to a three factor model.
Internal consistency	The internal consistencies of the three subscales were 0.81 for IMQ_OS, 0.83 for IMQ_SS, and 0.76 for IMQ_SO.
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	The ASQ score was strongly negatively correlated with three IMQ subscales, r = –0.31, p < 0.001 for IMQ_OS, r = –0.42, p < 0.001 for IMQ_SS, r = –0.19, p < 0.01 for IMQ_SO. The IMQ_OS was negatively correlated with psychopathy scores in the LSRP, r = –0.57, p < 0.001, and IMQ_SS were negatively correlated with psychopathy scores in the LSRP, r = –0.18, p < 0.05. In contrast, we observed a positive correlation between IMQ_SO and LSRP psychopathy, r = 0.24, p < 0.001. With regard to the time perspective scale, the Past Negative Hedonism (r = –0.32, p < 0.001), Present Hedonism (r = –0.51,p < 0.001) and Present Fatalism subscales (r = –0.58, p < 0.001) were negatively correlated with IMQ_OS. The Present Hedonism (r = 0.32, p < 0.001) and Present Fatalism (r = 0.21, p < 0.001) were all positively correlated with IMQ_SO. Moreover, the Future subscale (r = 0.31, p < 0.001) and Past Positive (r = 0.24, p < 0.001) strongly positively correlated with IMQ_SS. However, the EC – as measured by the IRI – did not show significant correlation with IMQ subscales after correction. In line with our hypotheses, these relationships imply that our subscales capture aspects of meta-cognition (e.g., a positive correlation with future confidence), can reflect social competence (a negative correlation with ASQ), and yet are divergent from others measures such as EC. Study 3 - Both the proposers and responders showed similar correlations with other questionnaire measures as in Study 2. For example, we replicated the negative correlations between IMQ_OS and psychopathy (measured by LSRP) both for proposers (r = –0.53, p < 0.001) and responders (r = –0.56, p < 0.001). We also reproduced the negative correlation between IMQ components and autism traits (for proposers, IMQ_OS to ASQ: r = –0.27, p < 0.05; IMQ_SS to ASQ: r = –0.55, p < 0.001; for responders, IMQ_SS to ASQ: r = –0.47, p < 0.001; IMQ_SO to ASQ: r = –0.29, p < 0.01), with the highest correlation with IMQ_SS.
Blinding/ Design	
Other characteristics	

Measure & Paper	**13. the Mentalization Imbalances Scale** (Gagliardini et al., 2018)
N	N=190
Age	Patients aged 18-65 M= 34 years, Therapists aged 27-68 M = 39 years
Gender	65% female patients, 40% female therapists

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Ethnicity/Nationality	All patients and therapists caucasian
Diagnoses	Patients with no psychotic disorder or psychotic symptoms for at least the previous six months, whom the therapist had seen for a minimum of six sessions and a maximum of 18 months, and who had a PD diagnosis or clinically relevant personality problem. Clinicians from the two largest Italian associations of psychodynamic and cognitive-behavioral psychotherapy, and from centers specialized in the treatment of PDs, with at least three years of post-psychotherapy licensure experience.
Content Validity	
Structural Validity	Confirmatory Factor Analysis suggested a 6 factor model, and a range of statistical fit indices indicated satisfactory fit.
Internal consistency	Cronbach's alpha values ranged from .70 to .89
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	all PDs had an association with one or more mentalization imbalances as predicted.
Blinding/ Design	To minimize selection biases, we directed clinicians to consult their calendars to select the last patient each had seen during the previous week who met the study criteria
Other characteristics	

Measure & Paper	**14. Modes of Mentalization Scale.** (Gagliardini & Colli, 2019)
N	N=190
Age	Patients aged 18-65 M= 34 years, Therapists aged 27-68 M = 39 years
Gender	65% female patients, 40% female therapists
Ethnicity/Nationality	All patients and therapists caucasian
Diagnoses	Patients with no psychotic disorder or psychotic symptoms for at least the previous six months, whom the therapist had seen for a minimum of six sessions and a maximum of 18 months, and who had a PD diagnosis or clinically relevant personality problem. Clinicians from the two largest Italian associations of psychodynamic and cognitive-behavioral psychotherapy, and from centers specialized in the treatment of PDs, with at least three years of post-psychotherapy licensure experience.
Content Validity	
Structural Validity	exploratory factor analysis suggested the presence of five factors that accounted for 62% of the variance
Internal consistency	Cronbach's alpha for the 5 factors ranged from .67-.91
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Correlations show that there is a moderate positive correlation between MMS prementalizing factors and the number of PD criteria and a moderate negative correlation with the good mentalization factor of the MMS and personality pathology.
Blinding/ Design	
Other characteristics	

Measure & Paper	**15. Reflective Functioning Scale** (Fischer-Kern et al., 2010) **Version 5, 2008**
N	N=92
Age	M = 27.7 18 –51
Gender	100% female
Ethnicity/Nationality	
Diagnoses	outpatients with BPD diagnoses
Content Validity	
Structural Validity	
Internal consistency	
Reliability	The interrater reliability of RF was K = 0.79
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	The level of personality organization measured by the STIPO was correlated with the RF overall score (r = .207, p = .048. No correlation was found between RF overall scores and the number of comorbid Axis I(r .143, p .174) and Axis II (r .039, p .710) diagnoses.
Blinding/ Design	
Other characteristics	

Measure & Paper	**16. Observable Social Cognition: A Rating Scale** (Healey et al., 2015)
N	clinical group N=62, non-clinical group N=50
Age	clinical group aged 25–60, non-clinical group aged 18–65
Gender	
Ethnicity/Nationality	
Diagnoses	clinical group: schizophrenia spectrum disorders (schizoaffective = 35; schizophrenia = 27) no current substance misuse or major nervous system disorder, 2 or greater on social functioning scale non-clinical group: no history of mental illness and no first-degree relatives with a psychotic disorder, bipolar disorder or autism
Content Validity	Initial pool of items rated as relevant by a panel of 5 experts
Structural Validity	Factor Analysis suggested a 2 factor model in the clinical group, and a three factor model in the non-clinical group. A range of statistical fit indices indicated reasonable to adequate fit for the clinical model and excellent fit for the non-clinical model.
Internal consistency	Cronbach's alpha) was .80 in patients and .78 in controls.
Reliability	Test-retest reliability of the eight OSCARS items ranged from .50 to .70 (mean = .62, SD = .07). OSCARS total score test-retest reliability was .86 (n = 47; patients only).
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Individuals with schizophrenia had significantly greater deficits on the OSCARS than NPCs, F(1,108), p < .001; Table 3, after controlling for IQ and education. in the schizophrenia sample, OSCARS total and SC Ability (factor 2) scores were significantly negatively associated with AIHQ aggression index scores, indicating that higher aggressive attribution ratings (i.e. increased tendency to report acting aggressively in ambiguous situations) are correlated with less impairment in SC. In controls, OSCARS total was significantly correlated with TASIT total score, indicating greater theory of mind performance is associated with less impairment in real-world SC. OSCARS total and SC Bias (factor

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
	1) were significantly associated with GSFS scores such that less impairment in SC was associated with higher global social functioning. OSCARS total and SC Bias (factor 1) scores were significantly associated with role functioning total scores; greater functionality was associated with lower deficits in SC. WASI Full Scale IQ was not significantly correlated with the OSCARS. The OSCARS was found to be significantly associated with real-world functioning after the effect of laboratory-based SC measures was statistically removed, as indicated by change in R2. weak evidence of convergent Table 7. Prediction of indices of functioning: regression based on objective measures of SC and OSCARS total score. R2 F df P Hierarchical regression predicting GSFS Objective measures of SC .105 .667 9,51 .735 OSCARS Total Score .181 4.591 1,50 .037* Hierarchical regression predicting RFS working productivity Objective measures of SC .236 1.749 9,51 .102 OSCARS Total Score .397 13.398 1,50 .001** Hierarchical regression predicting RFS independent living Objective measures of SC .230 1.695 9,51 .114 OSCARS Total Score .276 3.128 1,50 .083ˆ Hierarchical regression predicting RFS total Objective measures of SC .142 .939 9,51 .501 OSCARS Total Score .298 11.066 1,50 .002** Note: Bolded items indicate clinical significance. *p < 0.05; **p < 0.01; ˆp < .10. Cognitive Neuropsychiatry 21 validity with measures of SC. Correlational analyses of individual OSCARS items with measures in respective SC domains provided mild evidence of convergent validity. Finally, OSCARS indices were significantly correlated with various functional outcome measures.
Blinding/ Design	
Other characteristics	

Measure & Paper	**17. Metacognition Brief Rating Scale** (Pedone & Semerari, 2023)
N	N=75 students acting as informants, N=384 'targets' - speaking about someone they knew well
Age	M = 24.45 - informants, M = 33.70 - people spoken about
Gender	89.3% female - informants, 57% female - people spoken about
Ethnicity/Nationality	All informant participants and people they spoke about White Italian
Diagnoses	Exclusion criteria: psychiatric diagnoses, psychiatric or psychological treatment, severe brain injury, and/or substance-related disorders, taken psychotropic drugs in the month before the study
Content Validity	Modifying the 18 items of the self report MSAS by writing them in the third person, consensus between measure authors through discussion
Structural Validity	The CFA analysis was conducted using the Lavaan R Packages. A four-factor solution provided the best fit for the data, with a range of statistical fit indices indicating acceptable fit, and explaining 65.66% of the total variance. Factors closely resembled those found for the MSAS.
Internal consistency	α = 0.93; ω = 0.94
Reliability	
Criterion Validity	The difference in total scores between the MBRS and MSAS was small but statistically significant producing an essentially low effect size (d z = -0.10), suggesting that total scale scores are no different across the two versions. The magnitude of the correlation between the MBRS and MSAS total scores was relatively modest (r = 0.24, p < 0.001).
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	
Other characteristics	

Measure & Paper	**18. The brief version of the mentalization scale (ments-12)** (Stefana et al., 2024)
N	N=566
Age	18-61 M=30
Gender	58% female
Ethnicity/Nationality	Serbian
Diagnoses	
Content Validity	Items from the MentS
Structural Validity	CFA showed excellent fit for a one factor solution for each subscale, and a three factor model for the full scale had a good fit, with a range of statistical fit indices indicating this.
Internal consistency	approached McDonald's ω of .71
Reliability	
Criterion Validity	correlation between the total score of the MentS-12 and the MentS was r = .91
Construct validity/Convergent/ Divergent Validity	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.
Blinding/ Design	
Other characteristics	students and working adults

Measure & Paper	**19. Theory of Mind Assessment Scale** (Bosco et al., 2009) **37 item version**
N	N=44
Age	clinical group: M=39.59, non-clinical group M=38.5
Gender	12 females and 10 males in each group
Ethnicity/Nationality	Italian speakers
Diagnoses	clinical group: diagnosed with schizophrenia. Not acutely or floridly psychotic. IQ>70
Content Validity	
Structural Validity	
Internal consistency	
Reliability	

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	clinical group performed significantly worse than non-clinical group on all subscales and dimensions.
Blinding/ Design	
Other characteristics	

Measure & Paper	**20. Metacognition Assessment Scale abbreviated (MAS-A)** (Lysaker et al., 2005)
N	N=61
Age	M=47.7
Gender	100% male
Ethnicity/Nationality	
Diagnoses	outpatients with diagnoses of schizophrenia (n = 40) or schizoaffective disorder (n = 21)
Content Validity	
Structural Validity	
Internal consistency	
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	a negative symptom was linked to all three domains of metacognition.
Blinding/ Design	Interviews transcribed with identifying information removed. Raters of the transcripts were made blind to participant identity, test performance, symptom level and ratings of quality of life. Raters were not present during the IPII interviews, nor did they transcribe the audiotapes of the interviews.
Other characteristics	no hospitalizations or changes in medication or housing in the month prior to entering the study. Excluded from the study were participants with a learning disability or active substance abuse

Measure & Paper	**21. Mentalizing Values Scale** (Aival-Naveh et al., 2022)
N	N=360
Age	UK M=19.90, Israel M=23.10, Japan M=19.80
Gender	UK 51% female, Israel 63% female, Japan 57% female
Ethnicity/Nationality	3 groups of 120 participants born in the UK, Israel and Japan
Diagnoses	
Content Validity	
Structural Validity	EFA: principal axis factoring with varimax rotation extracted one factor with an eigenvalue of 5.91 that accounted for 49.25% of the variance. In CFA, a single factor model was a good fit for all three groups, with a range of statistical fit indices indicating this.
Internal consistency	Cronbach's α = .92, .91, and .87 in the Hebrew, English, and Japanese versions, respectively
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Consistent with our hypothesis, a cultural difference existed in MVS scores, with medium effect size.
Blinding/ Design	
Other characteristics	Social science students

Measure & Paper	**22. Mentalization Questionnaire (15)** (Hausberg et al., 2012)
N	total: N=434, group 1: N=97, group 2: N=337
Age	group 1: 19-74 M=40, group 2: 17-79 M=40
Gender	group 1: 62% female, group 2: 69% female
Ethnicity/Nationality	
Diagnoses	group 1: 45 inpatients and 52 day clinic patients in a psychiatric hospital
Content Validity	
Structural Validity	principal components analysis with varimax rotation suggested a four-factor solution accounting for 59% of the variance. CFA indicated a satisfactory model fit, with a range of statistical fit indices indicating this.
Internal consistency	Internal consistency of the Mentalization Questionnaire (MZQ) was .81, and of the sub-scales, ranged from .54 to .72.
Reliability	Test-retest coefficients were calculated on the basis of a subsample drawn from sample 2 (n min=109) and can be considered as sufficient for the total scale (r=.76)
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Significant group differences in suicide attempts (t -5.21; pB.001, df -385, d -.6) and self-harming behavior (t -7.66; p B.001, df -386, d-1.0) were found in both samples. patients with Borderline Personality Disorder showed significantly lower total scores on the MZQ compared to all other patients in the sample (t -2.14; pB.05, df -87, d -.6)patients classified as secure according to the BPEQ showed significantly higher MZQ scores compared to all other patients (t-3.71; p B.001, df -87, d-1.0). Furthermore, patients with multiple disorders scored significantly lower on the MZQ compared to those with a single diagnosis (t -2.43; p B.05, df -87, d-.5). All observed group differences show medium to large effect sizes.
Blinding/ Design	
Other characteristics	

Measure & Paper	**23. The Certainty About Mental States Questionnaire** (Müller et al., 2023)
N	study 1: German N=505, U.S. N=519 study 2: German N=401, U.S. N=403
Age	study 1: U.S. 18-74 M=37.9, German 18-81 M=46.4, study 2: U.S. 18-89 M=45.2, German 18-86 M=45.9
Gender	study 1: U.S. 56% female, German 51% female, study 2: U.S. 52% female, German 49% female

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	Ant Colony Optimization algorithm as implemented in the stuart package was used to derive a psychometrically optimized item selection from a pool of 40 items. A two factor solution approached the targeted criteria in the multigroup CFA in the first study. CFA indicated a satisfactory model fit in both groups in the second study, with a range of statistical fit indices indicating this.
Internal consistency	Internal consistency was high for both Self-Certainty, ω = .90/.88 (US/GER), and Other-Certainty, ω = .91/.89.
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Self-Certainty correlated most strongly with mentalizing measures that primarily pertain to the self (i.e., MentS Self, RFQ-6, MZQ), whereas Other-Certainty converged with MentS Other, which pertains to inferring others' mental states. Low Self-Certainty was indicative of maladaptive personality traits (i.e., PID5BF+M, LPFS-BF, D16 dark personality, JSI victim sensitivity) and symptom distress (i.e., SCL-K9). By contrast, Other-Certainty exhibited overall weak and divergent bivariate associations with psychopathology measures (e.g., small positive correlations with antagonism and psychoticism, small negative correlations with detachment and interpersonal dysfunction)
Blinding/ Design	
Other characteristics	

Measure & Paper	**24. Mentalization Questionnaire (6)** (Riedl et al., 2023)
N	N = 2487
Age	age ≥14, M = 46.0
Gender	n = 1321 were women and n = 1 diverse
Ethnicity/Nationality	
Diagnoses	
Content Validity	Overall acceptance of the MZQ items was satisfactory: only n = 90 (3.4%) of the participants did not answer one or more items, and only n = 12 (0.5%) participants left out more than one answer.
Structural Validity	Model fit was borderline acceptable, with a range of statistical fit indices indicating this.
Internal consistency	internal consistency was good (ω = 0.80)
Reliability	
Criterion Validity	The MZQ-15 and MZQ-6 total score had a correlation of r = 0.94 (p < 0.001).
Construct validity/Convergent/ Divergent Validity	Both the MZQ 15-item and MZQ 6-item version were correlated with the OPD-SQS total score and subscales to evaluate convergent validity
Blinding/ Design	5418 households within 258 predefined regions were selected by a random route procedure. In households with multiple persons, one person was randomly selected using the Kish-Selection-Grid.
Other characteristics	

Measure & Paper	**25. Social Cognition and Object Relations Scale (SCORS-G) as applied to the Thematic Apperception Test (TAT)** (Ridenour et al., 2022)
N	N=61
Age	Female M=29.73, male M=31.65, transgender or gender non-conforming M=21.75
Gender	37 female, 20 male, 4 transgender or gender non-conforming participants
Ethnicity/Nationality	96.7% European American
Diagnoses	adult psychiatric patients enrolled in long-term residential treatment. 85% of participants having at least one personality disorder diagnosis.
Content Validity	
Structural Validity	
Internal consistency	
Reliability	The SCORS-G dimensional ratings remained positively associated across time points as hypothesised, but average test–retest correlation (18 months between time points) across all eight SCORS-G dimensional means (r = .43) fell below acceptable range
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	Raters were blind to all identifying information as well as the administration time point for each protocol during the coding process.
Other characteristics	

Measure & Paper	**26. Interpersonal Reactivity Index (IRI): Perspective Taking subscale** (Davis, 1983)
N	sample 1 N=770, sample 2 N=460
Age	
Gender	sample 1: 392 males, 378 females, sample 2: 225 males, 235 females
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	
Internal consistency	All four scales have satisfactory internal and test-retest reliabilities (internal reliabilities range from .71 to .77; test-retest reliabilities range from .62 to .71)
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	It was anticipated that the four IRI subscales, representing different facets of the multidimensional construct empathy, would be differentially related to existing unidimensional measures of empathy. This was found with the exception od the personal distress scale.
Blinding/ Design	
Other characteristics	psychology undergraduates

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)

Measure & Paper	**27. Mental States Measure** (Bouchard et al., 2008)
N	N=73
Age	M=35.6
Gender	59% women
Ethnicity/Nationality	4 participants of 1st generation Asian origin, 3 refugees from African states, 3 dual heritage white and African-American, 2 UK born with African Caribbean heritage, 2 migrants from European countries outside the UK, 1 African American male, and the remainder reported as White.
Diagnoses	52% had a psychiatric diagnosis
Content Validity	
Structural Validity	a principal component factor analysis among the mental states variables produced a two-factor solution explaining 74% of the variance.
Internal consistency	
Reliability	The mean percentage of agreement for the identification of the thematic units was of 84%. Pearson product–moment correlations for each mental state between the two judges were very satisfactory, except for the infrequent reactive-impulsive mental state, which will be deleted from further analyses: concrete thinking (Co, r ,870), lower defensive (Lo-Def, r .792), reactive-impulsive (Rimp, r .592), higher defensive (Hi-Def, r .779), objective-rational (Obr, r .788), and reflective (Ref, r .807).
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	The extremes of the mental states continuum show significant relations with the RFS score (r - .49, p .01 with the concrete, r - .30, p .05 with the lower defensive, and r .42, p .01 with the reflective mental state
Blinding/ Design	
Other characteristics	

Measure & Paper	**28. Multidimensional Mentalizing Questionnaire** (Gori et al., 2021)
N	Study 1: N=349, Study 2: N=96 (clinical group N=46, non-clinical group N=50)
Age	Study 1: 16-20 M=38.6, Study 2: Clinical: 18 to 62 M = 33.33, Non-clinical: M = 20 to 76 38.86
Gender	Study 1: 81% female, Study 2: Clinical: 48% female Non-clinical: 58% female
Ethnicity/Nationality	Study 1: from central Italy 62.7%
Diagnoses	Study 2: clinical group recruited in private clinical settings
Content Validity	
Structural Validity	An exploratory factor analysis (EFA) with a principal axis factoring extraction method and confirmatory factor analysis indicated a factor structure with six principal dimensions, which explained 56.9% of the total variance. Goodness-of-fit indices indicated a satisfactory fit.
Internal consistency	total scale ($\alpha = 0.75$) and the subscales (factor 1, $\alpha = 0.89$; factor 2, $\alpha = 0.81$; factor 3, $\alpha = 0.82$; factor 4, $\alpha = 0.76$; factor 5, $\alpha = 0.74$; factor 6, $\alpha = 0.72$).
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	The MMQ subscales showed significant correlations with most of the measures used to assess construct validity. The independent-samples t-test showed significant differences in the MMQ total score and its subscales between the community and clinical samples, except for the factors of relational attunement and emotional dyscontrol
Blinding/ Design	Participants were randomly recruited through a snowball-like spreading strategy of an anonymous on-line link.
Other characteristics	

Measure & Paper	**29. Mind-Reading Belief Scale** (Realo et al., 2003)
N	Sample 1: N=280, 2: N=279, 3: N=802, 4: N=101 (2 judges per participant), 5: N=138
Age	Sample 1: 14-67 M=29.1, 2: 17-64 M=30.8, 3: 14-81 M=27.2, 4: participants 17-41 M=21.9 judges M=26.0, 5: 12-71 M=28.0
Gender	Sample 1: 212 women, 2: 193 women, 3: 556 women, 4: participants 81 women judges 153 women, 5: 82 women, 2 did not indicate
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	principal component analysis demonstrated the existence of only one factor, accounting for 33.5% of the total variance.
Internal consistency	Cronbach a across samples 1-5 .70 .81 .81 .82 .64
Reliability	61 participants in Sample 3 (53 females and 8 males) were re-tested at a time interval of approximately 3 and a half years. The long-term stability coefficient of the MBS was .61 (p = .000).
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	
Other characteristics	

Measure & Paper	**30. Beliefs about Emotions Scale** (Rimes & Chalder, 2010)
N	N=194 (clinical group N=121, non-clinical group N=73)
Age	clinical group: M=38.8, non-clinical group M=37.6
Gender	clinical group: 76.0% female, non-clinical group 67.1% female
Ethnicity/Nationality	clinical group: 76.9% caucasian, 14.9% other, 8.3% not given, non-clinical group: 76.7% cauasian, 19.2% other, 4.1% not given
Diagnoses	clinical group individuals attended a specialist Chronic Fatigue Syndrome unit
Content Validity	Items selected to represent the types of beliefs about the unacceptability of experiencing and expressing emotions that have been specified in clinical reports and cognitive models.
Structural Validity	Principal components analysis extracted only one factor, which accounted for 50.1% of the variance.
Internal consistency	The Cronbach's alpha for the BES, 0.91, indicated very good reliability.
Reliability	
Criterion Validity	

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Construct validity/Convergent/ Divergent Validity	there was a significant group difference, with the CFS group reporting more negative beliefs about emotions than the non-clinical group
Blinding/ Design	
Other characteristics	

Measure & Paper	**31. Narrative of Emotions Task** (Buck et al., 2014)
N	n=45 clinical and n=50 non-clinical
Age	age 20-65
Gender	32 f 63 m
Ethnicity/Nationality	63 white, 32 black
Diagnoses	Clinical group of participants meeting DSM-IV criteria for either schizophrenia or schizoaffective disorder.
Content Validity	
Structural Validity	
Internal consistency	α=0.80
Reliability	The intraclass coefficients between trained raters were acceptable to very good for the NET total score (ICC=0.84,0.91).
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Controls and participants with schizophrenia significantly differed (using an ANCOVA with Verbal IQ as a covariate)
Blinding/ Design	
Other characteristics	

Measure & Paper	**32. Mind Reading Motivation Scale** (Carpenter et al., 2016)
N	N = 194
Age	M=18.7
Gender	41 men, 149 women, 4 declined to specify
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	maximum likelihood exploratory factor analysis on MRM responses in a larger sample (N = 462) - 1 factor supported
Internal consistency	α = .81.
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Pilot testing also revealed sensible patterns of correlations with related variables. Those high in motivation for mind-reading were also more likely to have higher need for cognition (which has a shared emphasis on cognitive effort), higher general trait empathy and attributional complexity (shared emphasis on perspectives), and lower need for closure (an opposing desire to come to a quick conclusion rather than exploring alternatives. See Table 1. For the Big Five, MRM correlated positively with openness to experience and agreeableness, but not extraversion, as measured by the Big Five Inventory.
Blinding/ Design	
Other characteristics	

Measure & Paper	**33. The Porous Theory of Mind Scale** (van Elk et al., 2020)
N	study 1: N=71, Study 2: N=150, study 3: N=404, study 4: N=772
Age	study 1: 19-68 M=44.67 years, Study 2: 14–81 M=44.8 years, study 3: 18–70 M=54.4, study 4: 18–71 M=34.9 years
Gender	study 1: 14 males, Study 2: 40 males, study 3: 188 men, study 4: 361 females
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	
Internal consistency	α = .70-.82
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	correlated strongly with traditional religious beliefs (as measured with the SBS) and with paranormal beliefs (as measured with the RPBS).
Blinding/ Design	
Other characteristics	

Measure & Paper	**34. Grille de l'élaboration verbale de l'affect** (Bouchard et al., 2008)
N	N=73
Age	M=35.6
Gender	59% women
Ethnicity/Nationality	4 participants of 1st generation Asian origin, 3 refugees from African states, 3 dual heritage white and African-American, 2 UK born with African Caribbean heritage, 2 migrants from European countries outside the UK, 1 African American male, and the remainder reported as White.
Diagnoses	52% had a psychiatric diagnosis
Content Validity	
Structural Validity	the two subscales were not orthogonal, as a significant correlation was found between the two (r = .61, p < .05).
Internal consistency	

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Reliability	the GEVA categories yielded mean Kappas of .83 (range .67 to .94) for the modalities and of .71 (range .68 to .73) for the levels of tolerance/ abstraction. Evaluation of valence produced a mean kappa of .89 (.84 to .92). This indicates very good to excellent agreement beyond chance
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	an interesting convergence is seen between WGSNO (the degree of mental elaboration in the subject's verbal expression of affects attributed to others) and RF (r .44, p .01)
Blinding/ Design	
Other characteristics	

Measure & Paper	**35. Brief-Mentalized Affectivity Scale** (Greenberg et al., 2021)
N	Study 1: N=978, Study 2: N=230
Age	Study 1: 18-65 M=36.20, study 2: 18-65 M=28.36
Gender	Study 1: 64% female, study 2: 62% female, 2% other
Ethnicity/Nationality	Study 1: 73% White Caucasian, 9% African American or Black, 6% Latino, 2% Chinese. Study 2: 46% White, 17% Black, 16% ethnically and not racially, 23% Latino/Hispanic.
Diagnoses	Study 2: adults seeking psychological services at a community-based mental health clinic
Content Validity	
Structural Validity	PCA with varimax rotation showed a clear three-component structure in both studies
Internal consistency	Cronbach's alpha 0.76-0.78
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	study 1: Convergent correlations with the ERQ, DERS, FREE, RFQ, and TAS showed strong construct validity for the BMAS. For example, the B-MAS was highly correlated with both the DERS and ERQ study 2: MAS components were not associated with childhood experiences measured by the ACE. Hierarchical linear regressions showed that mentalized affectivity significantly contributed to predicting mental health, adding 28% more variance over demographics.
Blinding/ Design	
Other characteristics	

Measure & Paper	**36. Mental State Discourse analysis as applied to the Thematic Apperception Test** (Lee-Parritz, 2015)
N	N=63
Age	17-24 median age 19
Gender	31 female
Ethnicity/Nationality	81% Caucasian, 6.3% Asian, 3.2% African American, and 1.6% Hispanic
Diagnoses	
Content Validity	
Structural Validity	
Internal consistency	
Reliability	Reliability for both raters was strong, with in average intra-class correlation coefficient (ICC) of (.93). inter-rater reliability for the individual components of TMS was also strong, as evidenced by the following ICC's: Cognitive (.98) affective desire (.95) affective (.96), total mental state (non-ratio) (.97), and behavioral (.98).
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	All significant correlations are in the opposite direction of what was originally hypothesized. Recall that positive correlations were predicted between the tendency to engage in mental state discourse and scores on the GWB, the EMI, COGREP, and EMOREP, whereas a negative correlation was predicted between TMS scores and RULS.
Blinding/ Design	Raters will be kept unaware of demographic and other data associated with participants' TAT narratives.
Other characteristics	

Measure & Paper	**37. Mental-Physical Verb Norms** (Orr & Gilead, 2022)
N	Study 1: N=103, study 2: N=385
Age	Study 1: M=39.01, study 2: M=38.23
Gender	Study 1: female = 55, Gender-fluid/Non-identifying = 2, study 2: Female = 222, Gender-fluid/Non-identifying = 3, Other = 2
Ethnicity/Nationality	U.S. residents
Diagnoses	
Content Validity	
Structural Validity	
Internal consistency	
Reliability	ICC 0.979-0.982
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Statistical dependence in expected directions between MPVN and other dictionaries (MPD, LIWC, Brysbaert's concreteness norms). MPVN scores were significantly higher on the r/MentalHealth subreddit with a moderate effect size.
Blinding/ Design	
Other characteristics	

Measure & Paper	**38. The Moral Metacognition Scale** (McMahon & Good, 2016)
N	Study1: N=211, study 2: N=160, study 3: N=253
Age	Study 1: 19-50 M=21, study 2: 17-24 M=19, study 3: 19-88 M=33
Gender	Study 1: 52% female, study 2: 47% female, study 3: 54%
Ethnicity/Nationality	
Diagnoses	

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Content Validity	one faculty member from the psychology department and three faculty members from the philosophy department at a small mid-Atlantic public university took an online survey in which they indicated whether they felt each of the original 110 items was highly representative, somewhat representative, or not at all representative of metacognition in relation to ethical decision making.
Structural Validity	study 1: We elected to run two analyses, one using the Principal Component Analysis extraction method with the Varimax method of orthogonal rotation, the other using the Maximum Likelihood extraction method with the Promax method of oblique rotation. four metacognition factors demonstrated. study 2: Confirmatory factor analysis was conducted using SAS software, and the PROC CALIS program, on the covariance matrix. Goodness-of-fit indices indicated an acceptable fit of a four factor model.
Internal consistency	study 2: Cronbach's alpha for subscales ranged from .818 (Factor 2) to .825 (Factor 4).
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	study 3: The MMS mean was significantly correlated with Need for Cognition and not significantly correlated with the idealism factor of the EPQ or with the relativism factor as hypothesised. Contrary to hypotheses, it was not significantly correlated with the MAAS. study 4: As hypothesised, the MMS mean was significantly correlated to the mean of the MIS, MAS and knowledge of cognition factor
Blinding/ Design	
Other characteristics	

Measure & Paper	**39. Mindful Attention Awareness Scale** (Brown & Ryan, 2003)
N	Total N=1,253. 7 samples, Ns = 313, 327, 207, 187, 145, 74, 239
Age	17-77 Ms=19.0-43.3
Gender	55-66% female
Ethnicity/Nationality	63-93% Caucasian, 1-15% Asian, 1-9% African American, 0-4% hispanic, 3-6% other
Diagnoses	
Content Validity	nine experienced practitioners of mindfulness (i.e., "experts") who were provided with our exclusion criteria and a brief working definition of our construct. Items that were rated highly and consistently across raters ($p > .05$) were retained. eight faculty and graduate students in psychology rated these items, and the V statistic was again used to retain items. The resulting items were rated by yet another team of six faculty and graduate students, and their feedback was used to eliminate ambiguous items and revise others. 55 items that remained from this final round of ratings were included in several pilot studies with undergraduates. Items with non-normal (skewed or kurtotic) distributions were eliminated as were those showing less than a full range of response on a 6-point Likert scale.
Structural Validity	Study 1: An EFA suggested a single factor model, with one factor accounting for 95% of the total variation. CFAs were repeated across all samples and supported a single factor model.
Internal consistency	.80-.87 across 7 studies
Reliability	The temporal stability of the scale was examined in an independent sample of 60 introductory psychology students over a 4-week period. ICC =.81. Time 1 (3.78) and Time 2 (3.77) mean scale scores were not significantly different. Repeated across all samples
Criterion Validity	The MMS was, as expected, correlated with the MAAS.
Construct validity/Convergent/ Divergent Validity	Overall, the pattern of associations indicates that higher scorers on the MAAS tend to be more aware of and receptive to inner experiences and are more mindful of their overt behavior. They are more "in tune" with their emotional states and able to alter them, and they are more likely to fulfill basic psychological needs.
Blinding/ Design	
Other characteristics	

Measure & Paper	**40. Basic Empathy Scale** (Jolliffe & Farrington, 2006)
N	Study 1: N=363, study 2: N=357
Age	Study 1: M=14.8
Gender	Study 1: 169 females, study 2: 175 females
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	Study 1: EFA with the whole group and groups split by gender suggested a two factor model. Goodness-of-fit indices indicated a good fit.
Internal consistency	
Reliability	
Criterion Validity	The similarity between the total of the BES and the total of the IRI for males and females was reflected in their relatively high inter-correlations (r ¼ :53 males, r ¼ :43 females).
Construct validity/Convergent/ Divergent Validity	Generally, the results of the comparisons were in line with previous research and theoretical expectation, and therefore support the validity of the BES. For example, in line with previous empathy research females scored much higher than males on all scales of the BES and there was significant overlap evident between cognitive and affective empathy as measured by the BES.
Blinding/ Design	
Other characteristics	

Measure & Paper	**41. Empathy Quotient** (Lawrence et al., 2004) EQ-40 (with 20 additional filler items)
N	Study 1: N=53, study 2&4: non-clinical group N=110 clinical group N=62, study 3: N=24
Age	Study 1: M=32.5, study 2&4: clinical group M=34.6, study 3: N=24
Gender	Study 1: 52.8% female, study 2&4: non-clinical group 61.4% female clinical group 48.4% female, study 3: N=24
Ethnicity/Nationality	
Diagnoses	clinical group in study 2 & 4 people reporting depersonalization symptoms or diagnosed with depersonalization disorder (DPD)
Content Validity	
Structural Validity	study 2: exploratory factor analysis (PCA with varimax rotation suggested three factors, explaining 41.4% of the variance.
Internal consistency	
Reliability	study 3: test–retest correlation coefficient between EQs administered 10-12 months after study 1 was r=0.835 (n=25, p=0.0001).
Criterion Validity	

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Construct validity/Convergent/ Divergent Validity	study 2: as hypothesised, there were significant differences in EQ score by gender. Study 4: No significant differences were found on total EQ score between the psychologically healthy individuals and those with DPD
Blinding/ Design	
Other characteristics	

Measure & Paper	**42. Kentucky Inventory of Mindfulness—Describe and Act With Awareness subscales** (Baer et al., 2004)
N	study 1: N=11 study 2&6: non-clinical group N=420 clinical group N=26 study 3: N=49 study 4: N=130 study 5: N=115
Age	study 2&6: non-clinical 18-22 clinical 20-52 M=36 study 4: M=19.6
Gender	study 1: 5 psychologists and 3/6 doctoral students female, study 2&6: non-clinical about 60% female clinical 25/26 female study 4: 56% female
Ethnicity/Nationality	study 2&6: non-clinical about 85% caucasian clinical all caucasian study 4: 86% White, 9% Black, and 5% other races or ethnicities.
Diagnoses	study 2&6: clinical group people with BPD diagnoses
Content Validity	Study 1: Content validity was assessed using ratings by experts: 5 psychologists who were experienced DBT therapists and 6 doctoral students in clinical psychology who had completed a graduate course on DBT and had led DBT skills groups for at least 1 year under the supervision of a certified DBT therapist. One student had also co-led an MBCT group. inter-rater agreement on which skill items represented 45-100% M = 86%
Structural Validity	study 2: an exploratory factor analysis using data from the non-clinical group suggested 9 factors explaining 64% of the variance, but the scree plot indicated a four factor solution. A four factor solution explaining 43% of the variance was found. A confirmatory factor analysis using data from non-clinical and clinical groups compared single and 4 factor models. Goodness-of-fit indices indicated a good fit for the 4 factor model and poor fit for the single factor model.
Internal consistency	study 2: non-clinical group Alpha coefficients for the 4 subscales ranged from .76 to .91
Reliability	study 3: Paired samples t tests showed no significant differences between test-retest scores with a 14-17 day interval. Test-retest correlations were .65-.86
Criterion Validity	The expected pattern of correlations between KIMS and MAAS scores, with a non-significant correlation for the observe scale only.
Construct validity/Convergent/ Divergent Validity	Study 4: As expected, neuroticism (NEO-FFI) was negatively correlated with most KIMS scores, emotional intelligence (TMMS) showed significant positive correlations with mindfulness scores, alexithymia (TAS) showed significant negative correlations with mindfulness scores, and relationships between impression management (PDS-IM) and the KIMS scales were small. Correlations between KIMS scores and the DES were small although negative as predicted. Study 6: the BPD group scored significantly lower than the student groups.
Blinding/ Design	
Other characteristics	

Measure & Paper	**43. Toronto Alexithymia Questionnaire** (Bagby et al., 1994) TAS-20
N	Study 1: N=965 (test-retest sample N=72) Study 2: non-clinical N=401 clinical N=218
Age	Study 1: M=21.8 (test-retest sample M=20.8) study 2: non-clinical M=21.1 clinical M=35.2
Gender	study 1: 576 females (test-retest sample 48 females) study 2: non-clinical 242 females clinical 124 females
Ethnicity/Nationality	
Diagnoses	study 2: clinical group psychiatric outpatients
Content Validity	
Structural Validity	study 1: exploratory factor analysis found a 3 factor solution explaining 27.9% of the variance. Study 2: confirmatory factor analysis confirmed a three factor solution had the best fit.
Internal consistency	Cronbach's alpha = 0.81
Reliability	study 1: test-retest correlation of measures taken 3 weeks apart was 0.77 ($p < 0.01$)
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	
Other characteristics	

Measure & Paper	**44. Difficulties in Emotion Regulation Scale—Lack of Emotional Clarity/Awareness subscales** (Gratz & Roemer, 2004)
N	N=357 (test-retest sample N=21)
Age	18-55 years M=23.10 (test-retest sample 18-48 years M=25.95)
Gender	73% female (test-retest sample 62% female)
Ethnicity/Nationality	65% White, 17% Asian, 8% Black/African American, 4% Hispanic, 6% another or unspecified racial background. (test-retest sample 67% White, 24% Black/African American, 5% Asian/Pacific Islander, 5% unspecified racial/ethnic background.)
Diagnoses	
Content Validity	
Structural Validity	factor analysis with principal axis factoring method of extraction and promax oblique rotation suggested a 6 factor solution that explained 55.68% of the variance.
Internal consistency	alpha = .93, >.80 for each subscale
Reliability	good test–retest reliability over a period ranging from 4 to 8 weeks (=.88, p <.01), and adequate for subscales (.57-.89)
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Correlations between the DERS score and emotional regulation (NMR), experiential avoidance, emotional expressivity, self-harm and intimate partner abuse (for men but not women)were in the expected directions and statistically significant.
Blinding/ Design	
Other characteristics	

Measure & Paper	**45. Metacognition in Multiple Contexts Inventory.** (Allen & Armour-Thomas, 1993)
N	N=126
Age	9-11 grade high school students

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Gender	75 females
Ethnicity/Nationality	62% african american, 28% hispanic american, 10% anglo american
Diagnoses	
Content Validity	
Structural Validity	principal components analysis - 9 factors accounted for 62% of common variance among items. Factors were interdependent
Internal consistency	0.67 Cronbach's alpha
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	
Other characteristics	

Measure & Paper	**46. Metacognitions Questionnaire 30** (Cook et al., 2014) **MCQ-30**
N	N=182 (test-retest N=30)
Age	18–69 M=33.5 years (test-retest 18–53 M=27.24)
Gender	119 females, 3 did not indicate gender (test-retest 19 women)
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	EFA indicated a five factor model, goodness-of-fit indices indicated a good fit. Goodness-of-fit indices indicated an adequate fit in a CFA at 12 month follow up.
Internal consistency	Cronbach's alpha ranged from 0.73 to 0.89 pre-treatment and from .79 to .91 at 12 month follow-up
Reliability	Retested 22-118 days later, adequate test-retest reliability for four out of five subscales (ranging from r = 0.59 'Negative beliefs about worry' to r = 0.87 'Cognitive self-consciousness').
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Acceptable fit was found in latent variable SEM of hypothesised model of the relationship between metacognitive beliefs and concurrent anxiety and depression.
Blinding/ Design	Convenience sampling, students and university and health service employees.
Other characteristics	

Measure & Paper	**47. Metacognitive Activities Inventory (Interactive MultiMedia Exercises software package, IMMEX - not reviewed as experimental measure paradigm)** (Sandi-Urena, 2009) **MCA-I**
N	pilot N=151 main study N=290 replication study N=537
Age	>18
Gender	
Ethnicity/Nationality	
Diagnoses	
Content Validity	Initial pool of items was obtained using a panel-of-experts technique (four faculty members and four graduate students). Face validity was explored by asking students in to contribute comments and observations about the scale. Changes derived from these observations were implemented.
Structural Validity	Exploratory unrestricted factor analyses using an orthogonal Varimax rotation with Kaiser normalization were performed independently in both cases. The first one produced a seven factor solution with eigenvalues greater than one that accounted for 51% of the total variance. The two-factor solution proved to be the most interpretable, and it resulted in 29% of the cumulative variance. In the main study the two-factor solution accounted for 46% of the cumulative variance and was completely unambiguous.
Internal consistency	alphas ranged from .85 to .92 in the main and replication studies
Reliability	main study: pre and post MCA-I scores were not significantly different at the .05 level, replication study: paired t-test p-value of .07 and were highly correlated (Pearson's coefficient .53, significant at .01 level).
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	
Other characteristics	

Measure & Paper	**48. Swanson's 15-item metacognition questionnaire** (Swanson, 1990)
N	n=80
Age	children, grades 4-7
Gender	26 girls and 30 boys
Ethnicity/Nationality	predominantly Anglo (97%)
Diagnoses	
Content Validity	
Structural Validity	
Internal consistency	alpha was .87
Reliability	Independent interrater reliability for each item was above 96%
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	As reported in previous metacognition literature the correlation between the MCA-I mean scores for the main study pretest and GPA was low (Pearson's coefficient of .16) yet significant (p-value = .015). The proportion of high-metacognition users is larger for higher grades
Blinding/ Design	
Other characteristics	scores were above 120 or below 105 on the Cognitive Abilities Test (CAT)

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Measure & Paper	**49. Awareness of Independent Learning Inventory (Meijer et al., 2013) English version**
N	N=729 (test-rest N=34)
Age	
Gender	
Ethnicity/Nationality	
Diagnoses	
Content Validity	The translator did not just translate the items, but also followed the original steps of collecting equivalent terms and reproducing the aforementioned negative linguistic formulations.
Structural Validity	
Internal consistency	
Reliability	retest correlations after 2 years were 0.46, 0.39 and 0.25 for metacognitive knowledge, regulation and responsiveness, respectively. The former two correlations are statistically significant ($p<0.05$); the latter is not.
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	High correlations (0.67<<0.73) between the three AILI factors and the MSLQ factors Cognitive and Metacognitive Strategies. Non-significant correlations (0.06<<0.08) as expected between the three AILI factors and test anxiety, as measured by MSLQ
Blinding/ Design	
Other characteristics	
Measure & Paper	**50. Motivated Strategies for Learning Questionnaire (Pintrich, et al., 1993)**
N	N=291
Age	
Gender	66% female
Ethnicity/Nationality	5% African-American, 3.2% Asian - American, 90.3% Cacasian, 1.5% hispanic or spanish speaking.
Diagnoses	
Content Validity	
Structural Validity	2 confirmatory factor analyses for cognitive and metacognitive strategy items. Parameter estimates for the model specified were generated using maximum likelihood, and tests for goodness of fit were made. Lambda-ksi estimates range from .38 to .89, with an average of .68 (median .66).
Internal consistency	alphas on the scales ranged from .52 to .93
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Motivational subscales were significantly correlated in the expected direction with student's final grades.
Blinding/ Design	
Other characteristics	
Measure & Paper	**51. Assessment of Cognitive Monitoring Effectiveness (Osborne, 1998)**
N	N = 79
Age	5th grade children
Gender	42 girls
Ethnicity/Nationality	100% white
Diagnoses	
Content Validity	
Structural Validity	
Internal consistency	
Reliability	Correlations between the spelling tests (r (38) = .74, p <.0001) and between the social studies tests (r (27)= .35, p <.07) were significant and in the expected direction. Correlations between the math tests (r (24) = -.06) were not significant. Although analogous, this does not produce a pure test-retest reliability statistic, as test-retest reliability generally assumes the same instrument is used on both occasions.
Criterion Validity	Generally strong multivariate relationships between ACME scores and test performance, cumulative grades, and academic performance were found.
Construct validity/Convergent/ Divergent Validity	A multiple correlation was computed between the Index of Reading Awareness (ERA, a measure of knowledge of reading strategy usage), the Metacognitive Knowledge Monitoring Assessment (KMA, a measure o f knowledge monitoring), and teacher ratings of student cognitive monitoring. The results of the multivariate analyses indicated generally strong convergence among the measures. Of six possible multiple correlations, four were significant, ranging from R = .42 to R = .80. Univariate analyses also revealed good convergence.
Blinding/ Design	
Other characteristics	
Measure & Paper	**52. Metacognition Scale (Ford et al., 1998)**
N	N=93
Age	
Gender	
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	
Internal consistency	
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	As hypothesized, metacognition was positively related to:, mastery orientation (beta = .22), knowledge (beta = .24), final training performance (beta = .29) and self-efficacy (beta = .32).
Blinding/ Design	

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Other characteristics	

Measure & Paper	**53. Metacognitive Awareness Inventory** (Schraw & Dennison, 1994)
N	study 1: N=197 study 2: N=110
Age	
Gender	Study 1: 112 females study 2: 69 females
Ethnicity/Nationality	
Diagnoses	
Content Validity	
Structural Validity	Study 1: Exploratory factor analysis suggested a 6 factor model that explained 78% of the variance. These were not easily interpretable, so a confirmatory factor analysis was performed with a 2 factor model, which accounted for 65% of the variance. Study 2: confirmatory factor analysis was rerun and the 2 factor solution accounted for 58% of the variance.
Internal consistency	study 1: alpha = .90
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Mixed results were found in a series of studies testing the hypotheses that MAI score would be related to monitoring accuracy, test performance and pre-test judgements.
Blinding/ Design	
Other characteristics	

Measure & Paper	**54. Metacognitive subscale of the Cultural Intelligence Scale (CQ)** (Ang et al., 2007)
N	sample 1: N=235, sample 2: N=358
Age	sample 1: M=22, sample 2: M=19
Gender	sample 1: 45% female, sample 2: 76% female
Ethnicity/Nationality	
Diagnoses	
Content Validity	A panel of three faculty and three international executives (each with significant cross-cultural expertise) independently assessed the randomly ordered initial 53 items for clarity, readability and definitional fidelity. We retained the 10 best items for each dimension (40 items).
Structural Validity	Confirmatory factor analysis demonstrated acceptable goodness-of-fit indices in both samples of the four factor structure of the CQ.
Internal consistency	a = 0.76/0.70
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	As expected, the four CQ factors moderately and positively related to Emotional Intelligence. Eleven of the 16 correlations between the four factors of CQ and the four factors of the Cross cultural adaptability inventory were significant. Discriminant validity was demonstrated as the variance shared between the CQ and any other construct in the model (the squared correlation between the two constructs) was less than the variance that CQ shared with its measures.
Blinding/ Design	
Other characteristics	

Measure & Paper	**55. Toronto Structured Interview for Alexithymia** (Bagby et al., 2006)
N	main study: non-clinical N=136, clinical N=97 test-retest N=22
Age	non-clinical M=32.3 (men) M=32.9 (women), clinical M=42.4 (men) M=38.6 (women)
Gender	non-clinical 95 women, clinical 77 women
Ethnicity/Nationality	
Diagnoses	
Content Validity	Questions were eliminated when there was consensus among the interviewers/raters that the questions were difficult for the research participant to understand or for the interviewers and rater(s) to code.
Structural Validity	An exploratory factor analysis suggested a four factor solution, which was interpretable and accounted for nearly 50% of the variance. A confirmatory factor analysis compared 2, 3 and 4 factor models, with four factor models demonstrating the best fit with a range of goodness-of-fit indices.
Internal consistency	
Reliability	The inter-rater reliability estimates for the TSIA and subscales in both samples were statistically significant; and exceeded 0.60, indicating adequate interrater reliability. Test-rest reliability was assessed via intraclass correlations. Estimates for the TSIA and subscales were all statistically significant and all but two estimates exceeded 0.60.
Criterion Validity	In both samples, the TSIA and TAS-20 correlated significantly with one another.
Construct validity/Convergent/ Divergent Validity	
Blinding/ Design	An expert rater did not conduct any interviews but rated all transcripts. Other raters had conducted interviews.
Other characteristics	

Measure & Paper	**56. The Mentalizing Emotions Questionnaire** (Kasper et al., 2024)
N	Study 1: N=510, Study 2: N=509
Age	Study 1: 18-65 years, M=43.3, SD=13.8, Study 2: 18-65 years M=44.0, SD=13.2
Gender	Study 1: 50.0% female, 49.4% male, 0.6% diverse, Study 2: 53.3% female, 46.2% male, 0.6% diverse
Ethnicity/Nationality	
Diagnoses	Study 1: 383 participants did not suffer from mental disorder during the last year Study 2: 377 participants did not suffer from mental disorder during the last year, 132 participants did suffer from mental disorder during the last year, of these 78 went into treatment, and of these 57 were in outpatient treatment.

(*continued on next page*)

(*continued*)

Measure & Paper	1. Reflective Functioning Rating Scale (Meehan et al., 2009)
Content Validity	Questionnaire development: RF Scale used in initial development and an iterative, peer consensus (with MBT experts and certified RF-raters) process used and social desirability ratings considered.
Structural Validity	An exploratory factor analysis suggested a three factor solution, which accounted for 65% of the variance and was interpretable. A confirmatory factor analysis confirmed acceptable fit of this model.
Internal consistency	a = 0.95
Reliability	
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Significant correlations in expected directions with CAMSQ self and other, MZQ, ACS and ETMCQ trust and mistrust, and incremental validity assumed following finding of a significant multiple regression model with the CAMSQ, MZQ, RFQ-6, MEQ as independent variables and the EQ as a dependent variable.
Blinding/ Design	
Other characteristics	

Measure & Paper	**57. The Metacognition in Self-Control Scale** (Bürgler et al., 2022)
N	Study 1: N=315, Study 2: N=503
Age	Study 1: age groups 18-30 (N = 50), 31-40 (N = 55), 41-50 (N = 51), 51-60 (N = 51), 61-70 (N = 52), > 71 (N = 49) Study 2: M=25, SD = 6.5
Gender	Study 1: 153 females and 155 males Study 2: 69% were female, 30% male, and 1% other.
Ethnicity/Nationality	
Diagnoses	
Content Validity	Based on the MAI scale by Schraw and Dennison (1994), with use of expert feedback
Structural Validity	An exploratory factor analysis suggested a two factor solution, which accounted for 55% of the variance and was interpretable. A confirmatory factor analysis did not show a good enough fit. Items were reduced further and another confirmatory facto analysis compared one-factor, two-factor, orthogonal two-factor, and six-factor models, with the two factor model showing an adequate fit.
Internal consistency	a = 0.82
Reliability	Adequate test-retest reliability r = .54 (p < .001)
Criterion Validity	
Construct validity/Convergent/ Divergent Validity	Higher MISCS scores predicted more successful resolution of self-control conflicts in the 10 day experience sampling study. All correlations between the MISCS and other variables assessed at baseline are as hypothesized.
Blinding/ Design	
Other characteristics	

Appendix F. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2025.102657.

Data availability

Data available via OSF link https://osf.io/cwd6z/overview?view_only=f8287943cd4e4923aea54795340fd8c4

References

Aival-Naveh, E., Rothschild-Yakar, L., & Kurman, J. (2019). Keeping culture in mind: A systematic review and initial conceptualization of mentalizing from a cross-cultural perspective. *Clinical Psychology: Science and Practice, 26*(4), Article e12300. https://doi.org/10.1111/cpsp.12300

Aival-Naveh, E., Rothschild-Yakar, L., Park, J., & Kurman, J. (2022). The value of thinking about feelings across cultures: a preliminary investigation of the mentalizing values scale. *Journal of Cross-Cultural Psychology, 53*(3–4), 362–379. https://doi.org/10.1177/00220221221077355

Allen, B. A., & Armour-Thomas, E. (1993). Construct validation of metacognition. *Journal of Psychology, 127*(2), 203–211.

Allen, J. G. (2003). Mentalizing. *Bulletin of the Menninger Clinic, 67*(2), 91–112. https://doi.org/10.1521/bumc.67.2.91.23440

Allen, T. A., Rueter, A. R., Abram, S. V., Brown, J. S., & Deyoung, C. G. (2017). Personality and neural correlates of mentalizing ability. *European Journal of Personality, 31*(6), 599–613. https://doi.org/10.1002/per.2133

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Ang, S., Van Dyne, L., Koh, C., Ng, K. Y., Templer, K. J., Tay, C., & Chandrasekar, N. A. (2007). Cultural intelligence: Its measurement and effects on cultural judgment and decision making, cultural adaptation and task performance. *Management and Organization Review, 3*(3), 335–371. https://doi.org/10.1111/j.1740-8784.2007.00082.x

Baer, R. A., Smith, G. T., & Allen, K. B. (2004). Assessment of mindfulness by self-report: The Kentucky inventory of mindfulness skills. *Assessment, 11*(3), 191–206. https://doi.org/10.1177/1073191104268029

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Bagby, R. M., Taylor, G. J., Parker, J. D. A., & Dickens, S. E. (2006). The development of the toronto structured interview for alexithymia: Item selection, factor structure, reliability and concurrent validity. *Psychotherapy and Psychosomatics, 75*(1), 25–39. https://doi.org/10.1159/000089224

Ballespí, S., Vives, J., Sharp, C., Chanes, L., & Barrantes-Vidal, N. (2021). Self and other mentalizing polarities and dimensions of mental health: Association with types of symptoms, functioning and well-being. *Frontiers in Psychology, 12*(566254). https://doi.org/10.3389/fpsyg.2021.566254

Barnby, J. M., Bellucci, G., Alon, N., Schilbach, L., Frith, C. D., & Bell, V. (2023). *Beyond theory of mind: A formal interoperable framework for social inference and representation*. OSF. https://doi.org/10.31234/osf.io/cmgu7

Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Baron-Cohen, S. (2000). The evolution of a theory of mind. In M. Corballis, & S. E. G. Lea (Eds.), *The Descent of Mind: Psychological Perspectives on Hominid Evolution* (pp. 261–277). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780192632593.003.0013.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 241–251. https://doi.org/10.1111/1469-7610.00715

Barone, P., Corradi, G., & Gomila, A. (2019). Infants' performance in spontaneous-response false belief tasks: A review and meta-analysis. *Infant Behavior and Development, 57*, Article 101350. https://doi.org/10.1016/j.infbeh.2019.101350

Bateman, A., & Fonagy, P. (2004). Mentalization-based treatment of BPD. *Journal of Personality Disorders, 18*(1), 36–51. https://doi.org/10.1521/pedi.18.1.36.32772

Bateman, A., & Fonagy, P. (2016). *Mentalization-Based Treatment for Personality Disorders: A Practical Guide* ((1st ed.).). University Press.

Bateman, A., & Fonagy, P. (2019). *Handbook of Mentalizing in Mental Health Practice* ((2nd ed.).). American Psychiatric Publishing http://ebookcentral.proquest.com/lib/ucl/detail.action?docID=5790646.

Bender, D. S., & Skodol, A. E. (2007). Borderline personality as a self-other representational disturbance. *Journal of Personality Disorders, 21*(5), 500–517. https://doi.org/10.1521/pedi.2007.21.5.500

Bosco, F. M., Colle, L., Fazio, S. D., Bono, A., Ruberti, S., & Tirassa, M. (2009). Th.o.m.a.s.: An exploratory assessment of Theory of Mind in schizophrenic subjects.

Consciousness and Cognition, 18(1), 306–319. https://doi.org/10.1016/j.concog.2008.06.006
Bouchard, M.-A., Target, M., Lecours, S., Fonagy, P., Tremblay, L.-M., Schachter, A., & Stein, H. (2008). Mentalization in adult attachment narratives: Reflective functioning, mental states, and affect elaboration compared. *Psychoanalytic Psychology, 25*(1), 47–66. https://doi.org/10.1037/0736-9735.25.1.47
Bräuer, J., Hanus, D., Pika, S., Gray, R., & Uomini, N. (2020). Old and new approaches to animal cognition: There is not "one cognition". *Journal of Intelligence, 8*(3), 28. https://doi.org/10.3390/jintelligence8030028
Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822–848. https://doi.org/10.1037/0022-3514.84.4.822
Buck, B., Ludwig, K., Meyer, P. S., & Penn, D. L. (2014). The use of narrative sampling in the assessment of social cognition: The Narrative of Emotions Task (NET). *Psychiatry Research, 217*(3), 233–239. https://doi.org/10.1016/j.psychres.2014.03.014
Bürgler, S., Kleinke, K., & Hennecke, M. (2022). The metacognition in self-control scale (MISCS). *Personality and Individual Differences, 199*, Article 111841. https://doi.org/10.1016/j.paid.2022.111841
Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences, 12*(5), 187–192. https://doi.org/10.1016/j.tics.2008.02.010
Camerlinck, M.-F., Saunders, R., Fonagy, P., & Nolte, T. (2024). *A comparison of mentalizing capacity between clinical and non-clinical samples: A systematic review and meta-analysis of the movie for the Assessment of Social Cognition (MASC)* [Manuscript submitted for publication].
Canty, A. L., Neumann, D. L., Fleming, J., & Shum, D. H. K. (2017). Evaluation of a newly developed measure of theory of mind: The virtual assessment of mentalising ability. *Neuropsychological Rehabilitation, 27*(5), 834–870. https://doi.org/10.1080/09602011.2015.1052820
Carpenter, J. M., Green, M. C., & Vacharkulksemsuk, T. (2016). Beyond perspective-taking: Mind-reading motivation. *Motivation and Emotion, 40*(3), 358–374. https://doi.org/10.1007/s11031-016-9544-z
Chapple, M., Davis, P., Billington, J., Myrick, J. A., Ruddock, C., & Corcoran, R. (2021). Overcoming the double empathy problem within pairs of autistic and non-autistic adults through the contemplation of serious literature. *Frontiers in Psychology, 12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.708375.
Choi-Kain, L. W., & Gunderson, J. G. (2008). Mentalization: Ontogeny, assessment, and application in the treatment of borderline personality disorder. *American Journal of Psychiatry, 165*(9), 1127–1135. https://doi.org/10.1176/appi.ajp.2008.07081360
Colle, L., Gabbatore, I., Riberi, E., Borroz, E., Bosco, F. M., & Keller, R. (2019). Mindreading abilities and borderline personality disorder: A comprehensive assessment using the Theory of Mind Assessment Scale. *Psychiatry Research, 272*, 609–617. https://doi.org/10.1016/j.psychres.2018.12.102
Cook, S. A., Salmon, P., Dunn, G., & Fisher, P. (2014). Measuring metacognition in cancer: Validation of the metacognitions questionnaire 30 (MCQ-30). *PLoS ONE, 9*(9), Article e107302. https://doi.org/10.1371/journal.pone.0107302
Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology, 44*(1), 113–126. https://doi.org/10.1037/0022-3514.44.1.113
De Meulemeester, C., Lowyck, B., & Luyten, P. (2021). The role of impairments in self–other distinction in borderline personality disorder: A narrative review of recent evidence. *Neuroscience and Biobehavioral Reviews, 127*, 242–254. https://doi.org/10.1016/j.neubiorev.2021.04.022
Dennett, D. C. (1987). *The intentional stance*. MIT Press.
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
Dinsdale, N., & Crespi, B. J. (2013). The borderline empathy paradox: Evidence and conceptual models for empathic enhancements in borderline personality disorder. *Journal of Personality Disorders, 27*(2), 172–195.
Duschinsky, R., & Foster, S. (2021). *Mentalizing and Epistemic Trust: The Work of Peter Fonagy and Colleagues at the Anna Freud Centre*. Oxford University Press.
Dziobek, I., Fleck, S., Kalbe, E., Rogers, K., Hassenstab, J., Brand, M., … Convit, A. (2006). Introducing MASC: A movie for the assessment of social cognition. *Journal of Autism and Developmental Disorders, 36*(5), 623–636. https://doi.org/10.1007/s10803-006-0107-0
Dziobek, I., Preißler, S., Grozdanovic, Z., Heuser, I., Heekeren, H. R., & Roepke, S. (2011). Neuronal correlates of altered empathy and social cognition in borderline personality disorder. *NeuroImage, 57*(2), 539–548. https://doi.org/10.1016/j.neuroimage.2011.05.005
Ekdahl, D. (2023). The double empathy problem and the problem of empathy: Neurodiversifying phenomenology. *Disability & Society, 1–23*. https://doi.org/10.1080/09687599.2023.2220180
van Elk, M., Maij, D., & Rutjens, B. (2020). Development and validation of a porous theory of mind scale. *Journal of Cognition and Culture, 20*(1–2), 41–65. https://doi.org/10.1163/15685373-12340073
Euler, S., Nolte, T., Constantinou, M., Griem, J., Montague, P. R., Fonagy, P., & Personality and Mood Disorders Research Network. (2021). Interpersonal Problems in Borderline Personality Disorder: Associations With Mentalizing, Emotion Regulation, and Impulsiveness. *Journal of Personality Disorders, 35*(2), 177–193. https://doi.org/10.1521/pedi_2019_33_427
Fischer-Kern, M., Buchheim, A., Hörz, S., Schuster, P., Doering, S., Kapusta, N. D., … Fonagy, P. (2010). The relationship between personality organization, reflective functioning, and psychiatric classification in borderline personality disorder. *Psychoanalytic Psychology, 27*(4), 395–409. https://doi.org/10.1037/a0020862
Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science, 3*(4), 456–465. https://doi.org/10.1177/2515245920952393
Flavell, J. H. (1979). Metacognition and cognitive monitoring. A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906
Fonagy, P. (1991). Thinking about thinking: Some clinical and theoretical considerations in the treatment of a borderline patient. *International Journal of Psychoanalysis, 72*(4), 639–656.
Fonagy, P., & Bateman, A. W. (2006). Mechanisms of change in mentalization-based treatment of BPD. *Journal of Clinical Psychology, 62*(4), 411–430. https://doi.org/10.1002/jclp.20241
Fonagy, P., & Campbell, C. (2016). Attachment theory and mentalization. In A. Elliott, & J. Prager (Eds.), *The Routledge Handbook of Psychoanalysis in the Social Sciences and Humanities* (pp. 115–131). Routledge.
Fonagy, P., & Leigh.. (1996). The relation of attachment status, psychiatric classification, and response to psychotherapy. *Journal of Consulting and Clinical Psychology, 64*(1), 22–31. https://doi.org/10.1037/0022-006X.64.1.22
Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and Psychopathology, 21*(4), 1355–1381. https://doi.org/10.1017/S0954579409990198
Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., Howard, S., … Lowyck, B. (2016). Development and validation of a self-report measure of mentalizing: The reflective functioning questionnaire. *PLoS ONE, 11*(7), Article e0158678. https://doi.org/10.1371/journal.pone.0158678
Fonagy, P., Steele, H., & Steele, M. (1991). Maternal representations of attachment during pregnancy predict the organization of infant-mother attachment at one year of age. *Child Development, 62*(5), 891–905. https://doi.org/10.2307/1131141
Fonagy, P., & Target, M. (1996). Playing with reality: I. Theory of mind and the normal development of psychic reality. *The International Journal of Psychoanalysis, 77*(2), 217–233.
Fonagy, P., & Target, M. (1997). Attachment and reflective function: Their role in self-organization. *Development and Psychopathology, 9*(4), 679–700. https://doi.org/10.1017/S0954579497001399
Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-Functioning manual: Version 5 for application to adult attachment interviews* [Unpublished manual]. University College London
Ford, J. K., Weissbein, D. A., Smith, E. M., & Gully, S. M. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology, 83*(2), 218–233. https://doi.org/10.1037/0021-9010.83.2.218
Fossati, A., Borroni, S., Dziobek, I., Fonagy, P., & Somma, A. (2016). Thinking about assessment: Further evidence of the validity of the Movie for the Assessment of Social Cognition as a measure of mentalistic abilities. *Psychoanalytic Psychology, 35*(1), 127–141. https://doi.org/10.1037/pap0000130
Frith, C. D., Wolpert, D. M., Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218
Gagliardini, G., & Colli, A. (2019). Assessing mentalization: Development and preliminary validation of the modes of mentalization scale. *Psychoanalytic Psychology, 36*(3), 249–258. https://doi.org/10.1037/pap0000222
Gagliardini, G., Gullo, S., Caverzasi, E., Boldrini, A., Blasi, S., & Colli, A. (2018). Assessing mentalization in psychotherapy: First validation of the Mentalization Imbalances Scale. *Research in Psychotherapy : Psychopathology, Process, and Outcome, 21*(3), 339. https://doi.org/10.4081/ripppo.2018.339
Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology, 2*(3), 237–248. https://doi.org/10.1002/ajp.1350020302
Gopnik, A., & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language, 7*(1–2), 145–171. https://doi.org/10.1111/j.1468-0017.1992.tb00202.x
Gori, A., Arcioni, A., Topino, E., Craparo, G., & Lauro Grotto, R. (2021). Development of a new measure for assessing mentalizing: The multidimensional mentalizing questionnaire (MMQ). *Journal of Personalized Medicine, 11*(4), 305. https://doi.org/10.3390/jpm11040305
Gori, A., & Topino, E. (2023). Exploring and deepening the facets of mentalizing: The integration of network and factorial analysis approaches to verify the psychometric properties of the multidimensional mentalizing questionnaire (MMQ). *International Journal of Environmental Research and Public Health, 20*(6), Article 6. https://doi.org/10.3390/ijerph20064744
Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment, 26*(1), 41–54. https://doi.org/10.1023/B:JOBA.0000007455.08539.94
Greenberg, D. M., Kolasi, J., Hegsted, C. P., Berkowitz, Y., & Jurist, E. L. (2017). Mentalized affectivity: A new model and assessment of emotion regulation. *PLoS ONE, 12*(10), Article e0185264. https://doi.org/10.1371/journal.pone.0185264
Greenberg, D. M., Rudenstine, S., Alaluf, R., & Jurist, E. L. (2021). Development and validation of the Brief-Mentalized Affectivity Scale: Evidence from cross-sectional online data and an urban community-based mental health clinic. *Journal of Clinical Psychology, 77*(11), 2638–2652. https://doi.org/10.1002/jclp.23203
Hausberg, M. C., Schulz, H., Piegler, T., Happach, C. G., Klöpper, M., Brütt, A. L., … 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

Hawksdale, H.. Measures of reflective functioning. https://osf.io/cwd6z/overview?view_only=f8287943cd4e4923aea54795340fd8c4.
Healey, K. M., Combs, D. R., Gibson, C. M., Keefe, R. S. E., Roberts, D. L., & Penn, D. L. (2015). Observable social cognition – a rating scale: An interview-based assessment for schizophrenia. *Cognitive Neuropsychiatry, 20*(3), 198–221. https://doi.org/10.1080/13546805.2014.999915
Horváth, Z., Demetrovics, O., Paksi, B., Unoka, Z., & Demetrovics, Z. (2023). The Reflective Functioning Questionnaire–Revised–7 (RFQ-R-7): A new measurement model assessing hypomentalization. *PLoS ONE, 18*(2), Article e0282000. https://doi.org/10.1371/journal.pone.0282000
Hutchins, T. L., Lewis, L., Prelock, P. A., & Brien, A. (2021). The development and preliminary psychometric evaluation of the Theory of Mind Inventory: Self report-adult (ToMI:SR-Adult). *Journal of Autism and Developmental Disorders, 51*(6), 1839–1852. https://doi.org/10.1007/s10803-020-04654-6
Jańczak, M. (2018). Mentalization in borderline individuals: an attempt to integrate contradictory research results. *Current Issues in Personality Psychology, 6*(4), 266–278. https://doi.org/10.5114/cipp.2018.80196
Jańczak, M. (2021). *What's on your mind?* Wydawnictwo Rys: A guide for Mentalization Assessment in Adults. https://doi.org/10.48226/978-83-66666-85-6.
Jolliffe, D., & Farrington, D. P. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence, 29*(4), 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010
Kahneman, D. (2012). Two systems in the mind. *Bulletin of the American Academy of Arts and Sciences, 65*(2), 55–59.
Kasper, L. A., Hauschild, S., Berning, A., Holl, J., & Taubner, S. (2024). Development and validation of the Mentalizing Emotions Questionnaire: A self-report measure for mentalizing emotions of the self and other. *PLoS ONE, 19*(5), Article e0300984. https://doi.org/10.1371/journal.pone.0300984
Katznelson, H. (2014). Reflective functioning: A review. *Clinical Psychology Review, 34*(2), 107–117. https://doi.org/10.1016/j.cpr.2013.12.003
Kircanski, K., Joormann, J., & Gotlib, I. H. (2012). Cognitive aspects of depression. *WIREs Cognitive Science, 3*(3), 301–313. https://doi.org/10.1002/wcs.1177
Kivity, Y., Levy, K. N., Johnson, B. N., Rosenstein, L. K., & LeBreton, J. M. (2024). Mentalizing in and out of awareness: A meta-analytic review of implicit and explicit mentalizing. *Clinical Psychology Review.* , Article 102395. https://doi.org/10.1016/j.cpr.2024.102395
Kutney, A. M. (2006). An examination of psychiatric-mental health outcomes from the perspectives of logical positivism and phenomenology. *Journal of the American Psychiatric Nurses Association, 12*(1), 22–27. https://doi.org/10.1177/1078390306286443
Lakhani, S., Bhola, P., & Mehta, U. M. (2021). The conceptualization and assessment of social cognition in personality and common mental disorders. *Asian Journal of Psychiatry, 65*, Article 102829. https://doi.org/10.1016/j.ajp.2021.102829
Lawrence, E. J., Shaw, P., Baker, D., Baron-Cohen, S., & David, A. S. (2004). Measuring empathy: Reliability and validity of the Empathy Quotient. *Psychological Medicine, 34*(5), 911–920. https://doi.org/10.1017/S0033291703001624
Lecours, S., Bouchard, M.-A., St-Amand, P., & Perry, J. C. (2000). Assessing verbal elaboration of affect in psychotherapy: A preliminary report and single case study. *Psychotherapy Research, 10*(1), 47–56. https://doi.org/10.1080/713663593
Lee-Parritz, O. (2015). *Assessing Mentalization: Toward Construct Validation of Mental State Discourse Analysis as Applied to the Thematic Apperception Test* [Unpublished doctoral dissertation]. William James College https://www.proquest.com/docview/1906687787/abstract/71A590CEA7644B5BPQ/1
Leichsenring, F., Fonagy, P., Heim, N., Kernberg, O. F., Leweke, F., Luyten, P., … Steinert, C. (2024). Borderline personality disorder: A comprehensive review of diagnosis and clinical presentation, etiology, treatment, and current controversies. *World Psychiatry, 23*(1), 4–25. https://doi.org/10.1002/wps.21156
Linehan, M. (1993). *Cognitive-behavioral Treatment of Borderline Personality Disorder.* Guilford Press.
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
Luyten, P., Campbell, C., Moser, M., & Fonagy, P. (2024). The role of mentalizing in psychological interventions in adults: Systematic review and recommendations for future research. *Clinical Psychology Review, 108*, Article 102380. https://doi.org/10.1016/j.cpr.2024.102380
Luyten, P., Fonagy, P., Lowyck, B., & Vermote, R. (2012). Assessment of mentalization. In A. W. Bateman, & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 43–65). American Psychiatric Publishing, Inc.
Lysaker, P. H., Carcione, A., Dimaggio, G., Johannesen, J. K., Nicolò, G., Procacci, M., & Semerari, A. (2005). Metacognition amidst narratives of self and illness in schizophrenia: Associations with neurocognition, symptoms, insight and quality of life. *Acta Psychiatrica Scandinavica, 112*(1), 64–71. https://doi.org/10.1111/j.1600-0447.2005.00514.x
Lysaker, P. H., Hillis, J., Leonhardt, B. L., Kukla, M., & Buck, K. D. (2014). Chapter 6 - Metacognition in Schizophrenia Spectrum Disorders: Methods of Assessment and Associations with Psychosocial Function, Neurocognition, Symptoms, and Cognitive Style. In P. H. Lysaker, G. Dimaggio, & M. Brüne (Eds.), *Social Cognition and Metacognition in Schizophrenia* (pp. 99–113). Academic Press. https://doi.org/10.1016/B978-0-12-405172-0.00006-5.
MacLennan, C. (2021). *Measures of States of Mind: A Scoping Review [Unpublished manuscript]*.
Malda-Castillo, J., Browne, C., & Perez-Algorta, G. (2019). Mentalization-based treatment and its evidence-base status: A systematic literature review. *Psychology and Psychotherapy: Theory, Research and Practice, 92*(4), 465–498. https://doi.org/10.1111/papt.12195
Marty, P. (1990). *La Psychosomatique de l'Adulte.* Presses Universitaire de France.
McMahon, J. M., & Good, D. J. (2016). The moral metacognition scale: Development and validation. *Ethics & Behavior, 26*(5), 357–394. https://doi.org/10.1080/10508422.2015.1028548
Meehan, K. B., Levy, K. N., Reynoso, J. S., Hill, L. L., & Clarkin, J. F. (2009). Measuring reflective function with a multidimensional rating scale: Comparison with scoring reflective function on the AAI. *Journal of the American Psychoanalytic Association, 57*(1), 208–213. https://doi.org/10.1177/0003065109570011008
Meijer, J., Sleegers, P., Elshout-Mohr, M., van Daalen-Kapteijns, M., Meeus, W., & Tempelaar, D. (2013). The development of a questionnaire on metacognition for students in higher education. *Educational Research, 55*(1), 31–52. https://doi.org/10.1080/00131881.2013.767024
Milton, D. E. M. (2012). On the ontological status of autism: The 'double empathy problem'. *Disability & Society, 27*(6), 883–887. https://doi.org/10.1080/09687599.2012.710008
Mitchell, L. J., Gumley, A., Reilly, E. S., Macbeth, A., Lysaker, P., Carcione, A., & Dimaggio, G. (2012). Metacognition in forensic patients with schizophrenia and a past history of interpersonal violence: An exploratory study. *Psychosis, 4*(1), 42–51. https://doi.org/10.1080/17522439.2011.630098
Moessnang, C., Baumeister, S., Tillmann, J., Goyard, D., Charman, T., Ambrosino, S., … the EU-AIMS LEAP group. (2020). Social brain activation during mentalizing in a large autism cohort: The Longitudinal European Autism Project. *Molecular Autism, 11*(1), 17. https://doi.org/10.1186/s13229-020-0317-x
Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., … PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews, 4*, 1. https://doi.org/10.1186/2046-4053-4-1
Mokkink, L. B., Boers, M., van der Vleuten, C. P. M., Bouter, L. M., Alonso, J., Patrick, D. L., … Terwee, C. B. (2020). COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: A Delphi study. *BMC Medical Research Methodology, 20*, Article 293. https://doi.org/10.1186/s12874-020-01179-5
Morton, J., & Frith, U. (1995). Causal modeling: A structural approach to developmental psychology. In D. Cicchetti, & D. J. Cohen (Eds.), *Theory and methods: 1. Developmental psychopathology* (pp. 357–390). New York: John Wiley.
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
NICE. (2009). NICE guideline [CG78] Borderline personality disorder: Recognition and management. https://www.nice.org.uk/guidance/cg78.
Nijhof, A. D. (2023). Recent neural advances in studies on theory of mind and autism. *Social Neuroscience, 18*(4), 205–206. https://doi.org/10.1080/17470919.2023.2259535
Nolte, T., Bolling, D. Z., Hudac, C. M., Fonagy, P., Mayes, L., & Pelphrey, K. A. (2013). Brain mechanisms underlying the impact of attachment-related stress on social cognition. *Frontiers in human neuroscience, 7*, 816.
Orr, R. I., & Gilead, M. (2022). Development and validation of the Mental-Physical Verb Norms (MPVN): A text analysis measure of mental state attribution. *Behavior Research Methods, 55*, 2501–2521. https://doi.org/10.3758/s13428-022-01911-7
Osborne, J. W. (1998). *Measuring metacognition: Validation of the assessment of cognition monitoring effectiveness* [Unpublished doctoral dissertation]. State University of New York at Buffalo https://www.proquest.com/docview/304471755/abstract/2E46A27B3DC4C1CPQ/1
Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*(71). https://doi.org/10.1136/bmj.n71
Pedone, R., & Semerari, A. (2023). Preliminary development and psychometric evaluation of the metacognition brief rating scale: An informant form of the metacognition self-assessment scale. *Clinical Neuropsychiatry, 20*(6), 511–522. https://doi.org/10.36131/cnfioritieditore20230606
Pedone, R., Semerari, A., Riccardi, I., Procacci, M., Nicolo, G., & Carcione, A. (2017). Development of a self-report measure of metacognition: The metacognition self-assessment scale (MSAS). Instrument description and factor structure. Clinical. *Neuropsychiatry, 14*(3), 185–194.
Pilgrim, D. (2014). Some implications of critical realism for mental health research. *Social Theory & Health, 12*(1), 1–21. https://doi.org/10.1057/sth.2013.17
Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*(3), 801–813. https://doi.org/10.1177/0013164493053003024
Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512
Quek, J., Melvin, G. A., Bennett, C., Gordon, M. S., Saeedi, N., & Newman, L. K. (2019). Mentalization in adolescents with borderline personality disorder: A comparison with healthy controls. *Journal of Personality Disorders, 33*(2), 145–163. https://doi.org/10.1521/pedi_2018_32_336
Quesque, F., Apperly, I., Baillargeon, R., Baron-Cohen, S., Becchio, C., Bekkering, H., … Brass, M. (2024). Defining key concepts for mental state attribution. *Communications Psychology, 2*(1), 1–5. https://doi.org/10.1038/s44271-024-00077-6
Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science, 15*(2), 384–396. https://doi.org/10.1177/1745691619896607

Rameckers, S. A., Verhoef, R. E., Grasman, R. P., Cox, W. R., van Emmerik, A. A., Engelmoer, I. M., & Arntz, A. (2021). Effectiveness of psychological treatments for borderline personality disorder and predictors of treatment outcomes: A multivariate multilevel meta-analysis of data from all design types. *Journal of Clinical Medicine, 10* (23), 5622. https://doi.org/10.3390/jcm10235622

Realo, A., Allik, J., Nõlvak, A., Valk, R., Ruus, T., Schmidt, M., & Eilola, T. (2003). Mind-reading ability: Beliefs and performance. *Journal of Research in Personality, 37*(5), 420–445. https://doi.org/10.1016/S0092-6566(03)00021-7

Ridenour, J. M., Lewis, K. C., Siefert, C. J., & Stein, M. B. (2022). Longitudinal stability of Social Cognition and Object Relations Scale—Global Rating Method dimensional ratings, score ranges and narrative 'blandness' in a clinical sample. *Clinical Psychology & Psychotherapy, 29*(4), 1447–1456. https://doi.org/10.1002/cpp.2729

Riedl, D., Kampling, H., Nolte, T., Lampe, A., Beutel, M. E., Brähler, E., & Kruse, J. (2023). Measuring impairments of mentalization with the 15-item Mentalization Questionnaire (MZQ) and Introducing the MZQ-6 short scale: Reliability, validity and norm values based on a representative sample of the German population. *Diagnostics, 13*(1), Article 1. https://doi.org/10.3390/diagnostics13010135

Rifkin-Zybutz, R. P., Moran, P., Nolte, T., Feigenbaum, J., King-Casas, B., Fonagy, P., … London Personality and Mood Disorder Consortium. (2021). Impaired mentalizing in depression and the effects of borderline personality disorder on this relationship. *Borderline Personality Disorder and Emotion Dysregulation, 8*(1), 15. https://doi.org/10.1186/s40479-021-00153-x

Rimes, K. A., & Chalder, T. (2010). The Beliefs about Emotions Scale: Validity, reliability and sensitivity to change. *Journal of Psychosomatic Research, 68*(3), 285–292. https://doi.org/10.1016/j.jpsychores.2009.09.014

Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science, 24*(1), 27–33. https://doi.org/10.1177/0956797612447819

Rumeo, C. (2022). *A Critical Review of Measures of Mentalization from Peter Fonagy's Conceptual Framework* [Unpublished master's thesis]. University of Waterloo https://uwspace.uwaterloo.ca/handle/10012/18544

Rutimann, D. D., & Meehan, K. B. (2012). Validity of a brief interview for assessing reflective function. *Journal of the American Psychoanalytic Association, 60*(3), 577–589. https://doi.org/10.1177/0003065112445616

Sandi-Urena, G. S. (2009). *Design and validation of a multimethod assessment of metacognition and study of the effectiveness of Metacognitive Interventions* [Unpublished doctoral dissertation]. Clemson University https://www.proquest.com/docview/230661741/abstract/6785A8802DDC434DPQ/1

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460–475. https://doi.org/10.1006/ceps.1994.1033

Schwarzer, N.-H., Nolte, T., Fonagy, P., & Gingelmaier, S. (2021). Mentalizing and emotion regulation: Evidence from a nonclinical sample. *International Forum of Psychoanalysis, 30*(1), 34–45. https://doi.org/10.1080/0803706X.2021.1873418

Semerari, A., Carcione, A., Dimaggio, G., Falcone, M., Nicolò, G., Procacci, M., & Alleva, G. (2003). How to evaluate metacognitive functioning in psychotherapy? The metacognition assessment scale and its applications. *Clinical Psychology & Psychotherapy, 10*(4), 238–261. https://doi.org/10.1002/cpp.362

Semerari, A., Carcione, A., Dimaggio, G., Nicolò, G., & Procacci, M. (2007). Understanding minds: Different functions and different disorders? The contribution of psychotherapy research. *Psychotherapy Research, 17*(1), 106–119. https://psycnet.apa.org/doi/10.1080/10503300500536953.

Semerari, A., Cucchi, M., Dimaggio, G., Cavadini, D., Carcione, A., Battelli, V., … D'Angerio, S., Ronchi, P., Maffei, C., & Smeraldi, E.. (2012). The development of the Metacognition Assessment Interview: Instrument description, factor structure and reliability in a non-clinical sample. *Psychiatry Research, 200*(2), 890–895. https://doi.org/10.1016/j.psychres.2012.07.015

Shill, M. A., & Lumley, M. A. (2002). The Psychological Mindedness Scale: Factor structure, convergent validity and gender in a non-psychiatric sample. *Psychology and Psychotherapy: Theory, Research and Practice, 75*(2), 131–150. https://doi.org/10.1348/147608302169607

Slade, A., Grienenberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning, attachment, and the transmission gap: A preliminary study. *Attachment & Human Development, 7*(3), 283–298. https://doi.org/10.1080/14616730500245880

Stefana, A., Jolić Marjanović, Z., & Dimitrijević, A. (2024). The brief version of the Mentalization Scale (MentS-12): Evidence-based assessment of mentalizing capacity. *Journal of Personality Assessment, 1–10*. https://doi.org/10.1080/00223891.2024.2326884

Swanson, H. L. (1990). Influence of metacognitive knowledge and aptitude on problem solving. *Journal of Educational Psychology, 82*(2), 306–314. https://doi.org/10.1037/0022-0663.82.2.306

Taubner, S., Kessler, H., Buchheim, A., Kächele, H., & Staun, L. (2011). The role of mentalization in the psychoanalytic treatment of chronic depression. *Psychiatry: Interpersonal and Biological Processes, 74*(1), 49–57. https://doi.org/10.1521/psyc.2011.74.1.49

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., … Straus, S. E. (2018). PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. *Annals of Internal Medicine, 169*(7), 467–473. https://doi.org/10.7326/M18-0850

de Waal, F. B. M. (2019). Fish, mirrors, and a gradualist perspective on self-awareness. *PLOS Biology, 17*(2), Article e3000112. https://doi.org/10.1371/journal.pbio.3000112

de Waal, F. B. M., & Ferrari, P. F. (2010). Towards a bottom-up perspective on animal and human cognition. *Trends in Cognitive Sciences, 14*(5), 201–207. https://doi.org/10.1016/j.tics.2010.03.003

Ward, A., Ramsay, R., Turnbull, S., Steele, M., Steele, H., & Treasure, J. (2001). Attachment in anorexia nervosa: A transgenerational perspective. *British Journal of Medical Psychology, 74*(4), 497–505. https://doi.org/10.1348/000711201161145

Warnell, K. R., & Redcay, E. (2019). Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition, 191*, Article 103997. https://doi.org/10.1016/j.cognition.2019.06.009

Westen, D. (1991). Social cognition and object relations. *Psychological Bulletin, 109*(3), 429–455. https://doi.org/10.1037/0033-2909.109.3.429

Whiten, A. (1991). *Natural Theories of Mind: Evolution, Development, and Simulation of Everyday Mindreading.* B. Blackwell.

Williams, C. H. J. (2015). Improving Access to Psychological Therapies (IAPT) and treatment outcomes: Epistemological assumptions and controversies. *Journal of Psychiatric and Mental Health Nursing, 22*(5), 344–351. https://doi.org/10.1111/jpm.12181

Wobber, V., Herrmann, E., Hare, B., Wrangham, R., & Tomasello, M. (2014). Differences in the early cognitive development of children and great apes. *Developmental Psychobiology, 56*(3), 547–573. https://doi.org/10.1002/dev.21125

World Health Organisation. (2022). *International statistical classification of diseases and related health problems* (11th ed.) https://icd.who.int/.

Wu, H., Fung, B. J., & Mobbs, D. (2022). Mentalizing during social interaction: The development and validation of the interactive mentalizing questionnaire. *Frontiers in Psychology, 12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.791835.