



# Testing Children's Mentalizing in Middle Childhood: Adopting the Child and Adolescent Reflective Functioning Scale with Clinical and Community Children

Fabiola Bizzi<sup>1</sup> · Simone Charpentier-Mora<sup>1</sup> · Donatella Cavanna<sup>1</sup> · Jessica L. Borelli<sup>2</sup> · Karin Ensink<sup>3</sup>

Accepted: 15 July 2021  
© The Author(s) 2021

## Abstract

Beginning with Ensink's seminal study (2015), the field entered a new era in which we were able to measure mentalizing in school-aged children. The goal of this work is to continue developing the state of the research within this tradition by exploring the psychometric properties of the Child Reflective Functioning Scale (CRFS) - a measure applied to the Child Attachment Interview and designed to assess RF during middle childhood - within both clinical and normative groups, and to examine if it differentiates between both groups. Participants were 159 Italian children (age range 8–12 years,  $M_{age} = 10.66$ ,  $SD = 1.83$ ; 57% males) divided into two groups: 71 children with emotional-behavioral problems ('clinical group') and 88 children without emotional-behavioral problems ('community group'). Demographic data and Verbal Comprehension Index were collected. A two-factor model of CRFS (self- and other-focused RF) was confirmed in both groups, revealing that these intrapersonal and interpersonal indicators of children's RF are important domains of mentalizing abilities in middle childhood. The results revealed adequate inter-rater reliability of the CRFS. Controlling for the effect of potentially confounding demographic variables, significant differences on CRFS scales between clinical and community groups were found. The clinical group showed lower levels of CRFS scores than normative group, but importantly, the self-focused RF score uniquely predicted clinical/community status. Taken together, the findings showed that the CRFS is a reliable and validity measure for assessing RF in middle childhood with clinical and normative groups, contributing important information to the scientific literature on mentalizing in middle childhood.

**Keywords** Child and Adolescent Reflective Functioning Scale · Mentalizing · Psychological assessment · Middle childhood · Clinical and community children

## Highlights

- The CRFS is a reliable and valid measure used to assess RF in middle childhood, discriminating between clinical and community children (children with and without emotional behavioral problems).
- Testing the psychometric proprieties of CRFS through the use of clinical and community groups, we found that the measure achieved adequate internal consistency, high inter-rater reliability, with a two-factor solution of CRFS (self- and other-focused RF).

---

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1007/s10826-021-02057-4>.

✉ Fabiola Bizzi  
fabiola.bizzi@unige.it

<sup>1</sup> Department of Educational Sciences (DISFOR), University of Genoa, Genoa, Italy

<sup>2</sup> Department of Psychology and Social Behavior, University of California, Irvine, CA, USA

<sup>3</sup> Department of School of Psychology, University of Laval, Quebec, QC, Canada

- In both the clinical and the normative samples, CRFS scores were associated with age, gender and verbal competence, while they were not associated with SES and family composition. Clinical group showed lower levels of CRFS scores of all three dimensions (general, self- and other-focused RF) than the normative group, and only self-focused RF predicted clinical/community status in middle childhood.
- New evidence regarding the similar factorial invariance of CRFS for clinical and community children suggest that the two intrapersonal and interpersonal components of children's RF are reliable dimensions of mentalizing within clinical and community children during middle childhood.

One of the most fundamentally human capacities is the tendency to look to others' minds as sources of information and meaning about the world. Mentalizing refers to the individual's ability to hold others' minds in mind (Fonagy et al., 2002; Fonagy & Target, 1997). This capacity allows individuals to perceive both the self and others in terms of mental states (e.g., needs, desires, feelings, beliefs, goals, thoughts, intentions, and motivations), thereby rendering the contents of others' minds into something that is understandable, predictable, and contains meaningful sources of information. Mentalizing skills are thought to originate in the context of attachment relationships (Bateman & Fonagy, 2004; Fonagy & Target, 2006): secure attachment within early child-caregiver relationships provides a necessary precondition for mentalizing to emerge as a developmentally-acquired capacity that is enacted both explicitly and implicitly (Fonagy et al., 2007; Shai & Belsky, 2017). Conversely, disruptions in attachment relationships have been associated with impairments in mentalizing, which in turn confer risk for a variety of psychological disorders, such as borderline personality disorder (Bateman & Fonagy, 2004), antisocial personality disorder (Bateman & Fonagy, 2008), eating disorders (Skårderud, 2007), and major depressive disorder (Luyten et al., 2012).

From the attachment perspective, the Reflective Functioning (RF; Fonagy et al., 1991) framework provides a methodology for assessing mentalizing, specifically the capacity to consider close relationships and the self in terms of mental states. In other words, RF is inherently a dynamic skill used within interpersonal relationships to make meaning of behavior and interactions. RF may be especially important in the context of challenging life and interpersonal circumstances (Ensink et al., 2015). Indeed, researchers have argued that RF is a protective factor in the context of life adversity, perhaps because it enables people to reflect upon their thoughts and feelings following negative life experiences, facilitating deeper processing of such adversity (Borelli et al., 2015; 2019; Ensink et al., 2015).

Considering that RF does not develop and solidify early but emerges over time according to normative developmental milestones and the particular characteristics and circumstances of the child (Fonagy & Target, 1997), it is important to have several methods of measuring RF along the lifespan. In this respect, well-validated measures to assess mentalizing exist in adulthood through the

application of an elaborate coding procedure to transcripts of the Adult Attachment Interviews (AAIs; George et al., 1985), the Parent Development Interview (PDI; Slade, 2005), and the Insightfulness Assessment Procedure (IAP; Oppenheim & Koren-Karie, 2013). Conversely, there are a paucity of well-validated measures to assess mentalizing in childhood and in adolescence. Indeed, only the Reflective Function Questionnaire -Youth (RFQ-Y; Ha et al., 2013), a self-report measure adapted from the adult version of the RFQ (Fonagy et al., 2016) exists in adolescence to measure child's mentalizing. This instrument assesses two types of impairments in RF, namely hyper-mentalization (described as mentalizing content without appropriate contextual or observable supportive data; Frith, 1994) and hypo-mentalization (described as mentalizing content completely opaque and accordingly unapproachable; Fonagy et al. 2016). In middle childhood, only the Friends and Family Attachment Interview (FFI; Steele et al., 2005) and the Child and Adolescent Reflective Functioning Scale (CRFS; Ensink et al., 2013), are suitable measures of RF.

Youth in middle childhood are actively developing cognitively, emotionally, and socially and their narratives reflect this dynamic process in which, in contrast to adults, the verbalizations of one's own and other's states of mind are not yet fully elaborated (Ensink et al., 2013). Building upon the increasing language capacities and emotional-cognitive changes in middle childhood (Carr, 2017), the CRFS is an age-specific RF measure capable of distinguishing between developmentally determined limitations in narrative abilities and mentalizing difficulties (Ensink et al., 2015). It is expected that when speaking about oneself and one's close relationships, children can describe specific incidents with reveal something about the self, their interpersonal interactions, and their affective reactions. Generating these narrative descriptions requires that children engage in the retrieval of specific episodic memories. These episodic or autobiographical memories are expected to provide a good indicator of the child's knowledge of mental states and of both intrapersonal and interpersonal thinking (Ensink et al., 2013).

The CRFS was adapted from the Adult Reflective Functioning Scale (Fonagy et al., 1998) and it was designed for use with the Child Attachment Interview (CAI; Shmueli-Goetz et al., 2008). It investigates specific subdomains of RF, including awareness of qualities of mental states, explicit

efforts to understand mental states underlying behavior, recognizing that mental states develop in the contexts of developmental, psychobiological, and social processes, and mental states in relation to the interview (Ensink et al., 2013). It permits the assessment of multiple dimensions of RF (Allen et al., 2008; Fonagy et al., 2002; Suchman et al., 2010): children provide self-descriptions as well as descriptions of their attachment relationships, which raters score to obtain indicators of mentalizing regarding self (i.e., the self-focused RF) as well as attachment figures (i.e., the other-focused RF); these scores are then combined to provide a global RF indicator (i.e., the global-RF). The original validation study of CRFS, based on an American sample, confirmed the reliability and validity of the CRFS coding system: the excellent intra-class correlation coefficients for global-RF, self-focused RF and other-focused RF; the discriminant validity; and the association between maternal RF (measured with the PDI) and child RF (measured with the CRFS) were all demonstrated through this initial study (Ensink et al., 2015).

Curiously, as illustrated in greater detail in the Appendix, beyond this preliminary study, no previous studies have been conducted on the CRFS to establish its psychometric properties, leaving several questions unanswered in the literature. To date, the CRFS has been most frequently used to assess RF with children exposed to abuse (Ensink et al., 2015, 2016, 2017; Tessier et al., 2016) and in community samples of school-aged children (Bizzi et al., 2020a; Borelli et al., 2018; Rosso & Airoldi, 2016). Some studies using the CRFS have focused on specific clinical groups, as children with disruptive behavior disorders, somatic symptom disorders (Bizzi et al., 2019, 2020b), adolescents with borderline traits (Sharp et al., 2020; Vanwoerden et al., 2019), and children with Type 1 diabetes (Costa-Cordella et al., 2021). And yet, somewhat surprisingly, no studies to date have examined RF using the CRFS using a combined clinical-community sample. In other words, no studies have tested whether RF scores on the CRFS discriminate between clinical and non-clinical samples. This constitutes a major gap in the literature in that we do not yet know whether RF scores on the CRFS are contribute clinically meaningful information.

Therefore, beginning with Ensink's seminal study (2015), and considering the growing body of evidence of the associations between child RF and psychopathology, the goal of this work is to continue developing the state of the research within this tradition by exploring the psychometric properties of the Child Reflective Functioning Scale (CRFS) using a different sample consisting of children with and without emotional-behavioral problems ('clinical' and 'normative' groups). Specifically, we aim: (1) to confirm the internal structure (two-factor solution) of CRFS as suggested by Ensink et al. (2015) study, by testing it on clinical and community groups; (2) to test measurement invariance

of CRFS (self- and other-focused RF) in both clinical and community groups; (3) to test the inter-rater reliability of two independent coders; (4) to establish the CRFS' independence from potentially confounding demographic features (age, gender, socioeconomic status, family composition) and verbal competence in both clinical and community groups; and (5) to test whether CRFS scores discriminate between clinical and community groups.

The first two aims highlight an open question in the research literature as to whether RF is a multidimensional construct (Ensink et al., 2015). Within the adult mentalizing literature, scholars argue that RF is multidimensional, comprised of both self- and other-focused components (Suchman et al., 2010), with each of these dimensions uniquely associated with relevant constructs. For instance, self-focused RF is associated with maternal contingent behavior in substance dependent mothers of toddlers (Suchman et al., 2010) while child-focused RF is associated with lower overcontrol in a community sample of mothers of school-aged children (Borelli et al., 2017). Although the scoring procedure assumes the CRFS to be two-dimensional, having both self- and other-focused components, this assumption has never been tested empirically in both clinical and community groups. Thus, we use confirmatory factor analysis (CFA) to confirm the two-factorial solution of CRFS and multi-group CFA to test group invariance. Further, the inter-rater reliability has been demonstrated for the general RF score (Ensink et al., 2015) but not for two dimensions of CRFS (self- and other-focused RF), and thus is another important area in need of investigation. In addition, we contribute to the validity studies reported above by testing the associations between RF scores on the CRFS and demographics and verbal competence within both a clinical and community samples. On the basis of extant literature (Ensink et al., 2015; Rosso & Airoldi, 2016; Vanwoerden et al., 2019), we hypothesize that the CRFS assesses mentalizing independently of socio-demographic variables and verbal comprehension. Last, based on the argument that impairments in RF confer risk for a variety of specific psychological disorders (Bizzi et al., 2019, 2020b; Costa-Cordella et al., 2021; Ensink et al., 2015, 2016, 2017; Sharp et al., 2020; Vanwoerden et al., 2019), we hypothesize that CRFS will be significantly lower within the clinical as compared to the community groups, and thus will discriminate between the two.

## Method

### Participants

The overall sample consisted of 159 Italian children from North-West Italy, uniformly distributed with respect to age

**Table 1** Demographics

	Group	N	Mean (SD) Frequency percentage	Statistics
Age	Clinical	71	10.7 (1.81)	$t_{(157)} = -0.36, p = 0.72$
	Community	88	10.6 (1.85)	Cohen's $d = 0.06$
Gender	Clinical	71	52% males	$X^2_{(1)} = 1.32, p = 0.25$
	Community	87	62% males	Cramer's $V = 0.09$
SES	Clinical	71	<sup>a</sup> 52.1% yearly household income between 15.000 and 36.000 €	$X^2_{(3)} = 14.7, p < 0.01^b$
	Community	86	<sup>c</sup> 42% yearly household income between 15.000 and 36.000 €	Cramer's $V = 0.31$
Family composition	Clinical	71	<sup>d</sup> 69% both parents' families	$X^2_{(2)} = 15.9, p < 0.001^e$
	Community	86	<sup>f</sup> 87.5% both parents' families	Cramer's $V = 0.32$
Verbal Comprehension Index (VCI)	Clinical	69	110.3 (14.90)	$t_{(148)} = 4.92, p < 0.001$
	Community	81	121.9 (14.00)	Cohen's $d = 0.81$

<sup>a</sup>9.9% yearly house income greater than 36.000 € and 25.3% yearly house income lower than 15.000 €

<sup>b</sup>post-hoc test: significant difference ( $p < 0.05$ ) between the two groups in the following categories: yearly house income lower than 15.000 € – yearly house income greater than 36.000 €

<sup>c</sup>31.8% yearly house income greater than 36.000 € and 10.2% yearly house income lower than 15.000 €

<sup>d</sup>8.5% lives in stepfamilies and 22.5% lives in single-parent families

<sup>e</sup>post-hoc test: significant difference ( $p < 0.05$ ) between the two groups in the following categories: single-parent families – two parent families

<sup>f</sup>7.95% lives in stepfamilies and 2.3% lives in single-parent families

and gender (age range 8–12 years,  $M_{age} = 10.66$ ,  $SD = 1.83$ ; 57% males), and their families approached without any material incentive offered. The overall sample was divided into two groups: 71 children with parent-reported emotional-behavioral problems ('clinical group') and 88 children without parent-reported emotional-behavioral problems ('community group'). The 'community group' was recruited from urban and rural schools of Genoa (Italy), while the children of 'clinical group' were consecutively admitted outpatients from the Child Neuropsychiatry and Psychology Units of the Gaslini Children's Hospital (Genoa, Italy) in which the assessment report constituted a diagnostic deepening in addition to that conducted by the psychologists of the hospital.

The present study adopted the following inclusion criteria. For both groups: (a) children were between 8 and 12 years of age (middle childhood), and (b) used Italian as their primary language. Within the clinical group, (c) children exceeded the clinical cut-off for the emotional-behavioral problems (total problems scored  $t \geq 65$ ) of the *Child Behavior Checklist 6–18 Version* (CBCL; Achenbach and Rescorla, 2001) while children in the community group had scores below this clinical cut-off (total problems score  $t \leq 65$ ). Accordingly, children in community group did not show emotional-behavioral problems in their total subscale scores ( $M = 49.94$ ,  $SD = 8.55$ ), whereas the clinical group's scores were higher (CBCL total problems score:  $M = 68.51$ ,  $SD = 3.15$ ).

Gender, age, SES (high > 36.000 €/y, moderate from 15000 to 36000 €/y, and low < 15.000 €/y), family composition (two parent families, single parent families, families with step-parents), and verbal competence of the overall sample are shown in Table 1. The groups did not differ in terms of gender and age, but significantly differed in SES, family composition and verbal competence. As a result, we included SES, family composition and verbal competence as covariates in subsequent analyses.

## Measures

The *Child Attachment Interview* (CAI; Shmueli-Goetz et al., 2008) is a semi-structured interview used to assess the child's attachment representations for both parents. The current CAI protocol contains 19 questions (CAI revised edition VIII; Shmueli-Goetz et al., 2008) and the CAI manual recommends that both video-recordings and verbatim transcripts are used by coders in assigning ratings. The interview begins with a brief introduction in which the examiner explains to the child what he or she will be asked: "This is an interview about you and your family. I will ask you some questions first about you and then about your relationship with your parents. For each question I will ask you to give me some examples. This interview is not a test and there are no right or wrong answers. I just want you to tell me some things about you and your family from your point of view. The interview will last about half an hour,

**Table 2** Child attachment interview's questions for child reflective functioning scale's coding

Child reflective functioning's scales		Child attachment interview's questions	
Global-RF	Self-focused RF	2A. Self-description	"Tell me three words that describe yourself, that is not what you look like, but what sort of person you are."
		2B. Self-description	
		2C. Self-description	
		7. Upset	
	Other-focused RF	3A. Mother relationship	"Can you tell me three words to describe your relationship with your mum, that is, what it's like to be with your Mum?"
		3B. Mother relationship	
		3C. Mother relationship	
		4. Mother argue	"What happens when your Mum gets cross with you or tells you off?"
		5A. Father relationship	"Can you tell me three words to describe your relationship with your Dad, what it's like to be with your Dad?"
		5B. Father relationship	
		5C. Father relationship	
		6. Father argue	"What happens when your Dad gets cross with you or tells you off?"
		13. Parents argue	"Do/did your parents sometimes argue?"

maybe a little longer". The questions in the CAI tap the child's self-representation, representations of his/her primary caregivers, times of conflict, distress, illness, hurt, separation, and loss. Coders use eleven 9-point scales, then placing children into one of four best-fitting attachment classifications (secure, insecure-dismissing, insecure-preoccupied, disorganized) on the basis of the distribution of the scale scores as well as a consideration of the child's non-verbal behavior. The CAI's psychometric properties have been demonstrated through a series of studies exploring its reliability and construct validity in both clinical and community samples across multiple cultures (e.g., Bizzi, 2019; Bizzi & Pace, 2020; Bizzi et al., 2018, 2021a, 2021b; Cavanna et al., 2018; Shmueli-Goetz et al., 2008; Venta et al., 2014). In the current study, the CAI's inter-rater reliability (ICC (3, k) = 0.946) was demonstrated excellent in all cases ( $N = 159$ ).

The *Child and Adolescent Reflective Functioning Scale* (CRFS; Ensink et al., 2013) is a measure of RF designed for youth between the ages of 7 to 17 that was designed to be applied to CAI transcripts. The CRFS was adapted from Adult Reflective Functioning Scale (Fonagy et al., 1998) for use with school-aged children, purporting to assess children's ability to mentalize regarding themselves and their attachment figures. The CRFS manual (Ensink et al., 2013) contains descriptions and examples of different levels and types of CRF. Children's narratives are coded on an 11-point scale (-1 to 9) descriptively anchored at six points in terms of the degree to which children's responses reveal a propensity to consider interpersonal interactions and personal reactions in mental state terms. The different levels of RF are here reported. -1: bizarre, disorganized response or

avoided mentalization; 0: absence of mentalization; 1: self-description in terms of behavior, non-mental characteristics; 2: descriptions without explicit reference to mental states; 3: some vague, basic but unelaborated references to mental states; 4: recognition that the experience of negative affect may elicit responses from others, which in turn can help to soothe or regulate the affect in various ways; 5: clear description of mental states showing a solid understanding; 6: clear but intentionally communicates of mental states; 7: understanding that different people may perceive a given behavior or situation differently often based on differing knowledge of the situation or false belief; 8: unusually nuanced understanding of reactions of self and other that also incorporates a sense of feelings and reactions changing over time; 9: sophisticated mentalization capacity (Ensink et al., 2015). Examples of the CAI questions analyzed for Child Reflective Functioning Scale's coding are reported in Table 2. To obtain a general indicator of children's RF (CRF-G), we calculated the mean RF of all the coded responses. To obtain an indicator of self-focused RF (CRF-S), the mean RF for the four items eliciting self-descriptions and the child's reactions in response to upsetting events was used. Furthermore, an indicator of other-focused RF (CRF-O) was calculated based on the mean RF on the nine questions regarding the child's relationships with their parents and a description of parents' reactions when get upset or when they argue. The CRFS' psychometric properties have been demonstrated by Ensink's studies (Ensink, 2004; Ensink et al., 2015). In this study, the scale alpha was 0.94, and item-total correlations ranged from 0.57 to 0.79, confirming that the CRFS could be used as a good indicator of RF.

The *Wechsler Intelligence Scale for Children-IV* (WISC-IV; Wechsler, 2003) is a well-established and validated measure of the intellectual abilities of children aged 6–16. In this study, we employed the Verbal Comprehension Index (VCI) to measure the verbal competence, which is the sum of weighted points in similarities, vocabulary and comprehension subtests. In the similarities subtest, the child is required to explain in ways two words are similar that refer to common objects or concepts (e.g. “In what ways are apples and bananas alike?”); for the vocabulary subtest, the child verbally defines a set of words (e.g. “What is a hat?”); and for the comprehension subtest, the child responds verbally to questions about solutions to everyday problems expressing an understanding of rules and concepts (e.g. “What are the reasons for turning off the lights when no one uses them?”). In the Italian version (Orsini et al., 2012), Cronbach’s  $\alpha$  is 0.96 for the VCI, ranging 0.69 (comprehension) to 0.94 (vocabulary). In this study, Cronbach’s  $\alpha$  was 0.80 for the VCI.

The *Child Behavior Checklist* (CBCL 6–18; Achenbach & Rescorla 2001) is a parent-report questionnaire used to assess emotional and behavioral problems in children aged 8–16 years. It is comprised of 112 items; each item is scored on a 3-point scale, ranging from 0 to 2. The CBCL 6–18 provides a score for Total Problems and three main indexed scores: Internalizing problems (including anxious/depressed, withdrawn/depressed and somatic complaints), Externalizing problems (including rule-breaking behavior and aggressive behavior), and Other problems (including social problems, thought problems, and attention problems). The CBCL has good psychometric properties (Achenbach & Rescorla, 2001), and the Italian version was validated by Frigerio et al. (2004), displaying good validity and reliability with good internal consistency, with cut-offs comparable to the American population, and interrater agreement similar to the values reported by Achenbach. In our study, Cronbach’s  $\alpha$  was 0.92 for the Total Problems score.

## Procedure

The study was approved by the Ethics Committee of the Gaslini Children’s Hospital (Genoa, Italy) for the clinical group and the Ethics Committee of the Department of Educational Sciences (University of Genoa, Italy) for the community group, according with the Declaration of Helsinki. Data were collected in the last two years. Parents were provided with a written document describing the procedures and purpose of the study; all parents and their children provided informed consent and assent, respectively, and were informed that they could decline to participate in any part of the study. Only two participants of the community group elected not to participate because they were too busy

at the time of the data collection. Therefore, the final sample ( $N = 159$ ) represented 98% of those eligible. The assessment (individual sessions lasting around 60 min) was conducted (in a private room at the hospital for the clinical group and at participants’ homes for the community group) by a psychologist/researcher who had previously been trained in the administration of the CAI (directly by one of the authors for two consecutive days) and the WISC-IV. In a separate room, parents were asked to complete a socio-demographic questionnaire and the CBCL (for clinical group it was previously administered by the psychologist of the hospital).

In this study, CRFS coding was carried out by the first two authors; they were supervised and/or received consultation by the developer of the CRFS (the last author). These coders were naïve to all participant characteristics and did not administer the interviews.

The analyses were carried out with Statistical Package for the Social Sciences (SPSS) and JASP, an open-source statistics program that allows both classical and Bayesian analyses including CFA.

## Statistical Data Analyses

A priori statistical power analysis was conducted using G\*Power software (Faul et al., 2009). Considering differences between two independent means (two groups), effect size = 0.5, significance level ( $\alpha$ ) = 0.05, or power ( $1 - \beta$ ) = 0.90, 70 observations for each group were needed to test the psychometric properties of CRFS. Then, our aims were assessed following five steps. First, we examined the supposed two-factorial structure of CRFS with confirmatory factor analysis (CFA). Second, we used multi-group CFA to investigate the factorial invariance of CRFS in clinical and community groups testing for configural, metric, scalar and strict measurement invariance. We used fit indices to test model fit using cutoff values generally indicating a good fit (Hu & Bentler, 1999): chi-square/ $df$  ratio ( $\leq 2$  acceptable), comparative fit index (CFI over 0.90 acceptable), Tucker-Lewis fit index (TLI over 0.90 acceptable), root mean square error of approximation (RMSEA between 0.05 and 0.08 acceptable) and standardized root mean squared residual (SRMR < 0.08 acceptable). Comparison between different factor solutions has been made with chi-square difference test and a drop in CFI greater than 0.005 (Chen, 2007). Third, reliability of the CRFS was estimated with Cronbach’s alpha coefficient and mean inter-item correlation. Fourth, analyses of variance,  $t$ -tests, Pearson correlations were used to test the association between CRFS, demographic features and verbal competence. Fifth, analyses of covariance and logistic regression analyses were used to evaluate the ability of the CRFS to discriminate between clinical and community groups.

## Results

### Confirming the Internal Structure and Measurement Invariance of CRFS

Initially, we explored the items and score characteristics of the CRFS (see Table 3), demonstrating their adequate distributional characteristics.

In order to assess Aim One, a confirmatory factor analysis (CFA) was conducted on clinical and community groups combined to explore the two-factor solution (self- and other-focused RF) initially proposed in the realm of parental RF by Suchman et al. (2010) and introduced in the realm of children's mentalizing by Ensink et al. (2015). No missing values were reported. Table 4 presents the different models: First, we tested the proposed two-factor model (Model 1), which yielded a satisfactory factor solution. However, Model 1 significantly improved the fit by allowing error correlation between a few questions (2b – 2c; 6 - 7) with similar wording/content due to modification Indices values higher than 10 (Byrne 2016), resulting in Model 1a. All the standardized coefficients for the two-factor model were statistically significant ( $p < 0.001$ ) (see Table 6). Secondly, we tested a one-factor model (Model 2), which showed a worse fit than the previous two-factor model ( $X^2(3) = 40.2$ ,  $p < 0.001$ ,  $\Delta CFI = 0.04$ ).

In order to assess Aim Two, measurement invariance of model 1 between clinical ( $N = 71$ ) and community ( $N = 88$ ) groups was tested with a multi-group CFA elaborated with

increasingly restrictive models (see Table 5). First, we tested configural invariance (Model 3); note that here, obtaining a good model fit indicates that the two-factor model is associated with the same items across both groups. Secondly, we tested metric invariance (Model 4); here we obtained a good model fit, except SRMR value, but worse fit than the model 3 ( $X^2(13) = 23.35$ ,  $p < 0.05$ ,  $\Delta CFI = 0.01$ ) due to factor loadings that were not equal across groups. Moreover, all items had substantial and statistically significant loadings on their respective factors in both groups. Third, we tested scalar invariance (Model 5), obtaining a poor model fit that was significantly worse than Model 4 ( $X^2(11) = 29.24(11)$ ,  $p < 0.001$ ,  $\Delta CFI = 0.02$ ) due to different group means for clinical and community groups. In total, our findings confirm the two-factorial solution of CRFS proposed by Ensink et al. (2015) and indicate factorial invariance across both clinical and community groups.

Internal consistency reliability in the total sample for the subscales of the two-factor solution was acceptable both for self-focused RF ( $\alpha = 0.75$ , mean inter-item correlation = 0.44) and other-focused RF ( $\alpha = 0.90$ , mean inter-item correlation = 0.51). The same result occurred both in clinical (self-focused RF:  $\alpha = 0.67$ , mean inter-item correlation = 0.34; other-focused RF:  $\alpha = 0.89$ , mean inter-item correlation = 0.48) and community group (self-focused RF:  $\alpha = 0.66$ , mean inter-item correlation = 0.34; other-focused RF:  $\alpha = 0.85$ , mean inter-item correlation = 0.40). Internal consistency could not be improved by the deletion of any item. Thus, Aims One and Two were supported in

**Table 3** Items and scores characteristics of CRFS

Item	Descriptive statistics					
	<i>n</i>	<i>M</i>	<i>SD</i>	Range	Skew	Kurtosis
2A. Self-description	159	2.09	1.27	-1 to 6	0.14	-0.20
2B. Self-description	159	2.09	1.30	-1 to 6	0.23	-0.28
2C. Self-description	159	2.21	1.35	-1 to 6	0.35	-0.38
3A. Mother relationship	159	2.35	1.18	-1 to 5	-0.29	-0.09
3B. Mother relationship	159	2.38	1.36	-1 to 6	-0.21	-0.25
3C. Mother relationship	159	2.37	1.47	-1 to 6	-0.05	-0.45
4. Mother argue	159	3.32	1.50	0 to 7	-0.23	0.05
5A. Father relationship	159	2.29	1.35	-1 to 7	0.26	0.37
5B. Father relationship	159	2.32	1.46	-1 to 6	0.10	-0.55
5C. Father relationship	159	2.22	1.59	-1 to 6	0.23	-0.51
6. Father argue	159	2.92	1.77	-1 to 8	-0.38	-0.24
7. Upset	159	1.91	1.63	-1 to 7	0.47	-0.43
13. Parents argue	159	2.76	1.52	-1 to 7	-0.29	0.03
Score						
Self-focused RF	159	2.07	1.06	-1 to 5	0.22	0.29
Other-Focused RF	159	2.55	1.11	-0.33 to 5.22	-0.24	0.03
Global-RF	159	2.41	1.02	-0.54 to 5.08	-0.12	0.14

**Table 4** Fit statistics CFA models

	$X^2$	$df$	$p$	$X^2/df$	RMSEA	TLI	CFI	SRMR
Model 1	115	64	<0.001	1.79	0.07	0.93	0.94	0.04
Model 1a	92.8	62	0.01	1.49	0.05	0.96	0.97	0.04
Model 2	133	65	<0.001	2.04	0.08	0.91	0.93	0.05
Model 3	169.73	124	0.004	1.36	0.06	0.92	0.94	0.06
Model 4	193.08	137	<0.001	1.40	0.07	0.91	0.92	0.12
Model 5	222.32	148	<0.001	1.50	0.07	0.89	0.90	0.13

Model 1 = the baseline two-factors model; Model 1a = the final two-factors model adding error covariances between items 2b – 2c, 6 – 7; Model 2 = the one-factor model; Model 3 = configural measurement invariance; Model 4 = metric measurement invariance; Model 5 = scalar measurement invariance

**Table 5** Factor loadings and factor covariances of two-factors model

Factor	Item	Factor loadings (SE)	
		Community group	Clinical group
Self-focused RF	CRFS_2a	0.53 (0.10)	0.53 (0.10)
	CRFS_2b	0.50 (0.10)	0.54 (0.10)
	CRFS_2c	0.53 (0.10)	0.59 (0.10)
	CRFS_7	0.53 (0.12)	0.64 (0.12)
Other-focused RF	CRFS_3a	0.74 (0.08)	0.63 (0.08)
	CRFS_3b	0.77 (0.09)	0.73 (0.09)
	CRFS_3c	0.65 (0.10)	0.66 (0.10)
	CRFS_4	0.57 (0.10)	0.50 (0.10)
	CRFS_5a	0.68 (0.09)	0.66 (0.09)
	CRFS_5b	0.81 (0.10)	0.75 (0.10)
	CRFS_5c	0.70 (0.11)	0.68 (0.11)
	CRFS_6	0.52 (0.12)	0.50 (0.12)
	CRFS_13	0.66 (0.10)	0.64 (0.10)
Factor 1-Factor 2		Factor covariance (SE)	
		Community group	Clinical group
		0.94 (0.07)	0.82 (0.09)

Standardized parameters are used

All coefficients are statistically significant ( $p < 0.001$ )

confirming internal structure, measurement invariance and scales reliability of CRFS.

### Evaluating Inter-Rater Reliability of CRFS

In order to assess Aim Three, we tested inter-rater reliability of CRFS scores (self-focused RF and other-focused RF), and their composite indicator (global-RF) using two independent coders (first and second author, supervised by the last author) that were blind in terms of clinical status of participants. In line with Ensink's study (2015), inter-rater reliability (self-focused RF: ICC (3, k) = 0.98; other-focused RF: ICC (3, k) = 0.98; global-RF: ICC (3, k) = 0.98) was calculated on 30% of interviews. This value is indicated as "excellent" inter-rater reliability for all RF dimensions according to Cicchetti (1994). In sum, Aim Three was confirmed demonstrating a proper inter-rater reliability.

### Association of CRFS with Demographic Features and Verbal Competence

In order to assess Aim Four, we tested the relations of RF with demographic features and verbal competence through  $r$  Pearson correlations,  $t$  tests for independent samples and analyses of variance. Firstly, we tested the association between RF and age in the total sample. Here we found a positive association with all RF scales (self-focused RF:  $r = 0.20$ ,  $p < 0.05$ ; other-focused RF:  $r = 0.25$ ,  $p < 0.01$ ; global-RF:  $r = 0.25$ ,  $p < 0.01$ ). Specifically, self-focused RF ( $r = 0.34$ ,  $p < 0.01$ ), other-focused RF ( $r = 0.34$ ,  $p < 0.01$ ) and global-RF ( $r = 0.37$ ,  $p < 0.01$ ) were positively associated with age in the community group. Only other-focused RF ( $r = 0.24$ ,  $p < 0.05$ ) was positively associated with age in the clinical group, whilst no associations were found with self-focused RF ( $r = 0.10$ ,  $p > 0.05$ ) and global-RF ( $r = 0.22$ ,  $p > 0.05$ ).



**Table 6** Relation between CRF and gender in clinical and community groups

	Groups	Gender	Mean (SD)	Statistics
Self-focused RF	Clinical	Male	1.37 (0.92)	$t_{(69)} = -0.91$ , Cohen's $d = 0.22$
		Female	1.56 (0.76)	
	Community	Male	2.35 (0.88)	$t_{(85)} = -2.47^*$ , Cohen's $d = 0.53$
		Female	2.83 (0.94)	
	Total sample	Male	1.87 (1.02)	$t_{(156)} = -2.72^{**}$ , Cohen's $d = 0.44$
		Female	2.33 (1.07)	
Other-focused RF	Clinical	Male	1.72 (1.15)	$t_{(69)} = -2.54^*$ , Cohen's $d = 0.62$
		Female	2.39 (0.96)	
	Community	Male	2.83 (0.86)	$t_{(85)} = -2.09^*$ , Cohen's $d = 0.45$
		Female	3.21 (0.84)	
	Total sample	Male	2.29 (1.15)	$t_{(154.25)} = -3.55^{**}$ , Cohen's $d = 0.56$
		Female	2.89 (0.97)	
Global-RF	Clinical	Male	1.64 (0.99)	$t_{(69)} = -2.10^*$ , Cohen's $d = 0.51$
		Female	2.12 (0.80)	
	Community	Male	2.69 (0.82)	$t_{(85)} = -2.41^*$ , Cohen's $d = 0.52$
		Female	3.10 (0.77)	
	Total sample	Male	2.18 (1.04)	$t_{(156)} = -3.35^{**}$ , Cohen's $d = 0.54$
		Female	2.71 (0.91)	

\* $p < 0.05$ . \*\* $p < 0.01$ . \*\*\* $p < 0.001$

Second, we tested the association between RF and verbal competence in total sample, finding a positive association with all RF scales (self-focused RF:  $r = 0.35$ ,  $p < 0.001$ ; other-focused RF:  $r = 0.41$ ,  $p < 0.001$ ; global-RF:  $r = 0.41$ ,  $p < 0.001$ ). Specifically, self-focused RF ( $r = 0.27$ ,  $p < 0.05$ ), other-focused RF ( $r = 0.38$ ,  $p < 0.01$ ) and global-RF ( $r = 0.37$ ,  $p < 0.01$ ) were positively associated with verbal competence in the clinical group. On the contrary, there was no association between RF and verbal competence in the community group (self-focused RF:  $r = 0.11$ ,  $p > 0.05$ ; other-focused RF:  $r = 0.15$ ,  $p > 0.05$ ; global-RF:  $r = 0.16$ ,  $p > 0.05$ ).

Third, we tested the association between RF and gender in total sample, finding a significant association with all RF scales. Specifically, gender was associated with all RF scores in the community group, such that females had higher scores on other-focused and global-RF in the clinical group (see Table 6). Fourth, we tested the association between RF and SES in total sample, finding a significant association with all RF scales, such that a yearly house family income greater than 36,000 € was associated with greater scores in RF. Interestingly, no association was found between RF and SES in clinical or community groups either alone (see Table 7).

Finally, we tested the association between RF and family composition, finding no significant associations both in total sample (self-focused RF:  $F_{(2, 154)} = 1.75$ ,  $p > 0.05$ ,  $\eta^2 = 0.02$ ; other-focused RF:  $F_{(2, 154)} = 1.67$ ,  $p > 0.05$ ,  $\eta^2 = 0.02$ ; global-RF:  $F_{(2, 154)} = 1.88$ ,  $p > 0.05$ ,  $\eta^2 = 0.02$ ) or in

clinical (self-focused RF:  $F_{(2, 68)} = 0.33$ ,  $p > 0.05$ ,  $\eta^2 = 0.01$ ; other-focused RF:  $F_{(2, 68)} = 0.16$ ,  $p > 0.05$ ,  $\eta^2 = 0.00$ ; global-RF:  $F_{(2, 68)} = 0.12$ ,  $p > 0.05$ ,  $\eta^2 = 0.00$ ) and community group either alone (self-focused RF:  $F_{(2, 83)} = 0.09$ ,  $p > 0.05$ ,  $\eta^2 = 0.00$ ; other-focused RF:  $F_{(2, 83)} = 0.10$ ,  $p > 0.05$ ,  $\eta^2 = 0.00$ ; global-RF:  $F_{(2, 83)} = 0.12$ ,  $p > 0.05$ ,  $\eta^2 = 0.00$ ).

### Discriminating Clinical and Community Groups

In order to assess Aim Five, we tested if RF differentiated between clinical and community groups, controlling for SES, family composition and verbal competence. Self-focused RF ( $F_{(18, 129)} = 33.15$ ,  $p < 0.001$ ,  $\eta^2 = 0.19$ ), other-focused RF ( $F_{(18, 129)} = 18.79$ ,  $p < 0.001$ ,  $\eta^2 = 0.12$ ) and global-RF ( $F_{(18, 129)} = 27.52$ ,  $p < 0.001$ ,  $\eta^2 = 0.16$ ) differentiated between clinical and community groups (see Table 8 for mean values), in that the community group had significantly higher RF of all three types.

Next, we conducted logistic regression analyses to assess whether CRFS capacity (*i.e.*, self- and other-focused RF) discriminated between the two groups. Since analyses showed that SES, family composition, and verbal competence differed between the two groups, these variables were entered into the regression models as control variables. In the subsequent analyses, for each additional unit of self-focused RF, the odds of being categorized as a member of the clinical group decreased by 72% ( $B$  (SE) =  $-1.28$  (0.38), OR = 0.28, 95% CI = 0.13 – 0.56,  $p < 0.001$ ).

**Table 7** Relation between CRF and socioeconomic status in clinical and community groups

	Groups	Socioeconomic status	Mean (SD)	Statistics	
Self-focused RF	Clinical	1	1.30 (1.12)	$F_{(3, 67)} = 0.72, \eta^2 = 0.03$	
		2	1.45 (0.79)		
		3	1.86 (0.73)		
	Community	1	2.47 (0.59)		$F_{(3, 82)} = 1.80, \eta^2 = 0.06$
		2	2.37 (0.99)		
		3	2.90 (0.77)		
	Total sample	1	1.69 (1.11)		$F_{(3, 153)} = 6.28^{***}, \eta^2 = 0.11^a$
		2	1.91 (1.00)		
		3	2.69 (0.87)		
Other-focused RF	Clinical	1	1.70 (1.27)	$F_{(3, 67)} = .87, \eta^2 = 0.04$	
		2	2.02 (1.05)		
		3	2.50 (1.38)		
	Community	1	2.94 (0.55)		$F_{(3, 82)} = 2.11, \eta^2 = 0.07$
		2	2.79 (0.97)		
		3	3.32 (0.82)		
	Total sample	1	2.12 (1.22)		$F_{(3, 153)} = 5.70^{***}, \eta^2 = 0.10^a$
		2	2.40 (1.07)		
		3	3.16 (0.99)		
Global-RF	Clinical	1	1.58 (1.14)	$F_{(3, 67)} = 0.97, \eta^2 = 0.04$	
		2	1.87 (0.87)		
		3	2.28 (1.02)		
	Community	1	2.80 (0.54)		$F_{(3, 82)} = 2.31, \eta^2 = 0.08$
		2	2.66 (0.89)		
		3	3.19 (0.73)		
	Total sample	1	1.98 (1.13)		$F_{(3, 153)} = 6.69^{***}, \eta^2 = 0.12^a$
		2	2.27 (0.96)		
		3	3.00 (0.86)		

Single-parent families—two parent families

1 = yearly house income lower than 15.000 €; 2 = yearly household income between 15.000 and 36.000 €; 3 = yearly house income greater than 36.000 €

\* $p < 0.05$ . \*\* $p < 0.01$ . \*\*\* $p < 0.001$

<sup>a</sup>post-hoc test: significant differences ( $p < 0.05$ ) between groups in the following categories: yearly house income lower than 15.000 € – yearly house income greater than 36.000 €; yearly household income between 15.000 and 36.000 € = yearly house income greater than 36.000 €

**Table 8** Means and standard deviations for CRFS scores as a function of clinical status

Group	n	Mean	SD
Self-focused RF			
Clinical	71	1.44	0.86
Community	88	2.57	0.93
Other-focused RF			
Clinical	71	1.98	1.13
Community	88	3.01	0.86
Global-RF			
Clinical	71	1.83	0.95
Community	88	2.88	0.81

F statistics provided in text

Other-focused RF, however, did not predict clinical/community group membership ( $B$  (SE) =  $-0.12$  (0.30), OR = 0.89, 95% CI = 0.48, 1.61,  $p = 0.70$ ).

## Discussion

Over the past two decades, RF has emerged as an important transdiagnostic marker of psychological health, with significant ties to attachment security, emotion regulation, and clinical status (Katznelson, 2014). Although RF studies have mainly focused on parental and adult RF, children's RF is often neglected in the literature. Consequently, valid and reliable measures of RF for school-aged children have

been slow to emerge, representing a slowing progress in the field. The CRFS was developed to fill this need and researchers rushed to use the measure with children and adolescents with specific problems (Bizzi et al., 2019, 2020b; Costa-Cordella et al., 2021; Ensink et al., 2015, 2016, 2017; Sharp et al., 2020; Vanwoerden et al., 2019). However, a particular focus on the psychometric proprieties of CRFS was lacking, particularly when it comes to clinical samples. In this sense, our study aims to fill this gap, testing several psychometric proprieties of CRFS and considering children with and without emotional-behavioral problems.

Consistent with the Ensink et al. (2015) seminal study, our findings report adequate internal consistency of the CRFS as well as high inter-rater reliability. Additionally, these findings confirm the two-factor solution of CRFS (self- and other-focused RF) suggested, but not verified by Ensink's preliminary study. Moreover, this is the first study to add new evidence regarding the factorial invariance of CRFS, i.e., the factor structure (configural invariance) of the measure is similar for clinical and community children, suggesting that the two intrapersonal and interpersonal components of children's RF comprise internally consistent, reliable dimensions of mentalizing within clinical and community children during the middle childhood. This work constitutes an important contribution to the literature, adding to the body of work suggesting that RF can be conceptualized in terms of different components. Such a conceptualization, initially offered within the parental RF literature by Suchman et al. (2010), has advanced the field considerably, helping refine predictions.

Further, the analyses on demographic features and verbal competence found that CRFS scores were associated in both groups with age, gender and verbal competence, while they were independent of SES and family composition. More specifically, all CRFS scales were associated with age within the community group, while only other-focused RF was associated with age in the clinical group. Associations with children's age supported the idea that RF competencies are developmental acquisitions related to the increase of the child's meta-cognitive skills, as has previously been suggested by prior work (Vanwoerden et al., 2019). Indeed, as children reach middle childhood, their social network grows and includes more interpersonal experiences with individuals outside their family environment, which may act to further enrich social-cognitive abilities and consequently amplify their ability to reflect on their own minds and others' minds in a relational context (Borelli et al., 2016). Similarly, our study revealed gender differences; specifically, girls within the community group had higher RF than boys in all scales. However, in the clinical group, girls exhibited higher RF than boys only in other-focused and global-RF. This finding is consistent with the idea that girls

in this age-range are more able to understand thoughts and emotions in others than boys (Bosacki & Astington, 1999). In addition, this mentalizing profile fits with idea that females exhibit better social and cognitive functioning, have better insight into their own mental states and they are generally more empathic than males (Baron-Cohen, 2003). Further, although mental state talk becomes a specific competence that across development departs from general language skills and contextual characteristics (Hughes et al., 2006), we found an association between verbal competence and CRFS within the clinical group. This supports the idea that language plays a crucial role in the development of mentalizing due to the fact that language is typically the tool used to represent and communicate the mental states (Vanwoerden et al., 2019). However, this finding leads us to argue that future research should aim to replicate these findings and collect verbal competence measures alongside the CRFS to explore the potential impact that verbal competence may have on child's mentalizing in different populations.

Controlling for confounding demographic variables, our findings provide evidence that the CRFS is a reliable measure in discriminating between clinical and community children. The differences in RF scores between the two groups may suggest that the development of interpersonal problems, social information processing deficit isolation or aggression problems are due to difficulties to perceive both the self and others in terms of mental states (Midgley et al., 2017). In other words, this pattern suggests that poor RF may lead to emotional behavior problems, and mentalizing deficits may constitute a transdiagnostic vulnerability factor for child psychopathology (Ensink et al., 2016). This insight could be of great use to clinicians in their case conceptualizations, as mentalizing may be an important treatment target. However, further investigation in this area are needed.

In addition, these findings underscore the importance of conceptualizing RF as a multidimensional construct (Allen et al., 2008; Borelli et al., 2017; Fonagy et al., 2002; Suchman et al., 2010). Furthermore, the predictive power of only self-focused RF suggests that in middle childhood, the role of RF regarding others' mental states may not be mature enough to play a central role in child psychopathology. In other words, this suggests that emotional development become more apparent during adolescence and not earlier when social and peer relationships are more complex (Blakemore, 2017). Importantly, these findings are consistent with other work – in another study of RF in middle childhood (measured using an interview following a stressor) only self-focused RF was related to maternal overcontrol (Borelli et al., 2017).

Considering the paucity of mentalizing measures in school-aged children, our findings are of paramount

importance in that they permit the field to expand the previous studies already focused on the role of mentalizing in borderline personality pathology development during adolescence (Sharp et al., 2020; Vanwoerden et al., 2019) as well as in specific clinical conditions (Bizzi et al., 2019; 2020b; Costa-Cordella et al., 2021) or abuse contexts (Ensink et al., 2015, 2016, 2017) in middle childhood. The capacity of CRFS to distinguish clinical from community children in middle childhood allows the clinician and researcher to have a measure useful in the assessment process as well as in the design of interventions for populations at risk and in the evaluation of child psychotherapy. Beginning with the growing interest regarding the role of RF in the process of psychological treatments for children and their families (Midgley et al., 2017), the development of psychometrically-sound measures like the CRFS will help to advance understanding of mentalizing processes as potential risk and protective factors included in increasingly complex etiopathogenetic models. In addition, the use of the CRFS requires relatively less investment in training, transcribing and coding the interview compared to what is needed to rate attachment. It is likely to present a worthwhile investment for clinicians and researchers given the lack of alternative measures. Furthermore, many child researchers have already been trained to use the CAI to assess attachment in school-aged children and could use the CRFS to examine the specific contribution of mentalization capacities (Ensink et al., 2015).

Despite the notable strengths of this study, there are several limitations that should be considered before making conclusions based on these findings. The sample size is relatively small, although adequate for testing the hypotheses of this study; the status of clinical or normative group is based only on the CBCL scores, therefore a multi-method assessment including teacher report and clinical interviews could be important to include in future research. The focus on a heterogeneous clinical group is another limitation of this study; studies including specific clinical samples (e.g., a sample with children with major depressive disorder only) will be an important complement to the current study in order to elaborate upon the relationships between mentalizing capacities and specific psychological difficulties. A longitudinal perspective (test-retest analysis) is lacking; research should investigate the stability of this measure to enforce the psychometric proprieties of this tool. Nonetheless, this study provides mounting evidence that supports the notion that the CRFS is a valid method to assess mentalizing during the middle childhood in clinical and community samples.

**Author Contributions** F.B. contributed to the study conception and design. Material preparation, data collection and analysis were performed by F.B. and S.C. The first draft of the manuscript was written

by F.B. and S.C. and J.L.B. commented on previous versions of the manuscript. All authors read and approved the final manuscript. English has been revised by the author of native English (J.L.B.).

**Funding** Open access funding provided by Università degli Studi di Genova within the CRUI-CARE Agreement.

## Compliance with Ethical Standards

**Conflict of Interest** The authors declare no competing interests.

**Ethical Approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee of the Gaslini Children's Hospital (Genoa, Italy) for the clinical group and the Ethics Committee of the Department of Educational Sciences (University of Genoa, Italy) for the community group, according with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed Consent** Informed consent was obtained from all individual participants included in the study. Parents completed consent forms, and children completed assent forms. The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Achenbach, T., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont Research Center for Children Youth & Families.
- Allen, J. G., Fonagy, P., & Bateman, A. W. (2008). *Mentalising in clinical practice*. Washington, DC: American Psychiatric Publishing.
- Baron-Cohen, S. (2003). *The essential difference: men, women and the extreme male brain*. London: Penguin Book Ltd.
- Bateman, A., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder* (pp. 529–532). Oxford: Oxford University Press.
- Bateman, A., & Fonagy, P. (2008). Comorbid antisocial and borderline personality disorders: mentalization-based treatment. *Journal of Clinical Psychology*, 64(2), 181–194. <https://doi.org/10.1002/jclp.20451> PMID:18186112.
- Bizzi, F. (2019). The quality of attachment in children with somatic symptom disorders and related disorders. *Psicologia Clinica dello Sviluppo*, 28(1), 139–152. <https://doi.org/10.1449/93137>.

- Bizzi, F., Charpentier-Mora, S., Ensink, K., Cavanna, D., & Borelli, J. (2020a). Does children's mentalizing mediate the role of attachment and psychological maladjustment in middle childhood? *Journal of Child and Family Studies*, 29(7), 1793–1803. <https://doi.org/10.1007/s10826-020-01701-9>.
- Bizzi, F., Della Vedova, A. M., Prandi, E., Cavanna, D., & Manfredi, P. (2021a). Attachment representations to parents and emotional-behavioral problems: a comparison between children with type 1 diabetes mellitus and healthy children in middle childhood. *Clinical Child Psychology and Psychiatry*, 26(2), 393–405. <https://doi.org/10.1177/1359104520987871>.
- Bizzi, F., Ensink, K., Borelli, J., Charpentier-Mora, S., & Cavanna, D. (2019). Attachment and reflective functioning in children with somatic symptom disorders and disruptive behavior disorders. *European Child and Adolescent Psychiatry*, 28(5), 705–717. <https://doi.org/10.1007/s00787-018-1238-5>.
- Bizzi, F., & Pace, C. S. (2020). Attachment representations in children with Disruptive Behavior Disorders: a special focus on insecurity in middle childhood. *Clinical Child Psychology and Psychiatry*, 25(4), 833–846. <https://doi.org/10.1177/1359104520918637>.
- Bizzi, F., Riva, A., Charpentier Mora, S., Cavanna, D., & Nacinovich, R. (2020b). Attachment representation to caregivers and mentalizing ability in early adolescents with primary headache. *Clinical Psychology & Psychotherapy*, 1–10. <https://doi.org/10.1002/cpp.2540>.
- Bizzi, F., Shmueli-Goetz, Y., Castellano, R., & Cavanna, D. (2018). A multi assessment approach to attachment in middle childhood and early adolescence in two clinical groups. *Testing, Psychometrics, Methodology in Applied Psychology*, 25(3), 409–427. <https://doi.org/10.4473/TPM25.3.6>.
- Bizzi, F., Shmueli-Goetz, Y., Castellano, R., San Martini, P., & Cavanna, D. (2021b). Psychometric properties of the Child Attachment Interview: an Italian study with specific clinical groups. *Psychological Report*, 124(1), 382–410. <https://doi.org/10.1177/0033294120905515>.
- Borelli, J. L., Cohen, C., Pettit, C., Normandin, L., Target, M., Fonagy, P., & Ensink, K. (2019). Maternal and child sexual abuse history: an intergenerational exploration of children's adjustment and maternal trauma-reflective functioning. *Frontiers in Psychology*, 10, 1–11. <https://doi.org/10.3389/fpsyg.2019.01062>.
- Borelli, J. L., Compare, A., Navely, J. E., & Decio, V. (2015). Reflective functioning moderates the association between perceptions of parental neglect and attachment in adolescence. *Psychoanalytic Psychology*, 32(1), 23. <https://doi.org/10.1037/a0037858>.
- Borelli, J. L., Ensink, K., Hong, K., Sereno, A. T., Drury, R., & Fonagy, P. (2018). Reflective functioning predicts lower cardiovascular reactivity in children. *Frontiers in Medicine*, 5, 1–12.
- Borelli, J. L., Hong, K., Rasmussen, H. F., & Smiley, P. A. (2017). Reflective functioning, physiological reactivity, and overcontrol in mothers: links with school-aged children's reflective functioning. *Developmental Psychology*, 53, 1680–1693. <https://doi.org/10.1037/dev0000371>.
- Borelli, J. L., St John, H. K., Cho, E., & Suchman, N. E. (2016). Reflective functioning in parents of school-aged children. *American Journal of Orthopsychiatry*, 86(1), 24. <https://doi.org/10.1037/ort0000141>.
- Blakemore, S. (2017). Social-cognitive development during adolescence. In D. Skuse, H. Bruce & L. Dowdney (Eds.), *Child psychology and psychiatry: frameworks for clinical training and practice*. 3rd ed. (pp. 91–95). London: Blackwell.
- Bosacki, S., & Astington, J. W. (1999). Theory of mind in pre-adolescence: relations between social understanding and social competence. *Social Development*, 8, 237–255. <https://doi.org/10.1111/1467-9507.00093>.
- Byrne, B. M. (2016). *Structural equation modeling with AMOS: basic concepts, applications, and programming*. 3rd ed. New York, NY: Taylor & Francis Group.
- Cavanna, D., Bizzi, F., San Martini, P., & Castellano, R. (2018). Child Attachment Interview: a psychometric study of an Italian sample. *Psicologia Clinica dello Sviluppo*, 22(1), 115–141. <https://doi.org/10.1449/89588>.
- Carr, A. (2017). Social and emotional development in middle childhood. In D. Skuse, H. Bruce & L. Dowdney (Eds.), *Child psychology and psychiatry: frameworks for clinical training and practice*. 3rd ed. (pp. 83–90). London: Blackwell.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>.
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>.
- Costa-Cordella, S., Luyten, P., Cohen, D., Mena, F., & Fonagy, P. (2021). Mentalizing in mothers and children with type 1 diabetes. *Development and Psychopathology*, 33, 216–225. <https://doi.org/10.1017/S0954579419001706>.
- Ensink, K. (2004). *Assessing theory of mind, affective understanding and reflective functioning in primary school-aged children*. London: University College London. Unpublished PhD Dissertation.
- Ensink, K., Bégin, M., Normandin, L., & Fonagy, P. (2016). Maternal and child reflective functioning in the context of child sexual abuse: pathways to depression and externalising difficulties. *European Journal of Psychotraumatology*, 7, 30611. <https://doi.org/10.3402/ejpt.v7.30611>.
- Ensink, K., Bégin, M., Normandin, L., Godbout, N., & Fonagy, P. (2017). Mentalization and dissociation in the context of trauma: Implications for child psychopathology. *Journal of Trauma Dissociation*, 18, 11–30. <https://doi.org/10.1080/15299732.2016.1172536>.
- Ensink, K., Normandin, L., Target, M., Fonagy, P., Sabourin, S., & Berthelot, N. (2015). Mentalization in children and mothers in the context of trauma: an initial study of the validity of the Child Reflective Functioning Scale. *British Journal of Developmental Psychology*, 33(2), 203–217. <https://doi.org/10.1111/bjdp.12074>.
- Ensink, K., Target, M., & Oandasan, C. (2013). *Child reflective functioning scale scoring manual: for application to the Child Attachment Interview*. London: Anna Freud Centre - University College London. Unpublished manuscript.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149–1160. <https://doi.org/10.3758/brm.41.4.1149>.
- Fonagy, P., Gergely, G., & Target, M. (2007). The parent–infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry*, 48(3-4), 288–328. <https://doi.org/10.1111/j.1469-7610.2007.01727.x> PMID: 17355400.
- Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect regulation, mentalization, and the development of the self*. New York, NY: Other Press.
- Fonagy, P., Luyten, P., Moulton-Perkins, A., Lee, Y.-W., Warren, F., & Howard, S., et al. (2016). Development and validation of a self-report measure of mentalizing: the Reflective Functioning Questionnaire. *PLoS ONE*, 11(7), e0158678. <https://doi.org/10.1371/journal.pone.0158678>.
- 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. (2006). The mentalization-focused approach to self pathology. *Journal of Personality Disorders*, 20(6), 544–576. <https://doi.org/10.1521/pe.2006.20.6.544>.
- Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-functioning manual, version 5.0: For application to Adult Attachment Interviews*. London: University College London.

- Fonagy, P., Steele, M., Moran, G., Steele, H., & Higgitt, A. (1991). Measuring the ghost in the nursery: a summary of the main findings of the Anna Freud Centre-University College, London Parent-Child Study. *Bulletin of the Anna Freud Centre*, *14*(2), 115–131.
- Frigerio, A., Cattaneo, C., Cataldo, M., Schiatti, A., Molteni, M., & Battaglia, M. (2004). Behavioral and emotional problems among Italian children and adolescents aged 4 to 18 years as reported by parents and teachers. *European Journal of Psychological Assessment*, *20*(2), 124–133. <https://doi.org/10.1027/1015-5759.20.2.124>.
- Frith, C. D. (1994). Theory of mind in schizophrenia. In A. S. David & J. C. Cutting (Eds.), *The neuropsychology of schizophrenia* (pp. 147–162). Hove: Lawrence Erlbaum Associates.
- George, C., Kaplan, N., & Main, M. (1985). *Adult attachment interview*. Berkeley: University of California Department of Psychology. Unpublished manuscript.
- Ha, C., Sharp, C., Ensink, K., Fonagy, P., & Cirino, P. (2013). The measurement of reflective function in adolescents with and without borderline traits. *Journal of Adolescence*, *36*(6), 1215–1223. <https://doi.org/10.1016/j.adolescence.2013.09.008>.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- Hughes, C., Fujisawa, K. K., Ensor, R., Lecce, S., & Marfleet, R. (2006). Cooperation and conversations about the mind: a study of individual differences in 2-year olds and their siblings. *British Journal of Developmental Psychology*, *24*, 53–72. <https://doi.org/10.1348/026151005X82893>.
- Katznelson, H. (2014). Reflective functioning: a review. *Clinical Psychology Review*, *34*(2), 107–117. <https://doi.org/10.1016/j.cpr.2013.12.003>.
- Luyten, P., Fonagy, P., Lemma, A., & Target, M. (2012). Depression. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 385–417). Washington, DC: American Psychiatric Association.
- Midgley, N., Ensink, K., Lindqvist, K., Malberg, N., & Muller, N. (2017). *Mentalization-based treatment for children: A time-limited approach*. Washington, DC: American Psychological Association.
- Oppenheim, D., & Koren-Karie, N. (2013). The insightfulness assessment: measuring the internal processes underlying maternal sensitivity. *Attachment & Human Development*, *15*(5-6), 545–561. <https://doi.org/10.1080/14616734.2013.820901>.
- Orsini, A., Pezzuti, L., & Picone, L. (2012). WISC-IV: Contributo alla taratura italiana.
- Rosso, A. M., & Airoldi, C. (2016). Intergenerational transmission of reflective functioning. *Frontiers in Psychology*, *7*, 1903. <https://doi.org/10.3389/fpsyg.2016.01903>.
- Shai, D., & Belsky, J. (2017). Parental embodied mentalizing: How the nonverbal dance between parents and infants predicts children's socio-emotional functioning. *Attachment & Human Development*, *19*(2), 191–219. <https://doi.org/10.1080/14616734.2016.1255653>.
- Sharp, C., Penner, F., & Ensink, K. (2020). Reflective function and borderline traits in adolescents. *Journal of Personality Disorders*, *34*, Supplement B, 1–16. [https://doi.org/10.1521/pedi\\_2019.33416](https://doi.org/10.1521/pedi_2019.33416).
- Shmueli-Goetz, Y., Target, M., Fonagy, P., & Datta, A. (2008). The Child Attachment Interview: a psychometric study of reliability and discriminant validity. *Developmental Psychology*, *44*, 939–956. <https://doi.org/10.1037/0012-1649.44.4.939>.
- Skårderud, F. (2007). Eating one's words, part II: the embodied mind and reflective function in anorexia nervosa-theory. *European Eating Disorders Review: The Professional Journal of the Eating Disorders Association*, *15*(4), 243–252. <https://doi.org/10.1002/erv.778PMID:17676695>.
- Slade, A. (2005). Parental reflective functioning: an introduction. *Attachment & Human Development*, *7*, 269–281. <https://doi.org/10.1080/14616730500245906>.
- Steele, H., Steele, M., Kerns, K. A., & Richardson, R. A. (2005). The construct of coherence as an indicator of attachment security in middle childhood: the friends and family interview. *Attachment in middle childhood* (pp. 137–160). New York, NY: Guilford Press.
- Suchman, N. E., DeCoste, C., Leigh, D., & Borelli, J. L. (2010). Reflective functioning in mothers with drug use disorders: implications for dyadic interactions with infants and toddlers. *Attachment and Human Development*, *12*, 567–585. <https://doi.org/10.1080/14616734.2010.501988>.
- Tessier, V. P., Normandin, L., Ensink, K., & Fonagy, P. (2016). Fact or fiction? A longitudinal study of play and the development of reflective functioning. *Bulletin of the Menninger Clinic*, *80*(1), 60–79. <https://doi.org/10.1521/bumc.2016.80.1.60>.
- Vanwoerden, S., Greiner, I., Ensink, K., & Sharp, C. (2019). The relations between self-and caregiver-focused reflective function and theory of mind in the context of borderline pathology in adolescence. *Psychiatry Research*, *273*, 274–280. <https://doi.org/10.1016/j.psychres.2019.01.042>.
- Venta, A., Shmueli-Goetz, Y., & Sharp, C. (2014). Assessing attachment in adolescence: A psychometric study of the Child Attachment Interview. *Psychological Assessment*, *26*(1), 238–255. <https://doi.org/10.1037/a0034712>.
- Wechsler, D. (2003). *Wechsler intelligence scale for children-Fourth edition (WISC-IV)*. San Antonio, TX: The Psychological Corporation.