Validation study of the Roberts Apperception Test for Children (RATC) in an adolescents' forensic sample

Pedro Armelim Almiro [a,d,*], Pedro Rafael Fernandes Marques [b], Mónica Costa Duarte [b], Isabel Marques Alberto [b], Mário Rodrigues Simões [c]

[a] Center for Research in Neuropsychology and Cognitive Behavioral Intervention (CINEICC, FPCEUC), Portugal
[b] University of Coimbra (Portugal), Faculty of Psychology and Educational Sciences, Portugal
[c] University of Coimbra (Portugal), Center for Research in Neuropsychology and Cognitive Behavioral Intervention (CINEICC), Faculty of Psychology and Educational Sciences, Portugal
[d] Autonomous University of Lisbon (Portugal), Psychology Research Centre (CIP/UAL), Portugal

ABSTRACT

The *Roberts Apperception Test for Children* (RATC) is a projective method developed by McArthur and Roberts (1982) to assess children and adolescents' behavioural, social and emotional functioning, concerns, conflicts and emotional management strategies through their perceptions of common interpersonal situations of everyday life. The aim of the present study is to contribute to the validation of the RATC in a forensic sample ($N = 75$) [constituted by a group of juvenile delinquents detained in educational centres ($n = 40$, 12–17 years old, 1–10 years of schooling) and a group of maltreated adolescents integrated in residential care ($n = 35$, 11–16 years old, 5–10 years of schooling)], studying its psychometric properties, such as reliability and criterion validity (convergent and discriminant validity), considering the results obtained in other instruments as external validation criteria: Wechsler Intelligence Scale for Children – Third Edition (WISC-III) and Youth Self-Report (YSR). It also aims to search for some indicators based on means and standard deviations to interpret the scores obtained in RATC for these forensic contexts, through the comparisons within forensic groups and the comparisons of the forensic groups with a community sample. The RATC showed minimally acceptable reliability and adequate validity indices, considering that this instrument is a projective method. This limitation is compensated by the clinical value of the data obtained from the projection of individuals' thoughts, concerns, conflicts and problem-solving styles, which are useful to assess their emotional and behavioural characteristics and psychological functioning. The results also show statistically significant differences between the two forensic groups on RATC scales, as well as between them and the community sample, as expected, underlining their different characteristics.

1. Introduction

Psychological assessment is the discipline of scientific psychology which studies a given individual in a specific applied field (e.g., clinical, forensic), using scientific tools (tests and other measurement instruments), with the purpose of answering individual's demands that require scientific operations such as describing, diagnosing, predicting, and explaining (Cohen & Swerdlik, 2018).

Psychological assessment in the forensic settings allows the psychologist to inform the court regarding the psychological functioning of individuals, contributing to decision making in legal matters

(Ackerman, 2010; Heilbrun, 1992), and it is particularly critical in this context given the implications it may have on personal life (e.g., restriction of a person's freedom and well-being) and community (Heilbrun et al., 2009; Melton et al., 2007). The personality assessment, as a domain of psychological assessment, plays an important role to make informed decisions since it allows identifying what people are like and how they are likely to think, feel, and act (Cohen & Swerdlik, 2018; Weiner & Greene, 2017). Therefore, personality can be defined as the more or less stable and enduring organization of a person's character, temperament, intellect, and physique, which determines his unique adjustment to the environment; character denotes a person's more or less

stable and enduring system of conative behaviour (will); temperament, the system of affective behaviour (emotion); intellect, the system of cognitive behaviour (intelligence); physique, the system of bodily configuration and neuro-endocrine endowment (Eysenck & Eysenck, 1985). Or, in a simplest way, personality refers to psychological qualities that contribute to an individual's enduring and distinctive patterns of feeling, thinking, and behaving, which distinguish one person from another and are relatively stable in different situations and over time (Cervone & Pervin, 2018).

Consequently, the personality is a fundamental component of the human being and it is characterised in a personal way and manifests itself in all the attitudes, interests and behaviours of the individual. One of the means of access to the personality and its characteristics are the projective methods, based on the assumption that when faced with ambiguous situations (a relatively unstructured materials), the individual will respond according to his or her personality (Bellak, 1944; Cervone & Pervin, 2018; Frank, 1939, 1948). Therefore, projective methods are the indirect methods to get access of individual's personality characteristics and they aim to apprehend the psychic dynamics of the individual as a whole; whereas the objective methods, through the self-report scales or questionnaires, are the direct methods to get access of individual's personality characteristics (Anastasi & Urbina, 2000; Cohen & Swerdlik, 2018). Nevertheless, self-report scales do not always are sufficiently discriminant and do not always show differences between the normative and clinical groups regarding emotional functioning, such as depression, anxiety and self-esteem. Thus, when children keep these perceptions to themselves to avoid them, the only way to access them is through the use of projective methods (Leifer et al., 1991). For example, according to Joiner (1996), the Roberts Apperception Test for Children (RATC) appears to suffer less from the effects of defensiveness than Children's Depression Inventory (CDI) to evaluate depression, which may be an advantage; "self-report measures may be more influenced by self-attributed than implicit motives, and thus may be more affected by social desirability and defensiveness" (Joiner, 1996, p.804). Referring to self-report measures, social desirability refers to the bias or tendency of individuals to present themselves in a more favourable way toward others, giving answers that are in accordance with social norms, and defensiveness refers to the tendency to be sensitive to criticism or comments about the person's limitations or difficulties, and to counter or deny them (as defined in VandenBos, 2015; see Wetzel et al., 2016).

In the domain of non-self-report measures to assess personality, the *Roberts Apperception Test for Children* (RATC) is a well-known projective method developed by Dorothea McArthur and Glenn Roberts to assess children and adolescents' behavioural, social and emotional functioning, concerns, conflicts and emotional management strategies through their perceptions of common interpersonal situations of everyday life (Anastasi & Urbina, 2000; Cohen & Swerdlik, 2018; McArthur & Roberts, 1982). The RATC is based on the assumption that the child or adolescent participant responds according to his or her problems, characteristics and coping strategies. Since the RATC uses novel scenarios through cards, the participant responds with a minimum of distorting intention or defensiveness, and he or she is less likely to engage in social desirability bias. Aiken (1996, p.371) refers to the RATC as a "promising test which allows for the construction of stories on a wide range of topics, and which has a well-documented and easy to learn rating system".

The RATC is administered using 16 cards, 11 cards have different versions for male and female participants, while the other 5 cards are gender neutral and are administered to all participants. These cards were developed to be adjusted to children and adolescents, describing interpersonal situations of the various life contexts, involving children in their relationships with peers or with adults (e.g., parent-child relationships, sibling relationships, aggression situations). The child or adolescent participant is asked to develop a story with a beginning, middle and end about each card (McArthur & Roberts, 1982). Considering the "projective hypothesis" (attributed to Frank, 1939, 1948), it is

assumed that the individual projects his or her thoughts, concerns, conflicts and problem-solving styles onto these stories, therefore this instrument is a thematic approach to personality assessment.

The RATC consisting of 8 Adaptive Scales, 5 Clinical Scales and 3 Clinical Indicators, and this projective method was designed to allow for both quantitative and qualitative analyses (McArthur & Roberts, 1982). The Adaptive Scales are: Reliance on Others (REL), assesses individual's tendency to make up stories which characters reaches out to others for help in problem solving (it is an adaptative capacity to ask outside help); Support-Other (SUP-O), assesses the tendency to support others by giving help, emotional support, or material objects; Support-Child (SUP-C), assesses self-sufficiency and maturity as indicated by assertiveness or positive emotions; Limit Setting (LIM), assesses the extent to which authority figures place reasonable limits on the child when he or she break the rules; Problem Identification (PROB), assesses the ability to formulate concepts beyond the nature of the card (it is adaptive and requires lack of defensive behaviour to view others in dynamic interactions); Resolution-1 (RES-1), indicates a tendency to seek easy or unrealistic solutions to problems (defence, denial, naivety); Resolution-2 (RES-2), indicates a constructive resolution of a problem limited to the situation (of internal feelings, interpersonal relationship, external outcomes); Resolution-3 (RES-3), indicates a constructive resolution of a problem with new insight beyond current problem.

The Clinical Scales are: Anxiety (ANX), assesses the manifest anxiety or apprehension of the characters, and also worry, guilt, and remorse; Aggression (AGG), assesses the extent to which the characters express anger and engage in physical or verbal aggression; Depression (DEP), assesses sadness, despair, or physical symptoms of depression (fatigue, apathy, sleeplessness); Rejection (REJ), themes of separation, jealousy, discrimination, or feelings of being left out; Unresolved (UNR), when individual states a problem in the story which the characters are unable to or do not resolve.

And the Clinical Indicators are: Atypical Response (ATY), when individual combines a number of responses which indicate extreme deviation from the usual themes of the card, including distorted emotion, denial of obvious aspects of a picture, unrealistic content (primary process thinking); Maladaptive Outcome (MAL), describes an outcome of characters acting in socially disapproved way, show inappropriate use of defences, resolving a problem by withdrawing or by taking over autocratically, or acting out, deceiving, manipulating; Refusal (REF), when individual refuses to give a response to a card (defensive behaviour). In addition to the scales and indicators there are 3 supplementary measures: Ego Functioning Index, Aggression Index and Projection Levels (McArthur & Roberts, 1982). Defensive behaviour refers to the individuals' overuse of defence mechanisms operating at an unconscious level (as defined in VandenBos, 2015).

The original RATC provides normative data from a sample of 200 well-adjusted children (100 boys and 100 girls, aged between 6 and 15 years old) and the scores on each of the scales are converted into T-scores ($M = 50$, $SD = 10$), with significant deviation operationalised as a standard deviation in either direction (McArthur & Roberts, 1982).

Regarding on psychometric studies of the RATC, McArthur and Roberts (1982) obtained split-half reliability indices ranging from 0.86 to 0.44 using the Lord and Novick formula, and indices ranging from 0.86 to 0.48 using the Spearman-Brown correction ($N = 400$, 200 normative group and 200 clinical group). And Alberto (1999), in a Portuguese sample, obtained internal consistency indices ranging from 0.77 to 0.20 ($N = 92$).

McArthur and Roberts (1982) examined the construct validity through principal component analysis with varimax rotation. They obtained a three-factor solution: Factor I (accounted for 29.60 % of the variance) defined by Unresolved and six Adaptive Scales, Factor II (15.10 % of the variance) defined by four Clinical Scales, and Factor III (10.80 % of the variance) defined by Limit Setting and Resolution-1. Comparisons between the normative group and a clinical group showed significant differences for almost all scales. Considering the

Table 1
Studies with RATC based on a narrative review.

Authors	Objectives	Samples	Results/conclusions
Alberto (1999)	Assessing of symptomatology and PTSD in abused and neglected children.	92 children from the general population and 90 abused children in residential care (both sexes, aged 10–15 years old).	Group of abused children recorded higher values in Resolution 1 and in Support-Child and, on the other hand, lower on the Support-Other scale. Statistically significant results were identified for the group × gender interaction on the scales Limit Setting, Support-Child and Support-Other, with the male group of the abuse group registering higher values than the female gender, while in the control group the opposite happens. Statistically representative results were identified for the Problem Identification scale, with older children in the abuse group having higher values than younger children. No statistically significant differences were identified between the abuse and control groups on the Depression which corroborates to the lack of discriminant validity of the RATC between the control and clinical groups.
Bell and Nagle (1999)	Validation study of the RATC in a sample of children, analysing the adequacy of the standardization norms with non-clinical samples.	86 children from the general population (59 boys and 27 girls, mean age of 9 years and 11 months).	The authors concluded that the sample used for standardization is inadequate and should not be used for clinical diagnoses.
Burman et al. (1987)	Analysing the relationship between parental conflict and children's adjustment.	56 children (30 boys and 26 girls, aged 6–14 years old) and their parents.	Fathers' overall marital satisfaction is positively associated with the sons' adaptability, as scored on the RATC. Data from boys indicate that a warm mother-child relationship is associated with high adaptability, while a warm father-child relationship is related to low RATC clinical scale scores.
Canais (2012)	Comparing the results between a group of institutionalised and a group of non-institutionalised youths in the RATC scales and the Rosenberg Self-Esteem Scale (RSE).	30 institutionalised youths (13 boys and 17 girls) and 30 youths from the general population (21 boys and 9 girls), aged 10–15 years old.	The control sample has higher values in the Support-Other, Identification Problems and Resolution-2 scales, these youths have a higher tendency to provide support to other people and to analyse, identify and define problems because they have had more enriching and reassuring family and social experiences. Institutionalised youths have difficulties in terms of awareness of feelings, conflicts, or problems, and in discriminating the resolution steps, consequently having higher scores on the Unresolved scale.
Duncan (1993)	Studying differences between clinical and non-clinical groups of children, and their respective mothers, and relations between children and their mothers in RATC scales.	70 mother-child dyads.	Although no significant differences were found on clinical scales between groups, children in the clinical group, as well as their mothers, were less likely to project support-other, support-child, problem identification, or problem solving compared to with the control group. They were also more likely to report ineffective or abusive boundary-setting, maladaptive outcomes, rejection and atypical themes.
Friedrich and Share (1998)	Analysing the content analysis of stories developed from card 15 to identify sexual content.	93 children (59 boys and 34 girls, aged 4–13 years old), considering three groups: 59 children with no evidence of sexual abuse, 18 children possible victims of sexual abuse, and 16 children victims of sexual abuse.	The sexual responses are clearly not specific only to cases of possible or probable sexual abuse. Approximately 1 in 5 non-abused children referred for a psychological evaluation provided responses that were scored as sexual. The authors concluded that it is not possible to identify only one measure to accurately identify a child as sexually abused.
Gonçalves et al. (1999)	Normative study of RATC data based on a sample of well-adjusted children and comparison with a clinical sample.	80 children from the general population, stratified by ages (40 boys and 40 girls, aged 6–9 years old), and 44 children from a clinical sample (depression, anxiety, behavioural problems).	The authors obtained statistically significant differences between the two samples, with the clinical group having generally higher results on the clinical scales (e.g., Unresolved scale) and the normative sample registering higher results on the adaptive scales (e.g., Support-Other and Resolution-2 scales). There were no statistically significant differences in the Anxiety, Depression, Aggression and Rejection scales. Adaptive Scales increase significantly with age. Differences according to gender were recognized on the Anxiety scale, where girls have higher scores.
Headen (1986)	Studying the discriminative capacity of the RATC between a group of students with learning and/or behavioural problems and a group of children without these problems.	29 children with learning and/or behavioural problems (21 boys and 8 girls, aged 6–15 years old), and 29 children from a control group (14 boys and 15 girls, aged 6–12 years old).	Significant differences between the two samples in the type of resolution they gave to situations identified as problematic – children in the clinical sample gave answers that corresponded

Table 1 (*continued*)

Authors	Objectives	Samples	Results/conclusions
			to the lowest level of resolution. No differences were found between the two samples in the results obtained in the remaining adaptive and clinical scales. In addition, he found a positive correlation between cognitive performance measures and the Adaptive Problem Identification scale.
Joiner & Barnett (1994)	Analysing the correlation between the RATC scales and the effect of age on the results of the various scales.	53 male children/adolescents admitted to academic medical centres (aged 6–16 years old).	When checking the correlations between the results of the different RATC scales, they found that Aggression was strongly associated with Rejection, while Support-Other emerged as a moderating variable in the relationship between Depression and Rejection. It was found that the RATC Depression indexes interacted with the measure of interpersonal style, in this case Support-Other, to predict rejection rates.
Joiner (1996)	Analysing the susceptibility of self-report measures (CDI and RCMAS) and thematic measures (RATC) to simulation in the assessment of depression in children and adolescents admitted to psychiatric clinics.	44 children and adolescents, psychiatry inpatients (20 boys and 24 girls, aged 6–16 years old).	Defensiveness was more associated with self-report measures than thematic measures of depression in both genders, but being particularly relevant in females.
Lampel (1996)	Studying the personality and parenting characteristics of litigious couples and their children's patterns of preference for one parent in custody dispute cases.	20 families in dispute: one group of 10 children (6 boys and 4 girls) who have a preference for one of the parents, and another group of 10 children (6 boys and 4 girls) with no preference for one of the parents.	In the RATC results, was only found a statistically significant difference, on the Problem Identification scale, with the group of children who did not have a preference for either parent to narrate more situations as problematic.
Lavado (2008)	Studying the attachment patterns and the presence of psychopathological symptoms in children with atopic dermatitis.	5 children (4 girls and 1 boy, aged 9–10 years old) with atopic dermatitis.	In this study, the psychometric properties of the RATC were not analysed. However, high levels of anxiety were detected in all children in the sample. Lower values were found on the Support-Other scale in three children, revealing difficulties in recognizing others as possible allies in solving problems.
Louw and Ramkisson (2002)	Studying the adequacy of the RATC, the House-Tree-Person (H-T-P) test, and the Draw-A-Person test as measures of assessing sexual abuse.	23 sexually abused girls and 17 non-abused girls (aged 7–11 years old).	The two groups did not differ significantly with regards to the variables on the Adaptive Scales and Clinical Scales and Indicators. The sexually abused girls tended to reflect more sexual responses on the RATC cards, and the sexual content was defined as responses which reflected an explicit sexual act or an implied sexual act.
Mendes and Sani (2015)	Studying children's representations of inter-parental violence using RATC card 12 "parental conflict" and a "Dinnertime Conflict" Story (MacArthur Story Stem Battery).	9 children (6 girls and 3 boys, aged 8–15 years old) exposed to inter-parental violence.	The children projected in their narratives a parental relationship that was always conflicting, with events of physical and emotional violence, without resolution. The narratives reveal a diversity of emotional and cognitive reactions, and the results confirm the negative impact of violence on the child's adjustment level. The elaborated stories are disorganized and incoherent, presenting weak verbal resources.
Palomares et al. (1991)	Analysing the factor structure of the RATC in a sample of children with chronic diseases (comparing with the factor structure of the original study).	48 children with chronic diseases (aged 6–15 years old).	They obtained the three factors indicated in the original study, although they point to some differences in the composition of the three factors.
Wells et al. (2012)	Analysing the effectiveness of neurocognitive rehabilitation in children withdrawn from their parents due to substance exposure in the prenatal period.	40 children (27 boys and 13 girls, aged 6–11 years old) were removed from their biological families, and 38 children in the control group (26 boys and 12 girls).	Children in the intervention group demonstrated significant improvements in executive and emotional functioning when compared with the control group. There was a statistically significant decrease in the results of the Resolution-1 scale in the intervention group, because they adopted more realistic, complex, and elaborated responses in problem solving.
Worchel et al. (1992)	Studying the influence of suggestion on the results of various measures of depression in children.	50 paediatric children/adolescents (28 boys and 22 girls, aged 6–17 years old) with chronic illness (19 with cancer and 31 with diabetes).	Self-report of depressive symptoms by paediatric patients varied as a function of environmental cues, while scores on the RATC Depression scale did not change, providing important validity data.

construct validity, Palomares et al. (1991) carried out an principal component analysis and also obtain a three-factor solution with factor loadings >0.51 (comparing factor solutions for normative and clinical groups), but only Factor I is equivalent to the original factorial structure; the authors also performed a confirmatory factor analysis and yielded a satisfactory adjustment [normative group: $\chi^2(62) = 348.57$, $p < .001$, GFI = 0.89, RMS = 0.09; clinical group: $\chi^2(48) = 146.65$, $p < .001$, GFI = 0.72, RMS = 0.12].

In Portuguese context, Gonçalves et al. (1999) carried out an exploratory factor analysis ($N = 123$, 80 control group and 43 clinical group) and obtained a four-factor solution: Factor I (accounted for 24.28 % of the variance) defined by most Adaptive Scales, Factor II (14.56 % of the variance) defined by other Adaptive Scales (showing a less adequate adjustment), Factor III (12.55 % of the variance) defined by Depression and Unresolved, Factor IV (12.40 % of the variance) defined by Aggression and Rejection. Comparisons between the control group and a clinical group also showed significant differences for several scales.

And Alberto (1999) also carried out an exploratory factor analysis (varimax rotation) ($N = 92$, control group) and obtained a three-factor solution: Factor I (accounted for 21 % of the variance) defined by five Adaptive Scales and two Clinical Scales (positive social interaction), Factor II (15 % of the variance) defined by other three Adaptive Scales and three Clinical Scales (problem-solving), Factor III (13 % of the variance) defined by three Clinical Scales and two Clinical Indicators (depressive and maladaptive response). Comparisons between the control group ($n = 92$) and a clinical group ($n = 90$, victims of child abuse) also showed significant differences for several scales.

Table 1 presents the studies using the RATC found in literature review (narrative review).

As Table 1 shows, there are not many published (validation) studies. In fact, the lack of evidence for validity of the RATC is a problem that is well documented in the literature (e.g., Dupree & Prevatt, 2003; Frick et al., 2020). For this reason, RATC results cannot be used as the sole or main criterion in the diagnostic decision-making process. However, this limitation and reservation regarding the use of the RATC is common to these thematic projective techniques and, strictly speaking, to the use of any other assessment instrument (Frick et al., 2020). Yet the RATC has several advantages over other projective tests: unlike other thematic techniques, such as Children's Apperception Test (CAT) and Thematic Apperception Test (TAT), the RATC has an explicit, structured and standardized scoring system (although be rare the empirical research to test and guide the interpretations), which makes it easier to administer, quantify and compare results between individuals and allows the researchers to examine its validity; is one of the few storytelling instruments, where the themes of the cards are not specific to psychodynamic theory; it is more culture-fair as it does not use cultural references that may not be familiar to individuals from different backgrounds (Frick et al., 2020; Roberts, 1994; Teglasi, 2010).

The RATC is also used in the assessment of children in forensic settings given its potentialities to analyse emotional and behavioural functioning through the projective process, including victims of sexual abuse (see Louw & Ramkisson, 2002). However, as Joiner (1996) points out, the validity of projective data is not yet clearly demonstrated, and so they should be interpreted with caution, especially if they are not supported by other sources of information or assessment. In addition, based on a sample of well-adjusted children, Bell and Nagle (1999) found that the standardization of the RATC is inadequate, because of the possible misclassification of children in some cases. Therefore, the authors suggested not using the instrument for clinical diagnosis until a new standardization is completed. Therefore, more validation studies are needed.

In this sense, to meet this need, the present study aims to contribute to the validation of the Roberts Apperception Test for Children (RATC) assessing two forensic samples [a group of juvenile delinquents and a group of adolescent victims of maltreatment, with little empirical research with RATC], by (1) enhancing the information about the psychometric properties of the RATC [reliability (through the internal consistency and split-half methods) and criterion validity (convergent and discriminant validity), considering as external validation criteria the results obtained in other instruments], (2) searching for some preliminary indicators based on means and standard deviations for the interpretation of scores attained in the Adaptive Scales, Clinical Scales and Clinical Indicators of the RATC, for a more efficient use in these

forensic contexts, and (3) comparing indices obtained in the RATC among forensic groups and between the forensic groups and a community sample from Canais (2012)' study.

The Canais (2012)' study was considered for the present study, instead of the Alberto (1999)'s and Gonçalves et al. (1999)', because it is the most recent one and has similar age range; as far as is known, there are no studies with more recent normative data developed in Portugal.

Considering these two different forensic groups, juvenile delinquents and adolescent victims of maltreatment, the use of a projective method such as the RATC is very useful, because it assesses different areas of children and adolescents functioning/children and adolescents life contexts, including relationships within the family, peers, and school. The use of RATC in these contexts and the objectives of the present study are based on the fact that both groups have a personal history characterised by early adversity experiences, both in the family and in the community, particularly of abuse and/or neglect, and it is also very common for these juvenile delinquents detained in educational centres. Thus, it seems relevant to identify the narratives about family, peers and school, and the individuals' scores on the Adaptive Scales, Clinical Scales, and Clinical Indicators of the RATC to identify their emotional and behavioural styles, to identify the similarities and differences between these two groups, and to establish some preliminary indicators for the interpretation of RATC scores for clinical and forensic practice.

The hypotheses for the present study are:

H1. Roberts Apperception Test for Children (RATC) show adequate reliability indices.

H2. Roberts Apperception Test for Children (RATC) show adequate criterion validity indices, considering the results on Wechsler Intelligence Scale for Children – Third Edition (WISC-III) for convergent and discriminant validity.

H3. Roberts Apperception Test for Children (RATC) show adequate criterion validity indices, considering the results on Youth Self-Report (YSR) for convergent and discriminant validity.

H4. Considering the ages (11 to 17 years old) and years of schooling (1 to 10 years) range of whole forensic sample, it is expected to find an effect of these variables on the RATC results.

H5. There are some significant mean differences on RATC between two forensic groups, juvenile delinquents and adolescent victims of maltreatment (these groups have different characteristics and forensic issues, but both are forensic groups).

H6. There are significant mean differences on RATC between the forensic groups and the community sample [from Canais (2012)' study].

2. Method

2.1. Participants

Participants are 75 youths of two groups were assessed in the forensic context. The Group A, consisting of 40 juvenile delinquents detained in educational centres of juvenile justice system under Portuguese tutelar education law (Law number 4/2015 of 15th January 2015), all males, aged between 12 and 17 years old (mean age is 15.13 years old, $SD = 1.18$). These youths had between 1 and 10 years of schooling ($M = 5.36$, $SD = 1.78$).

The other group, the Group B, consisting of 35 adolescent victims of maltreatment integrated in residential care under the protection of Portuguese social care system, 19 males (54.3 %) and 16 females (45.7 %), aged between 11 and 16 years old (mean age is 14.31 years old, $SD = 1.37$). These youths had between 5 and 10 years of schooling ($M = 7.86$, $SD = 1.46$).

Another group of children and adolescents ($N = 30$; 21 males and 9 females, aged between 10 and 15 years old) from a community sample of Canais (2012)' study, was considered for comparisons with the two

forensic groups of the present study.

2.2. Instruments

The *Roberts Apperception Test for Children* (RATC; McArthur & Roberts, 1982; European Portuguese validations, Gonçalves et al., 1999, and Alberto, 1999) was used to assess children and adolescents' behavioural, social and emotional functioning. The RATC is a projective method which is composed by 16 cards, including 11 specific cards for boys and girls and 5 common cards, and can be administered to child or adolescent participants aged between 6 and 15 years old, and may be used with slightly older or younger children. The RATC consisting of several Adaptive Scales, Clinical Scales, and Clinical Indicators: Reliance on Others (REL), Support-Other (SUP-O), Support-Child (SUP-C), Limit Setting (LIM), Problem Identification (PROB), Resolution-1 (RES-1), Resolution-2 (RES-2), Resolution-3 (RES-3), Anxiety (ANX), Aggression (AGG), Depression (DEP), Rejection (REJ), Unresolved (UNR), Atypical Response (ATY), Maladaptive Outcome (MAL), and Refusal (REF).

The *Wechsler Intelligence Scale for Children – Third Edition* (WISC-III; Wechsler, 2003; European Portuguese version, Simões et al., 2006) was used to assess the cognitive abilities and functioning. The WISC-III is a measure that can be administered to evaluate children and adolescents aged between 6 and 16 years old and 11 months. It allows determining a general intelligence index (Full Scale Intellectual Quotient), two partial quotients according to verbal abilities (Verbal Intellectual Quotient) and non-verbal abilities (Performance Intellectual Quotient) and also three indicators or factorial indexes: Verbal Comprehension Index, Perceptual Organization Index and Processing Speed Index. It has 13 subtests: Picture Completion, Information, Coding, Similarities, Picture Arrangement, Arithmetic, Block Design, Vocabulary, Object Assembly, Comprehension, Symbol Search, Digit Span and Mazes (the latter three subtests are optional). The WISC-III is an instrument with several validation studies for the Portuguese population and with well-documented and established psychometric properties (Simões et al., 2006).

The *Youth Self-Report* (YSR; Achenbach, 1991; European Portuguese version, Fonseca & Monteiro, 1999) was used to assess the behavioural and emotional functioning. The YSR is a self-report instrument that can be administered to evaluate children and adolescents aged between 11 and 18 years old, aiming to measure their social skills, activities, and behavioural problems as perceived by themselves. The YSR is integrated into Achenbach's System of Empirically Based Assessment (ASEBA) which works as a structured process to collect information and assess the perceptions of different stakeholders, such as parents, teachers and the children or adolescents themselves. The YSR is composed by six scales: Behaviour Problems, Attention Deficit/Hyperactivity Problems, Anxious/Depressed, Withdrawn, Somatic Complaints, Thought Problems. This instrument has adequate psychometric properties (Fonseca & Monteiro, 1999).

2.3. Procedures

Participants were recruited from different forensic contexts, using the nonprobability sampling method, with a convenience sampling. For Group A, consisting of juvenile delinquents detained in educational centres of Portuguese juvenile justice system, the study was carried out with the permission from the Directorate-General of Reintegration and Prison Services (DGRSP) and from the Youth Detention Centres. For Group B, consisting of adolescent victims of maltreatment integrated in residential care under the protection of Portuguese social care system, the permission was obtained from the Technical Directors of the Children and Youth Care Homes. Participants were asked for voluntary participation and the objectives and relevance of the present study were explained to them (no incentives were offered in exchange for participation). They were informed that their responses would remain anonymous, the confidentiality and data protection was guaranteed, and the informed consent information was gathered, according to the

Declaration of Helsinki principles. All ethical principles were followed in the present study. For Group A, RATC and WISC-III were administered, and for Group B, RATC and YSR were administered. Data from Group A and Group B were subsequently compared with a community sample from a previous study, the Canais (2012)' study.

Coding for RATC was carried out for two different psychologists. The inter-rater concordance for RATC results was analysed by using as agreement/disagreement criteria the absence or presence of one of RATC scales in the coding for each card administered. The percentage of agreement ranged from 50 % to 100 % (the lowest percentage was registered in the absence of codes for problem-solving strategies in some cards interpreted by another examiner as unresolved), with a mean of 85 %, generally corresponding to a strong agreement between examiners.

2.4. Statistical analyses

To analyse the reliability of the Roberts Apperception Test for Children (RATC), two alternative methods were performed: the Cronbach's alfa (α), corresponding to internal consistency method (Cronbach, 1951), and the Guttman's second lower bond for test reliability ($\lambda 2$), based on split-half method and considering the square root of the sums of squares of the off diagonal elements of the scale (Guttman, 1945); the latter is a more robust reliability index when the instrument has multiple factors (see Callender & Osburn, 1979), which is the case of RATC. Criterion validity studies were performed through the Pearson's *r* correlation coefficients. The study of the influence of age and years of schooling on the RATC results was performed through ANOVA. Mean differences and comparisons were performed using independent-samples *t-test* and one-sample *t-test*, followed by the calculation of corresponding Cohen (1988)'s *d* indices (adapted to each method) for effect size. Statistical analyses were performed using the IBM *SPSS Statistics* (version 25.0; Armonk, New York, IBM Corp.) and JASP (version 0.16.1; University of Amsterdam) programs.

3. Results

3.1. Reliability

Table 2 presents the reliability coefficients obtained in this study for the whole sample ($N = 75$), which were estimated using the Cronbach's alfa (α) and Guttman's second lower bond for test reliability ($\lambda 2$) (Hypothesis 1). The reliability coefficients ranged for all scales from "minimally acceptable" to "unacceptable", according to the criterion established by DeVellis (2017) to interpret the reliability indices ("unacceptable", coefficients below 0.60; "undesirable", between 0.60 and 0.65; "minimally acceptable", between 0.65 and 0.70; "respectable" between 0.70 and 0.80; "very good", between 0.80 and 0.90): Cronbach's alphas ranged from 0.65 to 0.23, and Guttman's $\lambda 2$ coefficients, from 0.68 to 0.30. All coefficients are less than acceptable, with coefficients below the minimum of 0.70 pointed out by several authors (e. g., Kline, 1993; Nunnally & Bernstein, 1994).

3.2. Criterion validity

Criterion validity can be examined through the degree of correspondence, using the method of correlation, obtained between the scores on the measuring instrument and the individuals' achievement in external criteria associated with or dependent on the psychological dimension the instrument assesses (Nunnally & Bernstein, 1994). Thus, criterion validity should demonstrate that a test score is a predictor of the criterion information obtained from other test scores, measuring similar or related constructs. Relationships between test scores and other measures intended to assess the same or similar constructs provide convergent evidence (convergent validity), whereas relationships between test scores and different constructs provide discriminant evidence

(discriminant validity) (American Educational Research Association et al., 2014).

To analyse the convergent and discriminant validity of the RATC, the scores obtained by participants in the Wechsler Intelligence Scale for Children – Third Edition (WISC-III) and the Youth Self-Report (YSR) were considered as external validation criteria, and Pearson's *r* correlation coefficients were calculated. WISC-III was used for cognitive assessment of 40 juvenile delinquents from Group A, and its results were considered to analyse the convergent validity of RATC assessing individuals' problem-solving strategies [Adaptive Scales: Problem Identification (PROB), Resolution-1 (RES-1), Resolution-2 (RES-2), Resolution-3 (RES-3)] and the discriminant validity of RATC for Clinical Indicators [Atypical Response (ATY), Maladaptive Outcome (MAL)] (Hypothesis 2). YSR was used for emotional and behavioural assessment of the 35 adolescent victims of maltreatment from Group B, and its results were considered to analyse the discriminant validity of RATC for the Adaptive Scales and the convergent validity for the Clinical Scales and Clinical Indicators assessing individuals' behavioural, social and emotional functioning (Hypothesis 3).

Regarding problem-solving abilities, some significant and positive correlations were obtained between RATC scales and WISC subscales and indices, including Full Scale IQ, Verbal IQ, and Performance IQ, with a strength ranging from large ($1 > r > 0.50$) to medium ($0.50 > r > 0.30$), according to the criterion proposed by Cohen (1988) (see Table 3). The majority of correlations appear on RATC's Resolution-2 (RES-2) and Resolution-3 (RES-3) scales. These correlation coefficients ranged between 0.40 and 0.34 in Problem Identification (PROB) scale, between 0.52 and 0.36 in Resolution-2 (RES-2) scale, and between 0.75 and 0.43 in Resolution-3 (RES-3) scale. And a large and negative correlation of -0.54 was obtained between Maladaptive Outcome (MAL) scale and Mazes. Other correlations were not significant and were low or null.

Regarding emotional and behavioural problems, some significant positive and negative correlations were obtained between RATC and YSR scales, with a strength ranging from large ($1 > r > 0.50$) to medium ($0.50 > r > 0.30$) (see Table 4). The correlation coefficients were negative between RATC's Refusal (REF) and Maladaptive Outcome (MAL) scales and YSR scales ranging between -0.52 and -0.31, and were positive between Support-Other (SUP-O) and Somatic Complaints (0.34) and between Limit Setting (LIM) and Withdrawn (0.32). Other correlations were not significant and were low or null.

3.3. Influence of age and years of schooling

For each forensic group, the study of the effect of age and years of schooling on the RATC results was performed through ANOVA (Hypothesis 4). For age, there were no significant effects on RATC results for Group A and Group B, except for the latter on Atypical Response (ATY) [$F(5) = 4.945$, $p = .002$] and Rejection (REJ) [$F(5) = 2.841$, $p = .033$] scales. For years of schooling, only one significant effect for Group A on Unresolved (UNR) scale [$F(8) = 2.472$, $p = .034$], and for Group B on Atypical Response (ATY) scale [$F(5) = 2.627$, $p = .045$].

3.4. Mean differences and some preliminary indicators for use in forensic context

Table 5 presents the mean and standard deviations obtained for RATC scales in the forensic sample ($N = 75$) of the present study: 40 juvenile delinquents detained in educational centres from Group A, and 35 adolescent victims of maltreatment integrated in residential care from Group B.

The mean differences between Group A and Group B were examined using the independent-samples *t-test* and calculating the corresponding Cohen (1988)'s *d* indices (Hypothesis 5). The results indicate that the means were significantly and statistically different for Refusal (REF) [$t(44.13) = 4.281$, $p < .001$; $d = 1.29$] and Support-Child (SUP-C) [t

(62.90) $= 4.121$, $p < .001$; $d = 1.04$], both with a very large effect size, and for Depression (DEP) [$t(73) = 2.088$, $p = .040$; $d = 0.49$], with a medium effect size, showing that Group A tend to obtain higher scores in these scales. On the other hand, the means were significantly and statistically different for Unresolved (UNR) [$t(73) = -4.118$, $p < .001$; $d = -0.96$] and Atypical Response (ATY) [$t(73) = -3.371$, $p = .001$; $d = -0.79$], with a large effect size, showing that Group B tend to obtain higher scores in these scales.

Considering the results, some preliminary indicators (based on means and standard deviations) were stablished for the interpretation of scores in the Adaptive Scales, Clinical Scales and Clinical Indicators of the RATC in these forensic contexts (educational centres and residential care) (see Table 5).

3.5. Comparisons between forensic groups and community sample

Table 6 presents the mean differences comparing the forensic sample of the present study – Group A ($n = 40$) and Group B ($n = 35$) – with Canais (2012)' data from the community sample ($N = 30$) using the one-sample *t-test* and calculating the corresponding Cohen (1988)'s *d* indices (Hypothesis 6). The results indicate significant ($p < .001$, $p < .01$ and $p < .05$) and statistical differences for several scales, ranging from very large to medium effect size, with the same tendencies considering Group A and Group B, as expected. There are significant differences for the Support-Other (SUP-O) [Group A: $t = 3.015$, $p = .006$; $d = 0.63$; Group B: $t = 4.212$, $p < .001$; $d = 0.71$, both with large effect], Unresolved (UNR) [Group A: $t = 2.107$, $p = .047$; $d = 0.44$; medium effect; Group B: $t = 7.944$, $p < .001$; $d = 1.34$; very large effect], and Maladaptive Outcome (MAL) [Group A: $t = 5.514$, $p < .001$; $d = 1.15$; very large effect; Group B: $t = 5.027$, $p < .001$; $d = 0.85$; large effect] scales, showing that participants from the forensic groups tend to obtain higher scores in these scales. And there are significant differences for the Support-Child (SUP-C) [Group A: $t = -2.472$, $p = .022$; $d = -0.52$; with large effect; Group B: $t = -11.105$, $p < .001$; $d = -1.87$; very large effect], Problem Identification (PROB) [Group A: $t = -6.117$, $p < .001$; $d = -1.27$; Group B: $t = -8.547$, $p < .001$; $d = -1.45$; both with very large effect], Resolution-2 (RES-2) [Group A: $t = -17.603$, $p < .001$; $d = -3.67$; Group B: $t = -27.478$, $p < .001$; $d = -4.63$; both with very large effect], Anxiety (ANX) [Group A: $t = -2.105$, $p = .047$; $d = -0.44$; Group B: $t = -2.852$, $p = .007$; $d = -0.48$; both with medium effect], Aggression (AGG) [Group A: $t = -7.238$, $p < .001$; $d = -1.51$; very large effect; Group B: $t = -3.123$, $p = .004$; $d = -0.53$; large effect], Depression (DEP)

Table 2
RATC: Reliability indices considering adaptative and clinical scales (total sample, $N = 75$).

RATC scales	Cronbach's α (internal consistency method)	Guttman's λ2 (split-half method)
Adaptive Scales		
Reliance on Others (REL)	0.60	0.64
Support-Other (SUP-O)	0.46	0.51
Support-Child (SUP-C)	0.23	0.30
Limit Setting (LIM)	0.56	0.60
Problem Identification (PROB)	0.65	0.68
Resolution-1 (RES-1)	0.39	0.46
Resolution-2 (RES-2)	0.63	0.66
Resolution-3 (RES-3)	0.47	0.59
Clinical Scales		
Anxiety (ANX)	0.27	0.35
Aggression (AGG)	0.50	0.54
Depression (DEP)	0.58	0.61
Rejection (REJ)	0.47	0.50
Unresolved (UNR)	0.54	0.57

Note. RATC = Roberts Apperception Test for Children.

[significant differences only for Group B and Canais' comparison: $t = -3.961$, $p < .001$; $d = -0.69$; large effect], and Atypical Response (ATY) [significant differences only for Group A and Canais' comparison: $t = -3.199$, $p = .004$; $d = -0.67$; large effect] scales, showing that participants from the community sample tend to obtain higher scores in these scales. No significant differences were found for the remaining scales (Reliance on Others, Limit Setting, Resolution-1, Resolution-3, and Rejection); the comparison for Refusal (REF) scale was not performed due the score 0 obtained by Canais' community sample.

4. Discussion

The use of projective techniques in psychological assessment of children and adolescents has been the subject of some controversy. As a clinical tool, projective techniques allow for greater flexibility in administration and interpretation. However, with this flexibility, the interpretations that result from the assessment are much more susceptible to be influenced by the examiner's idiosyncrasies, and the interpretations of the same material may vary between clinicians (see Frick et al., 2020). The RATC was developed to overcome the weaknesses of projective tests, but also to correspond more than Children's Apperception Test (CAT) and Thematic Apperception Test (TAT) to the life contexts of children and adolescents (Roberts, 1994). Given the scarcity of studies aimed at analysing the psychometric qualities of the RATC and given its popularity (see Table 1), it is essential to deepen the knowledge about this instrument in various contexts of application, which motivated this study. The main objective of this study was to contribute to the validation of the RATC in forensic setting, considering two different forensic groups (juvenile delinquents and adolescent victims of maltreatment) and searching for some preliminary indicators based on means and standard deviations for the interpretation of RATC results in these contexts.

Hypothesis 1. Roberts Apperception Test for Children (RATC) show adequate reliability indices.

On the reliability analyses of RATC in forensic context, considering the Cronbach's α and Guttman's λ2, all reliability coefficients ranged from "minimally acceptable" to "unacceptable" (ranging between 0.68 and 0.23, in general, with several between 0.65 and 0.60 "undesirable"), with coefficients below 0.70 (which are less than acceptable). These results are similar to Alberto (1999)'s study in the Portuguese context (ranging between 0.77 and 0.20) and McArthur and Roberts (1982)' original study (ranging between 0.86 and 0.44), with normative and larger samples. In addition, some authors (e.g., Anastasi, 1988; Nunnally & Bernstein, 1994) refer that reliability indices can be influenced by the scale length and shorter scales have some tendency to show lower coefficients than the longer ones, which is the case of RATC (only 16 items per scale and low scores variability).

Nevertheless, RATC is a projective method used to assess children and adolescents' behavioural, social and emotional functioning and it has some psychometric limitations, namely for reliability, like some other important psychometric tests for personality assessment. At least these limitations can be studied, whereas in other thematic tests (such as CAT) this is not possible, due to the absence of an explicit, structured and standardized scoring system. Kline (1993) showed some examples of these limitations in the psychometric analyses of objective personality assessment instruments in general – for example, the Minnesota Multiphasic Personality Inventory (MMPI), one of the most widely used personality tests. As Anastasi (1988) pointed out, when the projective methods are evaluated as psychometric instruments (analysing reliability and validity indices), the large majority make a poor showing, however their popularity in clinical use still remains because of the richness of their analyses of psychological functioning and their contribution to treatment outcome. These aspects are also highlighted and discussed by Lilienfield et al. (2000), McGrath and Carroll (2012), and Piotrowski (2019), analysing the scientific status and the usefulness of projective methods, particularly the three major ones, which are the Rorschach Inkblot Test, Thematic Apperception Test, and human figure drawings, being the Rorschach the most studied one.

Although previous studies have found similar reliability coefficients

Table 3
Convergent and discriminant validity between the results obtained in RATC problem solving indices (Adaptative Scales and Clinical Indicators) and WISC-III ($n = 40$, subsample of Group A) (Pearson's r correlations).

WISC-III	RATC					
	Adaptive Scales				Clinical Indicators	
	Problem Identification (PROB)	Resolution-1 (RES-1)	Resolution-2 (RES-2)	Resolution-3 (RES-3)	Atypical Response (ATY)	Maladaptive Outcome (MAL)
Subscales						
Information	0.23	0.19	**0.40***	**0.55****	−0.23	−0.26
Similarities	0.23	0.28	**0.37***	**0.48****	−0.28	−0.12
Arithmetic	0.14	−0.25	0.29	**0.43***	0.05	−0.08
Vocabulary	**0.39***	0.17	0.30	**0.53****	−0.20	0.09
Comprehension	0.30	0.12	**0.39***	**0.75****	−0.17	−0.13
Digit Span	0.31	−0.03	**0.45***	0.34	−0.18	−0.35
Picture Completion	0.29	0.04	0.33	0.25	−0.18	−0.08
Coding	0.13	0.03	0.31	0.20	−0.03	−0.07
Picture Arrangement	0.25	0.03	0.14	0.31	−0.22	−0.06
Block Design	0.27	0.10	0.26	**0.64****	−0.29	−0.15
Object Assembly	−0.01	−0.02	0.01	**0.47****	−0.17	−0.16
Symbol Search	**0.40***	−0.06	0.27	0.22	−0.11	0.06
Mazes	0.11	0.34	0.27	0.24	−0.21	**−0.54****
Indices						
Verbal IQ	0.34	0.07	**0.45****	**0.66****	−0.17	−0.10
Performance IQ	0.28	−0.02	0.30	**0.48****	−0.24	−0.08
Verbal Comprehension Index	0.25	0.17	**0.36***	**0.68****	−0.21	−0.17
Perceptual Organization Index	0.23	0.07	0.23	**0.55****	−0.20	−0.13
Processing Speed Index	0.35	0.05	**0.52***	0.36	−0.01	−0.23
Full Scale IQ	**0.34***	0.07	**0.43***	**0.63****	−0.22	−0.10

Note. RATC = Roberts Apperception Test for Children; WISC-III = Wechsler Intelligence Scale for Children – Third Edition; * $p < .05$; ** $p < .01$. Significant Pearson's r correlations are presented in bold.

Table 4

Convergent and discriminant validity between the results obtained in RATC scales and YSR (*N* = 35, Group B) (Pearson's *r* correlations).

RATC	YSR					
	Behaviour Problems	Attention Deficit/Hyperactivity Problems	Anxious/ Depressed	Withdrawn	Somatic Complaints	Thought Problems
Adaptive Scales						
Reliance on Others (REL)	0.07	−0.26	−0.11	−0.21	0.02	−0.21
Support-Other (SUP-O)	0.20	0.03	−0.12	0.22	**0.34***	0.16
Support-Child (SUP-C)	−0.03	0.04	0.12	0.06	−0.06	−0.00
Limit Setting (LIM)	0.03	0.14	0.18	0.32	−0.02	−0.07
Problem Identification (PROB)	−0.22	−0.24	−0.11	−0.22	0.01	0.10
Resolution-1 (RES-1)	−0.03	−0.21	−0.07	0.02	−0.17	−0.32
Resolution-2 (RES-2)	−0.13	0.02	−0.08	0.07	0.16	−0.11
Resolution-3 (RES-3)	−0.12	−0.13	0.00	0.15	0.24	0.08
Clinical Scales						
Anxiety (ANX)	0.07	0.00	0.06	−0.09	0.02	0.16
Aggression (AGG)	−0.01	0.11	0.13	−0.04	−0.17	0.21
Depression (DEP)	0.04	−0.02	0.10	−0.03	0.04	0.08
Rejection (REJ)	−0.14	−0.20	−0.11	−0.10	−0.21	−0.06
Unresolved (UNR)	0.07	−0.12	0.15	−0.23	0.11	0.29
Clinical Indicators						
Atypical Response (ATY)	−0.17	0.19	0.08	0.07	−0.04	0.07
Maladaptive Outcome (MAL)	0.03	0.10	−0.27	−0.10	**−0.38***	−0.17
Refusal (REF)	**−0.45****	**−0.40***	**−0.36***	**−0.52****	−0.15	−0.31

Note. RATC = Roberts Apperception Test for Children; YSR = Youth Self-Report; * *p* < .05; ** *p* < .01. Significant Pearson's *r* correlations are presented in bold.

Table 5

RATC: Preliminary indicators for interpretation based on mean differences comparing Group A (*n* = 40, juvenile delinquents) and Group B (*n* = 35, adolescent victims of maltreatment).

RATC	Group A	Group B	*t*	df	*p*	Cohen's *d*
	M (SD)	*M* (SD)				
Adaptive Scales						
Reliance on Others (REL)	1.63 (1.90)	1.57 (1.52)	0.135	73	0.893	–
Support-Other (SUP-O)	3.58 (2.06)	3.94 (1.98)	−0.784	73	0.435	–
Support-Child (SUP-C)	1.25 (1.08)	0.43 (0.61)	4.121	62.90	**<0.001***	1.04
Limit Setting (LIM)	1.85 (1.59)	1.83 (1.72)	0.056	73	0.956	–
Problem Identification (PROB)	10.90 (2.71)	9.94 (3.13)	1.420	73	0.160	–
Resolution-1 (RES-1)	1.55 (1.45)	1.26 (1.42)	0.881	73	0.381	–
Resolution-2 (RES-2)	2.45 (2.35)	2.00 (1.82)	0.917	73	0.362	–
Resolution-3 (RES-3)	0.23 (0.62)	0.03 (0.17)	1.925	45.54	0.061	–
Clinical Scales						
Anxiety (ANX)	3.93 (1.70)	3.91 (1.84)	0.026	73	0.979	–
Aggression (AGG)	3.98 (1.64)	4.83 (2.16)	−1.939	73	0.056	–
Depression (DEP)	3.73 (2.05)	2.77 (1.88)	2.088	73	**0.040***	0.49
Rejection (REJ)	2.50 (1.66)	2.54 (2.01)	−0.101	73	0.920	–
Unresolved (UNR)	5.18 (2.32)	7.54 (2.66)	−4.118	73	**<0.001***	−0.96
Clinical Indicators						
Atypical Response (ATY)	0.85 (1.15)	1.89 (1.51)	−3.371	73	**0.001***	−0.79
Maladaptive Outcome (MAL)	1.58 (1.55)	1.06 (1.16)	1.649	71.39	0.104	–
Refusal (REF)	1.05 (1.34)	0.11 (0.32)	4.281	44.13	**<0.001***	1.29

Note. RATC = Roberts Apperception Test for Children; Group A = juvenile delinquents, Group B = adolescent victims of maltreatment; *M* = Mean, *SD* = standard-deviation, *p* = *p*-value, *t* = t-test, df = degrees of freedom, *d* = Cohen's *d* (effect size); * significant *p*-values for *t*-tests are presented in bold.

for RATC (and considering the limitations for other projective or objective instruments found in literature), the Hypothesis 1 of the present study must be rejected.

Hypothesis 2. Roberts Apperception Test for Children (RATC) show adequate criterion validity indices, considering the results on Wechsler Intelligence Scale for Children – Third Edition (WISC-III) for convergent and discriminant validity.

On the criterion validity studies of RATC, several significant correlations were obtained with WISC-III subscales and indices (convergent and discriminant validity), ranging from large to medium [coefficients between 0.75 and 0.34, all positive, except −0.54 between Maladaptive Outcome (MAL) and Mazes]. These results show the adequate association between the problem-solving abilities and managing strategies to deal with common interpersonal situations of everyday life assessed by RATC [particularly with Resolution-2 (RES-2) and Resolution-3 (RES-3),

which are the most adaptive way to solve problems], and the cognitive abilities and functioning assessed by WISC-III. This constitutes evidence of the convergent validity. As expected, the Clinical Indicators, Maladaptive Outcome (MAL) and Atypical Response (ATY) (the latter with no significant correlations), show a negative association with WISC-III subscales and indices, reflecting the difficulties in problem solving in individuals who present more maladaptive ways of managing emotions and behaviours. And this constitutes an evidence of the discriminant validity. Moreover, these juvenile delinquents' development of adaptive resolutions to the problems seems to be related to their cognitive functioning, particularly verbal skills. In this regard, Simões (2003) mentions that verbal deficits limit response options in ambiguous or threatening social situations, since they reflect a lower ability to classify perceptions about the outside world and the emotions expressed by others. Furthermore, the adaptation to unfamiliar situations, processing information speed, and reasoning, are all required skills for effective problem

Table 6
RATC: mean differences comparing the forensic groups with Canais (2012)'s data from the community sample.

RATC	Group A *N* = 40	Group B *N* = 35	Canais' *N* = 30	Group A × Canais' Community sample			Group B × Canais' Community sample		
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*t*	*p*	*d*	*t*	*p*	*d*
Adaptive Scales									
Reliance on Others (REL)	1.63 (1.90)	1.57 (1.52)	1.10 (1.21)	1.485	0.152	–	1.835	0.075	–
Support-Other (SUP-O)	3.58 (2.06)	3.94 (1.98)	2.53 (1.36)	3.015	**0.006***	0.63	4.212	**<0.001***	0.71
Support-Child (SUP-C)	1.25 (1.08)	0.43 (0.61)	1.57 (1.36)	−2.472	**0.022***	−0.52	−11.105	**<0.001***	−1.87
Limit Setting (LIM)	1.85 (1.59)	1.83 (1.72)	2.33 (1.49)	0.684	0.501	–	−1.721	0.094	–
Problem Identification (PROB)	10.90 (2.71)	9.94 (3.13)	14.47 (1.94)	−6.117	**<0.001***	−1.27	−8.547	**<0.001***	−1.45
Resolution-1 (RES-1)	1.55 (1.45)	1.26 (1.42)	1.50 (1.61)	0.355	0.726	–	−1.011	0.319	–
Resolution-2 (RES-2)	2.45 (2.35)	2.00 (1.82)	10.43 (3.48)	−17.603	**<0.001***	−3.67	−27.478	**<0.001***	−4.63
Resolution-3 (RES-3)	0.23 (0.62)	0.03 (0.17)	0.07 (0.25)	1.015	0.321	–	−1.450	0.156	–
Clinical Scales									
Anxiety (ANX)	3.93 (1.70)	3.91 (1.84)	4.80 (2.30)	−2.105	**0.047***	−0.44	−2.852	**0.007***	−0.48
Aggression (AGG)	3.98 (1.64)	4.83 (2.16)	5.97 (2.42)	−7.238	**<0.001***	−1.51	−3.123	**0.004***	−0.53
Depression (DEP)	3.73 (2.05)	2.77 (1.88)	4.03 (1.99)	−0.734	0.470	–	−3.961	**<0.001***	−0.69
Rejection (REJ)	2.50 (1.66)	2.54 (2.01)	3.03 (1.63)	−2.029	0.055	–	−1.437	0.160	–
Unresolved (UNR)	5.18 (2.32)	7.54 (2.66)	3.97 (2.58)	2.107	**0.047***	0.44	7.944	**<0.001***	1.34
Clinical Indicators									
Atypical Response (ATY)	0.85 (1.15)	1.89 (1.51)	1.73 (1.02)	−3.199	**0.004***	−0.67	0.610	0.546	–
Maladaptive Outcome (MAL)	1.58 (1.55)	1.06 (1.16)	0.07 (0.25)	5.514	**<0.001***	1.15	5.027	**<0.001***	0.85
Refusal (REF)	1.05 (1.34)	0.11 (0.32)	–	3.749	**0.001***	0.78	2.095	**0.044***	0.34

Note. RATC = Roberts Apperception Test for Children; Group A = juvenile delinquents, Group B = adolescent victims of maltreatment; *M* = Mean, *SD* = standard-deviation, *p* = *p*-value, *t* = t-test, *d* = Cohen's *d* (effect size); * significant *p*-values for *t*-tests are presented in bold.

solving, and they are assessed by intelligence tests like WISC-III (Gottfredson, 1997), particularly through the factorial indices (Perceptual Organization, Processing Speed) (Kaufman & Lichtenberger, 2000). And this is also true for projective tests, in a different way, like RATC, that assesses problem-solving strategies (considering the children resources for problem solving), identify cognitive deficits (e.g., poor solving skills) and problem-solving components (e.g., identify problems, understanding causal consequences) in unstructured social situation presented on the cards (Teglasi, 2010). Therefore, the Hypothesis 2 is confirmed.

Hypothesis 3. Roberts Apperception Test for Children (RATC) show adequate criterion validity indices, considering the results on Youth Self-Report (YSR) for convergent and discriminant validity.

On the criterion validity studies of RATC, some significant correlations were also obtained with YSR scales (convergent and discriminant validity), ranging from large to medium [negative coefficients between −0.52 and −0.31, and positive coefficients of 0.34 and 0.32, considering the maladaptive (convergent validity) and adaptive (discriminant validity) aspects of behaviour, respectively], showing the association between the behavioural, social and emotional functioning dealing with interpersonal situations assessed by RATC and the emotional and behavioural problems examined by YSR. In this case, most correlations appear on the Refusal (REF) scale, which is a Clinical Indicator. Therefore, the greater the tendency toward Refusal (REF), reflecting higher levels of defensive behaviour, the lower the tendency to identify in oneself the emotional and behavioural problems. Consequently, lower occurrence of Refusal (REF) is associated with the manifestation of these problems in this group of adolescent victims of maltreatment, such as attention deficit or hyperactivity problems, withdrawn, or thought problems (YSR).

Overall, no significant correlations were found between the Clinical Scales of RATC and the scales of YSR, but significant correlations were found for Clinical Indicators [Refusal (REF) and Maladaptive Outcome (MAL)], which partly constitutes evidence of the convergent validity. In addition, no significant correlations were found between the Adaptive Scales of RATC and the scales of YSR [except between Support-Other (SUP-O) and Somatic Complaints (0.34) and between Limit Setting (LIM) and Withdrawn (0.32)], which mainly constitutes evidence of the discriminant validity. Therefore, Hypothesis 3 is only partially confirmed.

Hypothesis 4. Considering the ages (11 to 17 years old) and years of schooling (1 to 10 years) range of whole forensic sample, it is expected to find an effect of these variables on the RATC results.

The study of the effect of age and years of schooling on RATC results shows only a few effects for Group A and Group B. For age, significant effects were found only for Atypical Response (ATY) and Rejection (REJ) scales in the Group B. And for years of schooling, one significant effect was found for Unresolved (UNR) scale in the Group A, and one significant effect was found for Atypical Response (ATY) scale in the Group B. These data show that the Hypothesis 4 must be rejected.

Hypothesis 5. There are some significant mean differences on RATC between two forensic groups, juvenile delinquents and adolescent victims of maltreatment.

To establish some preliminary indicators to interpret RATC data in these forensic contexts, considering our two groups [40 juvenile delinquents from Group A, and 35 adolescent victims of maltreatment from Group B] (and testing Hypothesis 5), means and standard deviations were calculated and comparisons within forensic groups were carried out (followed by the comparisons between the forensic groups and the community sample from Canais' study; see Table 6). Because of these groups' different characteristics and their forensic issues, some significant and statistical differences, with very large to medium effect size, were found between Group A and Group B, as expected (see Table 5). As these results showed, Group A (juvenile delinquents) tends to obtain higher scores in the Refusal (REF) (Clinical Indicator), Depression (DEP) (Clinical Scale), and Support-Child (SUP-C) (Adaptive Scale), whereas Group B (adolescent victims of maltreatment) tends to obtain higher scores in the Unresolved (UNR) (Clinical Scale) and Atypical Response (ATY) (Clinical Indicator). McArthur and Roberts (1982) refer that higher scores in Clinical Scales may suggest individuals' psychological difficulties and conflicts, and higher scores in Clinical Indicators point toward significant clinical information to be analysed. Therefore, these findings were expected considering the individuals' behavioural and emotional functioning in these forensic settings. The least expected result is the differences with Support-Child (SUP-C) scale, used when the individual shows self-sufficiency and maturity. Compared to Group B, Group A shows higher scores on this scale [but their scores are significantly lower when compared with community sample (*t* = −2.472, *p* = .022; *d* = −0.52), see Table 6], which may reflect the low self-esteem, low self-efficacy, and emotional problems felt by adolescents who were victims of maltreatment and abuse. The Hypothesis 5 is confirmed.

Hypothesis 6. There are significant mean differences on RATC between the forensic groups and the community sample [from Canais (2012)' study].

The study of mean differences between the forensic groups (Group A and Group B) of the present study and Canais' data from the community sample also showed several significant and statistical differences, with very large to medium effect size, as expected (see Table 6). As these results showed, the forensic groups tend to obtain higher scores in the Support-Other (SUP-O), Unresolved (UNR), and Maladaptive Outcome (MAL) scales, and the community sample tends to obtain higher scores in the Support-Child (SUP-C), Problem Identification (PROB), Resolution-2 (RES-2), Anxiety (ANX), Aggression (AGG), Depression (DEP), and Atypical Response (ATY) scales. As expected, the forensic groups generally score higher in Clinical Scales and Clinical Indicators, and score lower in Adaptative Scales.

Unresolved (UNR) is a Clinical Scale which scores when the individual states a problem in the story and the characters do not resolve it, and Maladaptive Outcome (MAL) is a Clinical Indicator which is used when the outcome of characters are inappropriate (socially disapproved behaviour, inadequate use of defences, withdrawing or taking over autocratically, deceiving, manipulating), and all those characteristics are coherent with the individual's behavioural, social and emotional functioning assessed in a forensic setting. Support-Other (SUP-O) is an Adaptive Scale which is used when the story entails giving help and emotional support; several cards (such as cards 2 "Maternal Support", 5 "Parental Affection", or 10 "Sibling Rivalry") show clearly supportive interaction, and high scores are frequent in both well-adjusted and clinical individuals (see Roberts, 1994). However, Refusal (REF) is a Clinical Indicator which is used when the individual rejects the card, and this is a significant clinical information, because it never happens in well-adjusted individuals, only in maladjusted ones (individuals with insufficient cognitive development, individuals with emotional blocking, or individuals with oppositional behaviours) (see Roberts, 1994). In the forensic sample of the present study, Group A show a mean score of 1.05 (SD = 1.34) for this scale, and Group B a mean score of 0.11 (SD = 0.32), whereas the Canais' community sample did not score for this scale. All these results were expected. In Group A, consisting of juvenile delinquents, the scores on Refusal may be related with emotional problems and/or oppositional behaviours; whereas in Group B, consisting of adolescent victims of maltreatment, the scores on Refusal may be related with emotional problems (such as emotional blocking).

Support-Child (SUP-C), Problem Identification (PROB), and Resolution-2 (RES-2) are Adaptive Scales. Support-Child (SUP-C) is used when the individual shows self-sufficiency and maturity, Problem Identification (PROB), when the individual formulates concepts beyond the nature of the card, and Resolution-2 (RES-2) when the individual indicates a constructive resolution for a problem (limited to the situation). The higher scores of the community sample on these scales were expected (Table 6), since Adaptive Scales indicate greater adaptive functioning (emotional, behavioural and socially) and they are frequent in well-adjusted individuals (see McArthur & Roberts, 1982). Yet, Anxiety (ANX), Aggression (AGG), and Depression (DEP) are Clinical Scales which are used when the characters show this kind of feelings and behaviours. According to McArthur and Roberts (1982), higher scores in Clinical Scales may reflect individuals' psychological difficulties and conflicts, but they do not entail necessarily a psychological maladjustment (further information is needed). Several cards can be interpreted with these themes (some are more explicit than others) and they should be analysed from case to case. Anxiety (ANX) is used when the character expresses anxiety or reveals apprehension, surprise, guilt, embarrassment, doubts about his or her own competence, illness, etc. Gonçalves et al. (1999) studied the most frequent themes for each of the sixteen cards with clinical and well-adjusted Portuguese children; for example, in card 4 "Support/Aggression" the most frequent theme was "The girl lying down is hurt or sick, and the other girl will help her or seek help"

(75 % of responses, same for the original data, see Roberts, 1994), and this is coded with Anxiety (ANX) scale (among others), hence it accounts for Anxiety (ANX) (among other examples). Aggression (AGG) is used when the character expresses anger (well managed or not) or when is involved in verbal or physical aggression, destruction of objects, etc. In Gonçalves et al. (1999)' study, for example, in card 13 "Aggression Release" the most frequent theme was "Child expressing anger; resolved by another person" (30 % of responses, and a further 50 % for other types of resolution; corresponding to a total of 80 % of the most expect responses for this card; see Roberts, 1994), and this is coded with Aggression (AGG) scale (among others). Therefore, these results can also be evidenced in individuals from a normative sample.

Depression (DEP) mean scores are indeed higher in the community sample (Table 6), and these are the least expected results of the present study. Comparing among the forensic groups, Depression (DEP) mean scores are higher in Group A than Group B [$t(73) = 2.088$, $p = .040$; $d = 0.49$] (Table 5). Depression (DEP) is used when the character expresses sadness, disappointment, fatigue, nightmares, boredom, suffering, apathy, etc. In Gonçalves et al. (1999)' study, for example, in card 12 "Parental Conflict/Depression" the second most frequent theme was "Family sadness by event or problem" (20 % of responses; see Roberts, 1994), and this is coded with Depression (DEP) scale (among others). Therefore, these results are less frequent, but can also be evidenced in a normative sample. Atypical Response (ATY) was also not expected for this group, because it is a Clinical Indicator which reflects an extreme deviation from the usual themes of the card (distorted emotion, denial of obvious aspects of a picture, unrealistic content). This is a significant clinical information and may reflect some psychological difficulties or, more likely, a defensive behaviour in some participants of this community sample. Even individuals from a normative sample may also use denial (which is different from refusal) to manage strong negative emotions, for example, in situations of explicit aggression, such as cards 9 "Physical Aggression" and 13 "Aggression Release". Or may be a regular situation, if they tell stories about "ghosts" or "monsters" from a nightmare in cards 7 "Dependency/Anxiety" and 11 "Fear", which are also coded with Atypical Response (ATY) scale (among others). Therefore, the Hypothesis 6 is confirmed.

The fact that some significant differences between Group A and Group B, and several significant differences between the forensic groups and the Canais' community sample were found, supports the establishment of some preliminary indicators for the interpretation of the RATC data to enhance the assessment of the adolescents in these forensic contexts, particularly in educational centres and residential care.

In general, the RATC showed minimally acceptable reliability indices and adequate criterion validity indices, considering that this instrument is a projective method and a thematic approach, despite lower psychometric robustness when compared with the psychometric tests, as expected, including objective personality assessment instruments. This limitation is compensated by the clinical value of the data that this type of instrument allows to get access through the projection of individual's thoughts, concerns, conflicts and problem-solving styles dealing with stimuli presentation. In spite of these limitations, RATC showed to be useful to assess children and adolescents' behavioural, social and emotional functioning in different forensic contexts, as Alberto (1999) and Gonçalves et al. (1999) evidenced in previous validation studies of this instrument with well-adjusted children and clinical data. However, as with other tests, data from RATC need to be complemented with inputs from other assessment instruments [e.g., Child Behavior Checklist (CBCL), Teacher's Report Form (TRF), Youth Self-Report (YSR), and other projective and objective tests] and techniques (e.g., interview, observation), and other sources of information (e.g., parents, teachers).

Two limitations of the present study are the relatively small sample size (in total), and the large portion of male participants. A larger forensic sample is needed. And a more representative sample of delinquent female adolescents from the educational centres and from residential protection is also needed (all participants in Group A are male,

and 54.3 % of the participants in Group B are also males). This should be improved in future validation studies to analyse a more representative forensic sample to establish indicators of interpretation and carry out factor analysis studies. Another limitation is the use of a reference sample for comparison (Canais' study). Therefore, it is desirable to extent the validation studies with a larger, representative, and normative sample for comparisons. WISC-III and YSR were used in different subsamples and this also constitutes a limitation; for future studies, the evaluation protocols should include more instruments to become more wide-ranging and more effective for criterion validity studies and analysing the RATC potentialities. It is also desirable to develop further criterion validity studies of the RATC using other personality assessment instruments, including projective and objective ones [e.g., projective methods, such as the Children's Apperception Test (CAT), Thematic Apperception Test (TAT), House-Tree-Person (HTP); objective tests: the Minnesota Multiphasic Personality Inventory – Adolescent (MMPI-A), Piers-Harris Children's Self-Concept Scale (PHCSCS)], as well as some behavioural and emotional functioning measures, already mentioned, including symptoms checklists [e.g., Children's Depression Inventory (CDI)].

Declaration of competing interest

We declare that we have no conflicts of interest.

Data availability

The data that has been used is confidential.

References

Achenbach, T. M. (1991). *Manual for the Youth Self-Report and 1991 Profile*. University of Vermont.

Ackerman, M. J. (2010). *Essentials of forensic psychological assessment* (2nd ed.). John Wiley & Sons.

Aiken, L. R. (1996). *Personality assessment: Methods and practices* (2nd ed.). Hogrefe and Huber Publishers.

Alberto, I. (1999). *Avaliação da Perturbação Pós-Stress Traumático (PTSD) e suas consequências psicológicas em crianças e adolescentes vítimas de maltrato. [Assessment of Post-Traumatic Stress Disorder (PTSD) and its psychological consequences in children and adolescents victims of abuse]*. Doctoral dissertation (unpublished). Portugal: Faculty of Psychology and Educational Sciences, University of Coimbra https://estudogeral.sib. uc.pt/handle/10316/949?locale=pt.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). Standards for educational and psychological testing. https://www.apa.org/science/programs/testing/standa rds.

Anastasi, A. (1988). *Psychological testing* (6th ed.). MacMillan.

Anastasi, A., & Urbina, S. (2000). Testagem psicológica. In *Psychological testing*. Artmed.

Bell, N. L., & Nagle, R. J. (1999). Interpretive issues with the Roberts apperception test for children: Limitations of the standardization group. *Psychology in the Schools, 36* (4), 277–283. https://doi.org/10.1002/(SICI)1520-6807(199907)36:4<277::AID-PITS1>3.0.CO;2-Q

Bellak, L. (1944). The concept of projection. *Psychiatry, 7*(4), 353–370. https://doi.org/10.1080/00332747.1944.11022519

Burman, B., John, R. S., & Margolin, G. (1987). Effects of marital and parent- child relations on children's adjustment. *Journal of Family Psychology, 1*, 91–108. https://doi.org/10.1037/h0080439

Callender, J. C., & Osburn, H. G. (1979). An empirical comparison of coefficient alpha, Guttman's lambda-2, and MSPLIT maximized split-half reliability estimates. *Journal of Educational Measurement, 16*(2), 89–99. https://doi.org/10.1111/j.1745-3984.1979.tb00090.x

Canais, J. C. P. (2012). *Diferenças no funcionamento emocional, comportamental e social (Roberts Apperception Test for Children) e na autoestima (Rosenberg Self-Esteem Scale) entre jovens institucionalizados e não institucionalizados [Differences in emotional, behavioural and social functioning (Roberts Apperception Test for Children) and self-esteem (Rosenberg Self-Esteem Scale) between institutionalised and non-institutionalised youths]*. Master's thesis (unpublished). Coimbra, Portugal: Instituto Superior Miguel Torga https://repositorio.ismt.pt/handle/123456789/108.

Cervone, D., & Pervin, L. A. (2018). *Personality: Theory and research* (14th ed.). Wiley.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cohen, R. J., & Swerdlik, M. E. (2018). *Psychological testing and assessment: An introduction to tests and measurement* (9th ed.). McGraw-Hill.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. https://doi.org/10.1007/BF02310555.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE.

Duncan, L. A. (1993). *The Roberts apperception test for children: A study of characteristics of non-clinical children and their mothers and clinical children and their mothers*. Master's thesis (unpublished). Los Angeles, US: California School of Professional Psychology https://www.proquest.com/openview/3ba8c4b82f506ce795bdeb229928de68/1? pq-origsite=gscholar&cbl=18750&diss=y.

Dupree, J. L., & Prevatt, F. (2003). Projective storytelling techniques. In C. R. Reynolds, & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Personality, behavior, and context* (2nd ed., pp. 66–90). The Guilford Press.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. Plenum Press.

Fonseca, A. C., & Monteiro, M. C. (1999). Um inventário de problemas de comportamento Para crianças e adolescentes: O youth self-report de achenbach [An inventory of behavioural problems for children and adolescents: Achenbach's youth self-Report]. *Psychologica, 21*, 79–96.

Frank, L. K. (1939). Projective methods for the study of personality. *The Journal of Psychology: Interdisciplinary and Applied, 8*(2), 389–413. https://doi.org/10.1080/00223980.1939.9917671

Frank, L. K. (1948). *Projective methods*. Charles C. Thomas.

Frick, P. J., Barry, C. T., & Kamphaus, R. W. (2020). *Clinical assessment of child and adolescent personality and behavior* (4th ed.). Springer.

Friedrich, W. N., & Share, M. C. (1998). The Roberts apperception test for children: An exploratory study of its use with sexually abused children. *Journal of Child Sexual Abuse, 6*(4), 83–91. https://doi.org/10.1300/J070v06n04_06

Gonçalves, M., Morais, A. P., Pinto, H., & Machado, C. (1999). Teste Aperceptivo de Roberts para Crianças [Roberts apperception test for children]. Master's thesis (unpublished). In M. R. Simões, M. M. Gonçalves, & L. S. Almeida (Eds.), *Vol.2. Testes e provas psicológicas em Portugal* (pp. 185–198). APPORT/SHO.

Gottfredson, L. S. (1997). Editorial: Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24*(1), 13–23. https://doi.org/10.1016/S0160-2896(97)90011-8

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*, 255–282. https://doi.org/10.1007/BF02288892.

Headen, E. H. (1986). *Roberts' apperception test for children: Referred and nonreferred student profiles*. Master's thesis (unpublished). US: University of Nebraska at Omaha https://digitalcommons.unomaha.edu/cgi/viewcontent.cgi?article=1259&context =studentwork.

Heilbrun, K. (1992). The role of psychological testing in forensic assessment. *Law and Human Behavior, 10*(3), 256–272. https://doi.org/10.1007/BF01044769

Heilbrun, K., Grisso, T., & Goldstein, A. M. (2009). *Foundations of forensic mental health assessment*. Oxford University Press.

Joiner, T. E. (1996). The relations of thematic and nonthematic childhood depression measures to defensiveness and gender. *Journal of Abnormal Child Psychology, 24*(6), 803–813. https://doi.org/10.1007/BF01664741

Joiner, T. E., & Barnett, J. (1994). A test of interpersonal theory of depression in children and adolescents using a projective technique. *Journal of Abnormal Child Psychology, 22*, 595–609. https://doi.org/10.1007/BF02168940.

Kaufman, A. S., & Lichtenberger, E. O. (2000). *Essentials of WISC-III and WPPSI-R assessment*. John Wiley & Sons.

Kline, P. (1993). *Personality: The psychometric view*. Routledge.

Lampel, A. K. (1996). In *, 34. Children's alignments with parents in highly conflicted custody cases Family and Conciliation Courts Review* (pp. 229–239). https://doi.org/10.1111/j.174-1617.1996.tb00416.x

Lavado, A. M. (2008). *Relações de vinculação em crianças com dermatite atópica [Attachment relationships in children with atopic dermatitis]*. Lisboa, Portugal: Instituto Superior de Psicologia Aplicada. https://repositorio.ispa.pt/bitstream/10400.12/598/1/DM%20LAVA-A1.pdf.

Leifer, M., Shapiro, J., Martone, M. W., & Kassem, L. (1991). Rorschach assessment of psychological functioning in sexually abused girls. *Journal of Personality Assessment, 56*, 14–28. https://doi.org/10.1207/s15327752jpa5601_2

Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The scientific status of projective techniques. *Psychological Science in the Public Interest, 1*(2), 27–66. https://doi.org/10.1111/1529-1006.002

Louw, A. E., & Ramkisson, S. (2002). The suitability of the Roberts apperception test for children (RATC), the house-tree-person (H-T-P) and draw-A-person (D-A-P) scales in the identification of child sexual abuse in the indian community: An exploratory study. *Southern African Journal of Child and Adolescent Mental Health, 14*, 91–106. https://doi.org/10.1080/16826108.2002.9632431

McArthur, S., & Roberts, G. (1982). *Roberts apperception test for children manual*. Western Psychological Services.

McGrath, R. E., & Carroll, E. J. (2012). The current status of "projective tests". In H. Cooper (Ed.), *Vol.1. APA handbook of research methods in psychology* (pp. 329–348). American Psychological Association.

Melton, G. B., Petrila, J., Poythress, N. G., & Slobogin, C. (2007). *Psychological evaluations for the courts: Handbook for mental health professionals and lawyers* (3rd ed.). Guilford Press.

Mendes, T. C., & Sani, A. (2015). Representações de crianças expostas a violência interparental através de provas projetivas [Representations of children exposed to interparental violence through projective tests]. *Revista de Psicologia da Criança e do Adolescente, 6*, 171–192.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Palomares, R. S., Crowley, S. L., Worchel, F. F., Olson, T. K., & Rae, W. A. (1991). The factor analytic structure of the Roberts apperception test for children: A comparison of the standardization sample with a sample of chronically ill children. *Journal of Personality Assessment, 56*(3), 414–425. https://doi.org/10.1207/s15327752jpa5603_4

Piotrowski, C. (2019). Projective techniques are not moribund [editorial]. *Journal of Projective Psychology & Mental Health, 26*, 73–76.

Roberts, G. (1994). *Interpretive handbook for the Roberts Apperception Test for Children.* Western Psychological Services.

Simões, M. R. (2003). Avaliação neuropsicológica e comportamentos anti-sociais: Notas breves em torno do papel das competências verbais e das funções executivas [Neuropsychological assessment and antisocial behaviour: Brief notes around the role of verbal skills and executive functions]. In I. Alberto, A. C. Fonseca, C. P. Albuquerque, A. G. Ferreira, & J. Rebelo (Eds.), *Comportamento anti-social: Escola e família [Antisocial behaviour: School and family]* (pp. 129–143). Centro de Psicopedagogia da Universidade de Coimbra.

Simões, M. R., Seabra-Santos, M. J., Albuquerque, C. P., Pereira, M. A., Almeida, L. S., Rocha, A. M., Ferreira, C., Lopes, A. F., Gomes, A. A., Xavier, R. E., Rodrigues, F., Lança, C., Matoso, J., Nunes, J., Filipe, M., & Eusébio, C. (2006). Escala de Inteligência de Wechsler para Crianças – Terceira Edição (WISC-III) [Wechsler Intelligence Scale for Children – Third Edition]. In M. M. Gonçalves, M. R. Simões, L. S. Almeida, & C. Machado (Eds.) (2ª ed.,, *Vol.1. Avaliação psicológica: Instrumentos validados para a população portuguesa [Psychological assessment: Instruments validated for the Portuguese population]* (pp. 199–232). Quarteto.

Teglasi, H. (2010). *Essentials of TAT and other storytelling instruments assessment* (2nd ed.). Wiley.

VandenBos, G. R. (2015). *APA dictionary of psychology* (2nd ed.). American Psychological Association.

Wechsler, D. (2003). *Escala de Inteligência de Wechsler para Crianças – Terceira Edição (WISC-III): Manual* [Wechsler Intelligence Scale for Children – Third Edition (WISC-III): Manual]. CEGOC-TEA.

Weiner, I. B., & Greene, R. L. (2017). *Handbook of personality assessment* (2nd ed.). Wiley.

Wells, A. M., Chasnoff, I. J., Schmidt, C. A., Telford, E., & Schwartz, L. D. (2012). Neurocognitive habilitation therapy for children with fetal alcohol spectrum disorders: An adaptation of the alert Program®. *American Journal of Occupational Therapy, 66,* 24–34. https://doi.org/10.5014/ajot.2012.002691

Wetzel, E., Böhnke, J. R., & Brown, A. (2016). Response biases. In F. Leong, D. Bartram, F. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (pp. 349–364). Oxford University Press.

Worchel, F. F., Rae, W. A., Olson, T. K., & Crowley, S. L. (1992). Selective responsiveness of chronically ill children to assessments of depression. *Journal of Personality Assessment, 59*(3), 605–615. https://doi.org/10.1207/s15327752jpa5903_13