Psychometric properties of Basic Empathy Scale among female juvenile delinquents and school youths☆,☆☆

Pedro Pechorro [a,d,*], Rachel E. Kahn [b], Rui Abrunhosa Gonçalves [a], James V. Ray [c]

[a] School of Psychology, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal
[b] Sand Ridge Secure Treatment Center, Research Unit, MMHI Bld 14, 301 Troy Drive, Madison, WI, 53704, USA
[c] University of Central Florida, Department of Criminal Justice, College of Health and Public Affairs, 12805 Pegasus Drive, Bldg. 80, Suite 311, Orlando, FL 32816, USA
[d] Research Centre for Spatial and Organizational Dynamics, University of Algarve, 8005-139 Faro, Portugal

ARTICLE INFO

ABSTRACT

Research on empathy often distinguishes between affective and cognitive empathy, but there is limited knowledge regarding the application or measurement of these two dimensions of empathy among female youth, especially forensic samples of female youth. The main aim of the present study was to examine the psychometric properties of the Basic Empathy Scale (BES) among a Portuguese sample of female youths ($N = 377$), composed of incarcerated female juvenile offenders ($n = 103$) and school youths ($n = 274$). The two-factor structure of the BES obtained a good fit among the school sample, but the fit among the forensic sample was poor. Both samples demonstrated adequate psychometric properties in terms of Cronbach's alpha, omega coefficient, mean inter-item correlations, corrected item-total correlation range, and criterion validity. However, some caution is advised when using the BES with female youth involved in the juvenile justice system, particularly with incarcerated female youth.

1. Introduction

Due to its complexity, empathy has been defined in a variety of a ways. From a developmental standpoint, empathy is typically defined as an affective response or arousal that is derived from understanding another's emotional state or feelings in a particular situation (Eisenberg, Shea, Carlo, & Knight, 1991). These and other common definitions of empathy emphasize the affective components. However, it is widely accepted that empathy includes both affective and cognitive components that differ in their developmental trajectories, each exerting various influences on empathic behavior (Ang & Goh, 2010; Baron-Cohen & Wheelwright, 2004; Davis, 1980; Decety & Jackson, 2004; Eisenberg & Eggum, 2009). In general, affective features are typically defined as arousal to or resonation and congruence with another's emotional state (Blair, 2005; Hoffman, 1987; Singer & Lamm, 2009). In

contrast, cognitive empathy is often considered synonymous with perspective taking abilities, such as being able to imagine or take the perspective of another in order to understand what they may be feeling (Davis, 1980, 1983; Decety, 2010).

In general, empathy is believed to play an important role in social cognition and prosocial behavior (Decety, 2010). For instance, empathy is thought to be important in the inhibition of aggression and promotion of prosocial behavior (Eisenberg & Eggum, 2009). To this end, shared negative arousal between individuals often results in distress, and serves as a signal that activates empathic concern and thus promotes prosocial behavior. When this shared arousal is absent, there is no motivation to act in order to decrease any discomfort promoted by the negative arousal (Decety & Michalska, 2010). So for those individuals who do not experience this arousal or distress, they may continue to engage in antisocial or aggressive behavior as they cannot understand or experience the distress they may be afflicting on others (Jolliffe & Farrington, 2006).

1.1. Development of the Basic Empathy Scale

Given the relevance of empathy in understanding antisocial or aggressive behavior, it is important to have an adequate measure that captures the multidimensionality of empathy. A measure that can capture both affective and cognitive components of empathy may be especially useful since cognitive empathy may demonstrate a stronger association with offending behavior (van Langen, Wissink, van Vugt, Van der

* Corresponding author.
E-mail addresses: ppechorro@gmail.com (P. Pechorro), rachel.kahn@wisconsin.gov (R.E. Kahn), rabrunhosa@psi.uminho.pt (R. Abrunhosa Gonçalves), james.ray@ucf.edu (J.V. Ray).

Stouwe, & Stams, 2014). Over the years, a variety of self-report measures of empathy have been developed including, but not limited to, the Hogan Empathy Scale (HES; Hogan, 1969), the Questionnaire Measure of Emotional Empathy (QMEE; Mehrabian & Epstein, 1972), and more recently the Interpersonal Reactivity Index (IRI; Davis, 1980). However, these measures of empathy have several noted limitations. They often equate sympathy with empathy, are not designed to or do not adequately capture the cognitive component of empathy, and the validation of many of these measures has mainly relied on university samples (Jolliffe & Farrington, 2006). This third limitation is especially important as validation within university samples may have resulted in the creation of measures that fail to capture the components of empathy that may be most relevant to antisocial or offending behaviors or generalizable to an adolescent population.

Thus, in an attempt to address the limitations of previous instruments, Jolliffe and Farrington (2006) developed the Basic Empathy Scale (BES). They validated a 20-item BES scale in a mixed gender adolescent sample of high school students in England, finding strong support for a two factor structure (cognitive and affective empathy) as well as adequate construct validity with the BES demonstrating expected associations with other measures of empathy or personality constructs such as conscientiousness or agreeableness (Jolliffe & Farrington, 2006). Since that time, the BES has been validated in a wide variety of mixed gender school and community samples from France (D'Ambrosio, Olivier, Didon, & Besche, 2009), Italy (Albiero, Matricardi, Speltri, & Toso, 2009), and China (Geng, Xia, & Qin, 2012). Confirmatory factor analyses across these studies consistently supports the two factor structure of the BES with internal consistencies across samples ranging from α = 0.66–0.81 and 0.73–0.85 for cognitive and affective factors respectively. In addition, across these different cultures, the BES scales demonstrate the expected positive associations with other empathy scales (Albiero et al., 2009; D'Ambrosio et al., 2009), and measures of prosocial behaviors (Albiero et al., 2009; Geng et al., 2012) and show negative associations with measures of internalizing disorders (D'Ambrosio et al., 2009) or emotional problems (Geng et al., 2012). More recently, a Portuguese 16-item adapted version of the BES has been validated in a large community sample of adolescents, with four items having to be removed in order to achieve an acceptable two-factor measurement model (Anastacio, Vagos, Nobre-Lima, Rijo, & Jolliffe, 2016).

However, when the BES has been studied among high risk or delinquent samples of youth, this measure does not always conform to the original two factor structure. For instance, in a mixed gender sample of high risk Hispanic youth involved in gangs, results supported a two factor, 7-item adapted BES scale that excluded negatively worded items in order to achieve a better fit to the data (Salas-Wright, Olate, & Vaughn, 2012). Using a sample of male Portuguese juvenile offenders, Pechorro, Ray, Salas-Wright, Maroco, and Gonçalves (2015) found support for both the original 20-item version of the BES as well as the adapted 7-item BES produced by Salas-Wright et al. (2012). Despite research validating the BES in high risk samples of males, no one has examined the psychometric properties among incarcerated female youth. Given the broad literature base suggesting empathy differences between males and females as well as the increasing rates of aggression and violence among female adolescents over the last two decades (Moretti, Catchpole, & Odgers, 2005; Tracy, Kempf-Leonard, & Abramoske-James, 2009), it is important to explore whether the BES appropriately captures the underlying components of empathy in an incarcerated female adolescent sample.

1.2. Gender Differences in Empathy and Antisocial Outcomes

Gender differences in empathy are apparent from a very early age. For example, at ages 3 to 4 months, females are able to discriminate facial expressions better than their male counterparts as evidenced by their responses to maternal still-face paradigms (McClure, 2000). In addition, throughout childhood and adolescence, females tend to

demonstrate higher levels of empathy and corresponding prosocial behavior (for review see Chaplin & Aldao, 2013). The transition into adolescence, especially around puberty, widens the gender gap even further between males and females (Lam, Solmeyer, & McHale, 2012) suggesting this may be an important period for empathy development. Notably, these gender differences in empathy appear to be developmentally stable throughout the lifespan (Michalska, Kinzler, & Decety, 2013) with females consistently demonstrating higher levels of empathy than males and individuals who demonstrate higher levels of empathy earlier on in development continue to remain higher in empathy throughout development (Eisenberg et al., 1999).

This well-established gender difference in levels of empathy and associated prosocial behavior between males and females has been cited by criminologists and psychologists to help explain why males engage in criminal offenses, especially violent offenses, at higher rates than females and also have higher rates of recidivism (Broidy, Cauffman, Espelage, Mazerolle, & Piquero, 2003; Katsiyannis, Zhang, Barrett, & Flaska, 2004). However, while males have a substantially higher prevalence rate of antisocial behavior than females (Lahey et al., 2000; Moffitt & Caspi, 2001; Rutter, Giller, & Hagell, 1998), rates have been increasing in the past years for women (e.g., Tracy et al., 2009). This is in part due to the field acknowledging that there may be differences in phenotypic expressions of antisocial behavior between males and females. For example, females may be more likely to evidence their aggression towards family members or be aggressive in the home (Robbins, Monahan, & Silver, 2003) and may also be more likely to use manipulative behavior in their criminal acts, while men are more likely to engage in aggressive behavior that results in more violent crimes (Forouzan & Cooke, 2005). In addition, there is evidence to suggest that female youth who engage in delinquent or antisocial behavior experience more disparate outcomes, such as persistently high rates of comorbid psychological symptoms, increased instance of substance dependence, poor quality of romantic relationships, and physically abusive relationships (Moffitt, Caspi, Rutter, & Silva, 2001; Pajer, 1998; Schnittker & John, 2007). For example, Moffitt et al. (2001) found among those female youth in a community sample diagnosed with Conduct Disorder (CD), 72% also met criteria for a depressive or anxiety disorder diagnosis between the ages of 11 to 21. High comorbidity rates are also present in incarcerated samples, where female youth may have higher rates of almost all diagnoses including affective, anxiety, and substance dependence diagnoses (Karnik et al., 2009). In addition, the effects of juvenile delinquency for females carries over into their role as mothers evidenced by increased use of substances during pregnancy and raising children who are more physically aggressive (Tzoumakis, Lussier, & Corrado, 2012). Taken together, there is strong evidence for gender differences in empathy as well as antisocial behavior and its associated outcomes. Thus, it is important to determine whether empathy, as measured by the BES, can be validly assessed in female populations, particularly those that exhibit higher forms of antisocial or aggressive behavior.

Consistent with extant research showing clear differences in empathic behavior between males and females, previous validation studies of the BES have also demonstrated that females tend to score higher in empathy, with effect sizes being more pronounced for the affective empathy component (Albiero et al., 2009; Anastacio et al., 2016; D'Ambrosio et al., 2009; Geng et al., 2012; Jolliffe & Farrington, 2006; Salas-Wright et al., 2012) and no differences have been found between males and females in the factor structure of the BES. However, while past research has only examined the psychometric properties of the BES in community samples or small sub-samples of high risk female adolescents (e.g., Salas-Wright et al., 2012), there has yet to be an examination of the BES in a detained sample of female adolescents.

1.3. Current Study

The aim of the present study is to examine the psychometric properties of the BES, extending its cross-cultural application among a

Portuguese sample of incarcerated female juvenile delinquents and community youths. To our knowledge this is the first study examining psychometric properties of the BES among incarcerated female youths. Psychometric properties such as internal consistency, factor structure, and criterion validity of the BES among a sample of incarcerated and community female adolescents will be examined. Specifically, we will examine the associations between cognitive and affective empathy with theoretically relevant criteria such as psychopathic personality traits, aggression, social desirability, and CD symptoms. It was predicted that: (1) the two-factor structure of the BES would be replicated among the current sample of incarcerated female youths and school youths; (2) the BES would show adequate levels of internal consistency; (3) the BES would show negative associations with existing measures of psychopathic traits, aggression, social desirability and CD symptoms.

2. Material and methods

2.1. Participants

The sample was composed of 377 female participants ($N = 377$; $M_{age} = 16.23$ years; $SD_{age} = 1.38$ years; range = 14–19 years) recruited from forensic and school contexts. Of this total, 103 participants ($n = 103$; $M_{age} = 16.41$ years; $SD_{age} = 1.19$ years; range = 14–18 years) formed the forensic sample and 274 participants ($n = 274$; $M_{age} = 16.17$ years; $SD_{age} = 1.44$ years; range = 14–19 years) formed the school sample. The female inmates were recruited from the three juvenile detention centers managed by the Portuguese Ministry of Justice that admit female detainees. They were all detained by the court's decision, the strictest sanction a Portuguese court can impose. The females from a school context were recruited from public schools of the Lisbon, Algarve, and Coimbra regions. All the participants were informed about the nature of the study and asked to voluntarily participate.

The participants were mainly white Europeans (forensic sample = 59.2%; school sample = 90.1%) from an urban background (forensic sample = 97.1%; school sample = 100%) with a low socioeconomic status (forensic sample = 60.2%; school sample = 39.1%). The detained youths had their crime onset ($M = 12.50$ years; $SD = 1.56$ years) and first criminal problems with the law ($M = 13.27$ years; $SD = 1.55$ years) early in their lives. Most were detained before they were 16 years old ($M = 15.90$ years, $SD = 1.04$ years) due to having committed serious and violent crimes (e.g., robbery, assault).

2.2. Measures

The Basic Empathy Scale (BES; Jolliffe & Farrington, 2006) is a 20-item self-report measure designed to assess empathy in youths. The BES was developed as a concise and coherent scale with the aim of measuring two distinct factors: affective empathy (11 items), and cognitive empathy (9 items). Each item is scored on a five-point ordinal scale (ranging from 1 = *Strongly disagree* to 5 = *Strongly agree*). Scores are calculated by reverse-scoring the appropriate items and then summing the items to obtain the total score and the factor scores. Higher scores indicate an increased presence of the associated characteristics. The Portuguese validation of the BES (Pechorro, Ray, et al., 2015) was used. Internal consistency reliability statistics for the BES are presented below.

The Youth Psychopathic Traits Inventory (YPI; Andershed, Gustafson, Kerr, & Stattin, 2002; Andershed, Kerr, Stattin, & Levander, 2002) is a 50-item self-report measure designed to assess the core personality traits of the psychopathic personality constellation in youth aged 12 years old and up. Each item is scored on an ordinal 4-point Likert scale (ranging from 0 = *Does not apply at all*, to 3 = *Applies very well*). The YPI consists of 10 subscales (with 5 items each) designed in line with Cooke and Michie's (2001) three-dimensional conceptualization of the psychopathy construct, namely: the Grandiose-Manipulative dimension, the Callous-Unemotional dimension, and the Impulsive-Irresponsible dimension. More specifically, the Grandiose-Manipulative dimension consists of the Dishonest charm, Grandiosity, Lying, and Manipulation subscales; the Callous-Unemotional dimension consists of the Callousness, Unemotionality, and Remorselessness subscales; the Impulsive-Irresponsible dimension consists of the Impulsivity, Thrill-seeking, and Irresponsibility subscales. Higher scores reflect an increased presence of the characteristics associated, namely psychopathic traits. The Portuguese version of the YPI was used (Pechorro, Andershed, Ray, Maroco, and Gonçalves 2015; Pechorro, Ribeiro da Silva, Andershed, Rijo, & Gonçalves, 2016). The internal consistency for the current study, estimated by Cronbach's alpha, was 0.94.

The Inventory of Callous-Unemotional Traits (ICU; Essau, Sasagawa, & Frick, 2006; Kimonis et al., 2008) is a 24-item self-report scale designed to assess callous and unemotional traits in youths derived from the callous-unemotional (CU) subscale of the Antisocial Process Screening Device (APSD; Frick & Hare, 2001; Pechorro, Hidalgo, Nunes, & Jiménez, 2016). Each item is scored on a four-point scale (ranging from 0 = *Not at all true*, to 3 = *Definitely true*). Scores are calculated by reverse-scoring the appropriate items and then summing the items to obtain a total score. Using confirmatory factor analysis it was possible to identify three independent factors, namely: Callousness (11 items), Uncaring (8 items), and Unemotional (5 items). Higher scores indicate an increased presence of the associated characteristics. The Portuguese validation of the ICU (Pechorro, Hawes, Gonçalves, & Ray, 2017; Pechorro, Ray, Barroso, Maroco, & Gonçalves, 2016) was used. The internal consistency for the current study, estimated by Cronbach's alpha, was 0.86.

The Reactive-Proactive Aggression Questionnaire (RPQ; Raine et al., 2006) is a self-report measure appropriate for use with youths in late adolescence and young adults that distinguishes between reactive and proactive aggression. The RPQ consists of 23 items rated on a 3-point ordinal scale (0 = *Never*, 1 = *Sometimes*, and 2 = *Often*). A total of 11 items assess reactive aggression (e.g., "Reacted angrily when provoked by others") and 12 items assess proactive aggression (e.g., "Hurt others to win a game"). Summed scores provide measures of reactive or proactive aggression, as well as total aggression. Higher scores indicate higher levels of aggression. The Portuguese validation of the RPQ (Pechorro, Kahn, Ray, Raine, & Gonçalves, 2017; Pechorro, Ray, Raine, Maroco, & Gonçalves, 2017) was used. Internal consistency for the present study, estimated by Cronbach's alpha, was 0.90.

The Socially Desirable Response Set-5 (SDRS-5; Hays, Hayashi, & Stewart, 1989) is a 5-item unidimensional self-report measure designed to assess the degree to which self-report responses may be influenced by social desirability, i.e., the tendency to give socially desirable responses. The items were drawn from the Marlowe–Crowne (MC) form A (Reynolds, 1982), an 11-item short form measure developed from the 33-item Marlowe–Crowne Scale (Crowne & Marlowe, 1960). An advantage of the SDRS-5 relative to many existing measures is the subtle nature of the items. The Portuguese validation of the SDRS-5 (Pechorro, Ayala-Nunes, Nunes, Oliveira, & Gonçalves, 2016) was used. Internal consistency for the present study, estimated by Cronbach's alpha, was 0.70.

A self-report CD scale was also created based on the 15 criteria used to assess CD (see e.g., Skilling, Quinsey, & Craig, 2001). The 15 dichotomous items (coded 0 = No; 1 = Yes) were summated to obtain a total continuous score. Thus, higher scores indicate a higher number of positively endorsed indicators of CD. Based on the Kuder-Richardson coefficient, the internal consistency of the CD scale was considered good (0.89). DSM-5's Conduct Disorder diagnosis (CD; American Psychiatric Association, 2013) was assessed only regarding the forensic sample, using the official diagnostic criteria (i.e., the standard method described in the DSM-5).

In addition, a questionnaire was constructed to describe the sociodemographic and criminal characteristics of the participants. This questionnaire included variables such as participants' age, ethnic group, geographic classification of residence (rural vs. urban), level of schooling

completed, socioeconomic status, parental marital status. Socioeconomic status (SES) was measured by considering both parental level of education and profession, appropriate to the Portuguese context (Simões, 1994).

2.3. Procedures

Authorization to translate and validate a Portuguese version of the BES was obtained from the first author of the scale (Jolliffe & Farrington, 2006). Appropriate procedures (e.g., avoiding item bias or differential item functioning) were followed during the translation and retroversion (Hambleton, Merenda, & Spielberger, 2005). The initial translation from English into Portuguese was completed by the first and last authors of this article, who made sure that young people would be able to properly understand the meaning of the items. The questionnaire was then independently back-translated into English by a native English speaker with considerable professional experience in translating psychology-related scientific texts. The original and the back-translated items were compared for non-equivalence of meaning, and discrepancies were revised until no semantic differences were detected between the English version and the Portuguese version (i.e., the translated items had the same or very similar meanings as the original English items).

Authorization to assess detained female youths was obtained from the General Directorate of Reintegration and Prison Services of the Portuguese Ministry of Justice. The detainees were informed about the nature of the study and asked to voluntarily participate. The participation rate was approximately 89%. Motives for not participating included refusal to participate (6%), inability to participate due to not understanding the Portuguese language (4%) and inability to participate due to security issues (1%). Authorization to assess youths in the school context was obtained from the General Directorate of Education of the Portuguese Ministry of Education, and parental permission was obtained for all children. The participants, students from public schools of the Lisbon, Algarve and Coimbra regions, were informed about the nature of the study and asked to voluntarily participate. The participation rate was approximately 84%. Participants who were unwilling or unable to collaborate were excluded. The measures were administered by means of individual face-to-face interviews in an appropriate setting. Some of the information (e.g., sociodemographic variables) was obtained from self-reports, and institutional files were also used to complement the information obtained (e.g., prior criminal activity and detentions).

2.4. Analytic Strategy

The data were analyzed using SPSS v24 (IBM Corp, 2016) and EQS 6.3 (Bentler & Wu, 2015). The factor structure of the Portuguese language version of the BES was assessed with Confirmatory Factor Analysis (CFA) performed in EQS 6.3 (Bentler & Wu, 2015), with the robust estimation methods. Goodness of fit indices were calculated, including Satorra-Bentler chi-square/degrees of freedom, comparative fit index (CFI), incremental fit index (IFI), root mean square error of approximation (RMSEA), and Akaike Information Criterion (AIC). A chi-square/degrees of freedom value <5 is considered adequate, ≤ 2 is considered good and values = 1 are considered very good (Maroco, 2014; West, Taylor, & Wu, 2012). A CFI ≥ 0.90 and RMSEA ≤0.10 indicate adequate fit, whereas a CFI ≥ 0.95 and RMSEA ≤0.06 indicate good model fit (Byrne, 2006). The incremental fit index is also known as Bollen's IFI; values that exceed 0.90 are regarded as acceptable. In terms of the AIC, lower values indicate a better relative quality of the model (West et al., 2012).

The CFA was performed on the original scale items and items with standardized loading above 0.30 were considered. Modification indexes were used to improve the measurement model. Polychoric correlations were used with robust methodologies to perform the CFA because they

provide more accurate estimates for ordinal items (Byrne, 2006). Cronbach's alpha (α) and omega (ω) coefficients (considered satisfactory if above 0.70), mean inter-item correlations (MIIC; considered good if within the 0.15–0.50 range), and corrected item-total correlation ranges (CITCR; considered adequate if above 0.20) were used to assess reliability (Clark & Watson, 1995; Nunnally & Bernstein, 1994). The omega coefficient was used in the present research because it is currently considered a better estimator of reliability than alpha (see Revelle & Zinbarg, 2009). Pearson correlations were used to analyze associations between scale variables and Spearman correlations were used to analyze associations between ordinal variables and scale variables (Leech, Barrett, & Morgan, 2015). Correlations were considered low if below 0.20, moderate if between 0.20 and 0.50, and high if above 0.50.

3. Results

The first step in assessing the psychometric properties of a Portuguese version of the BES among females was to examine the descriptive statistics for each of the BES items (see Table 1). Some of the items of the forensic sample indicated potential problems such as the median coinciding with the lower value of the item range (e.g., item 4), low mean (e.g., item 4), incomplete item ranges (e.g., item 9 and 15), and very high kurtosis (e.g., item 20). Floor and ceiling effects for the BES total (20/100 and 100/100, respectively), the Affective dimension (11/55 and 55/55, respectively) and the Cognitive dimension (9/45 and 45/45, respectively) were also examined and were considered present if >15% of respondents achieved the lowest or highest possible score, respectively (Terwee et al., 2007). Results indicated that none were found for the BES total scale (forensic sample: floor = 0%, ceiling = 0%; school sample: floor = 0%, ceiling = 0%), the Affective dimension (forensic sample: floor = 0%, ceiling = 0%; school sample: floor = 0%, ceiling = 0%), and the Cognitive dimension (forensic sample: floor = 0%, ceiling = 11.7%; school sample: floor = 0.04%, ceiling = 5.8%).

The second step was to attempt to replicate, by means of CFA using the ML Robust method, the factor structures proposed for this instrument. Table 1 displays the goodness-of-fit indexes that were obtained. We were able to find support in terms of goodness-of-fit indexes for the original two-factor model of the BES for the school sample. That is, across all fit indices, the two-factor model of the BES showed superior fit compared to the one-factor model and all fit indices met the recommended cutoffs after modifications were made to include error covariances among some of the items (see Table 2). Alternatively, however, while the two-factor model showed superior fit compared to the one-factor model for the forensic sample, the fit indices did not meet conventional cutoffs even when modifications were made to the model. The model was then modified by excluding the items with loadings below 0.30 (i.e., item 4 and 15) but the fit indices did not significantly improve. We also tested a 16 item adapted version of the BES by Anastacio et al. (2016), but the model also did not present a good fit.

Table 3 shows the item loadings for the two-factor original model structure estimated with the ML Robust method for both samples. The majority of the items for the two-factor model met cutoff criteria (>0.30) across both samples. That is, the items consistently loaded on their respective scales (i.e., affective and cognitive empathy scales). However, items 4 and 15 failed to reach the 0.30 cutoff value in the forensic sample. Due to the low loadings, items 4 and 15 were excluded from any subsequent analysis regarding the forensic sample.

Table 4 presents the correlations between the BES total and its dimensions among the two samples. Not surprisingly, the dimensions correlated somewhat stronger with the total score across both samples. However, while the correlations between the Affective dimension and the Cognitive dimension of the BES were significant, the strength of the correlations were moderate for the school sample and weak and non-significant for the forensic sample. Tests for equality of the correlations for two independent samples were conducted to determine if the correlations between the Affective and Cognitive scales were

Table 1
Descriptive statistics for BES items among the forensic and school samples.

Forensic/School	Median	Mean(SD)	Range	Skewness	Kurtosis
Item 1	4 / 4	4.06(1.02) / 3.91(0.98)	1–5 / 1–5	−0.89 / -0.47	0.34 / -0.64
Item 2	4 / 4	3.69(0.91) / 3.50(1.02)	1–5 / 1–5	−0.42 / -0.42	0.13 / -0.08
Item 3	5 / 4	4.27(0.99) / 4.11(0.91)	1–5 / 1–5	−1.37 / -0.98	1.33 / 0.92
Item 4	1 / 3	1.85(1.23) / 3.06(1.52)	1–5 / 1–5	1.27 / -0.03	0.48 / -1.43
Item 5	3 / 3	2.89(1.23) / 2.62(1.18)	1–5 / 1–5	−0.04 / 0.26	−0.74 / -0.77
Item 6	4 / 4	3.92(0.98) / 3.77(0.95)	1–5 / 1–5	−0.90 / -0.48	0.69 / -0.22
Item 7	4 / 4	3.63(1.10) / 3.55(1.11)	1–5 / 1–5	−0.38 / -0.39	−0.51 / -0.47
Item 8	4 / 3	3.68(1.02) / 3.39(1.07)	1–5 / 1–5	−0.21 / -0.33	−0.60 / -0.40
Item 9	4 / 4	3.77(0.93) / 3.69(1.00)	2–5 / 1–5	−0.33 / -0.47	−0.71 / -0.30
Item 10	4 / 4	3.85(0.95) / 3.62(0.98)	1–5 / 1–5	−0.67 / -0.45	0.24 / -0.07
Item 11	3 / 3	2.69(1.24) / 3.20(1.28)	1–5 / 1–5	0.08 / -0.14	−1.06 / -1.00
Item 12	4 / 3	3.52(0.94) / 3.43(1.01)	2–5 / 1–5	−0.10 / -0.35	−0.87 / -0.25
Item 13	3 / 3	3.01(1.28) / 2.57(1.17)	1–5 / 1–5	0.03 / 0.30	−1.04 / -0.76
Item 14	4 / 4	3.95(0.85) / 4.00(0.88)	2–5 / 1–5	−0.48 / -0.97	−0.37 / 1.36
Item 15	2 / 2	2.21(0.99) / 2.38(1.16)	1–4 / 1–5	0.16 / 0.58	−1.14 / -0.52
Item 16	4 / 4	4.13(0.86) / 3.96(0.85)	2–5 / 1–5	−0.90 / -0.66	0.39 / 0.47
Item 17	3 / 2	2.83(1.10) / 2.33(1.06)	1–5 / 1–5	−0.06 / 0.26	−0.28 / -0.87
Item 18	4 / 3	3.61(1.12) / 3.28(1.16)	1–5 / 1–5	−0.28 / -0.30	−0.81 / -0.68
Item 19	4 / 4	4.12(1.01) / 3.76(0.89)	1–5 / 1–5	−1.04 / -0.51	0.51 / 0.29
Item 20	4 / 4	4.26(0.85) / 4.09(0.85)	1–5 / 1–5	−1.50 / -0.70	3.15 / 0.24

Note. Mean(SD) = Mean (Standard-deviation).

significantly different between the two samples (Cohen & Cohen, 1983). However, the results showed that the Affective and Cognitive dimensions for the two samples were not significantly different from one another ($z = -1.84$, $p = 0.07$).

Table 5 displays Cronbach's alphas, omega coefficients, mean inter-item correlations, and corrected item-total correlation range for the BES for both samples. Values for the Cronbach's alphas and omegas can be considered satisfactory and these were consistent across the two samples. For the most part, the mean inter-item correlation and the corrected item to total correlations confirmed what was found in terms of internal consistency suggesting good internal reliability. Consistent with the factor loadings reported above, the removal of items 4 and 15 improved the results obtained in terms of internal consistency.

The criterion validity of the BES with the YPI, the ICU, the RPQ, the SDRS-5, and CD symptoms are presented in Table 6 separately for the forensic and school samples. We discuss these associations for the forensic and school samples separately. For the forensic sample, a significant negative correlation emerged between the YPI Callous-Unemotional dimension and the BES Total score. This is not too surprising given the conceptual overlap between CU traits and lacking empathy. Somewhat unexpectedly, however, the BES Total score was positively correlated with scores on the RPQ Total and RPQ Reactive as

well as CD symptoms. In terms of the BES subscales, the Affective dimension was negatively correlated with the Callous-Unemotional dimension of the YPI and positively correlated with the RPQ Reactive scale. There were positive correlations between the YPI Total, Grandiose-Manipulative, and Impulsive-Irresponsible scales and the BES Cognitive subscale. Additionally, the BES Cognitive subscale showed positive correlations with the ICU Uncaring subscale, the RPQ Total, and CD symptoms.

For the school sample the correlations between the BES total and the dimensions of the YPI were significant and negative with the exception of the Impulsive-Irresponsible scale which was unrelated to the total BES score. The Callous-Unemotional dimension showed the strongest

Table 2
Goodness of fit indices for the different models of the BES.

	S-Bχ²/df	IFI	CFI	RMSEA(90% CI)	AIC
Forensic sample					
BES 1-factor	4.71	0.57	0.57	0.19(0.18–0.20)	461.85
BES 2-factor[a]	2.40	0.84	0.84	0.12(0.10–0.13)	65.93
BES 2-factor (i4, i15)[b]	2.53	0.87	0.86	0.12(0.11–0.14)	68.96
BES 2-factor (16i)[c]	2.85	0.84	0.84	0.14(0.12–0.15)	85.19
School sample					
BES 1-factor	11.32	0.76	0.76	0.20(0.19–0.21)	1585.75
BES 2-factor[d]	2.83	0.96	0.96	0.08(0.07–0.09)	138.18

S-Bχ² = Satorra-Bentler chi-square; df = degrees of freedom; IFI = Incremental Fit Index; CFI = Comparative Fit Index; RMSEA (90% CI) = Root Mean Square Error of Approximation (90% confidence interval); AIC = Akaike Information Criterion; ML = Maximum Likelihood; BES 2-factor (i4, i15) = items 4 and 15 excluded due to loadings below 0.30; BES 2-factor (16i) = BES 16 items version adapted by Anastacio et al. (2016).
[a] Included error covariance between items 3 and 14, 1 and 6, 5 and 16.
[b] Included error covariance between items 3 and 14, 6 and 1, 5 and 16.
[c] Included error covariance between items 9 and 12, 3 and 12, 3 and 14.
[d] Included error covariance between items 5 and 17, 15 and 17, 6 and 10, 9 and 10.

Table 3
Item loadings for the confirmatory two-factor robust structure of the BES.

BES items	Factor 1 F/S	Factor 2 F/S
Affective dimension		
1. My friend's emotions don't affect me much.	0.48/0.74	
2. After being with a friend who is sad about something I […].	0.31/0.76	
4. I get frightened when I watch characters in a good scary movie.	−−/0.31	
5. I get caught up in other people's feelings easily.	0.52/0.45	
7. I don't become sad when I see other people crying.	0.75/0.83	
8. Other people's feelings don't bother me at all.	0.87/0.84	
11. I often become sad when watching sad things on TV or […].	0.34/0.58	
13. Seeing a person who has been angered has no effect on […].	0.42/0.41	
15. I tend to feel scared when I am with friends who are afraid.	−−/0.51	
17. I often get swept up in my friend's feelings.	0.60/0.47	
18. My friend's unhappiness doesn't make me feel anything.	0.67/0.65	
Cognitive dimension		
3. I can understand my friend's happiness when she/he […].		0.58/0.76
6. I find it hard to know when my friends are frightened.		0.56/0.77
9. When someone is feeling 'down' I can usually understand […].		0.64/0.74
10. I can usually work out when my friends are scared.		0.63/0.75
12. I can often understand how people are feeling even […].		0.70/0.75
14. I can usually work out when people are cheerful.		0.70/0.86
16. I can usually realize quickly when a friend is angry.		0.90/0.81
19. I am not usually aware of my friend's feelings.		0.85/0.85
20. I have trouble figuring out when my friends are happy.		0.77/0.86

Note. BES = Basic Empathy Scale; F/S = forensic/school samples.

Table 4
Correlations matrixes for the forensic and school samples.

	BES total	BES Affective	BES Cognitive
Forensic sample			
BES total	1		
BES Affective	0.79***	1	
BES Cognitive	0.74***	.16^{ns}	1
School sample			
BES total	1		
BES Affective	0.86***	1	
BES Cognitive	0.79***	0.36***	1

Note. BES = Basic Empathy Scale; BES Affective = BES Affective dimension; BES Cognitive = BES Cognitive dimension.
*** significant at the 0.001 level; ns = non-significant.

negative correlation with the BES. At the BES dimension level, the Affective scale was negatively correlated with all of the YPI scales with the exception of the Impulsive-Irresponsible dimension, to which it was unrelated. The Affective dimension was also negatively correlated with the ICU total and all three of its subscales. The Cognitive dimension showed the same pattern of correlations with the ICU and its subscales as the Affective dimension; however, the only correlation that emerged as significant with regard to the YPI was the YPI Callous-Unemotional subscale (negatively). Interestingly, for the school sample the BES was unrelated to the RPQ and CD symptoms.

Comparisons between the forensic and school samples revealed no significant differences between BES total scores ($F = 1.736$, $p = 0.19$, $M_{forensic} = 69.96$, $SD_{forensic} = 9.47$, $M_{school} = 68.22$, $SD_{school} = 12.06$), Affective empathy scores ($F = 0.167$, $p = 0.68$, $M_{forensic} = 34.17$, $SD_{forensic} = 6.83$, $M_{school} = 33.80$, $SD_{school} = 7.96$), and Cognitive empathy scores ($F = 3.436$, $p = 0.07$, $M_{forensic} = 35.80$, $SD_{forensic} = 5.76$, $M_{school} = 34.42$, $SD_{school} = 6.66$). Mean affective empathy scores from school samples of both English (Jolliffe & Farrington, 2006) and French (D'Ambrosio et al., 2009) female adolescents were markedly higher (English: $M = 40.3$, $SD = 5.8$; French: $M = 41.90$, $SD = 4.71$) to those reported in both our forensic and school samples ($d = 0.93–1.32$); however, there was little difference in cognitive empathy scores (English: $M = 35.0$, $SD = 3.9$; French: $M = 36.17$, $SD = 3.94$; $d = 0.07–0.31$).

4. Discussion

The primary aim of this study was to assess the psychometric properties of the BES among school-based and forensic samples of Portuguese female adolescents. In order to examine the construct validity of the BES we tested the factor-structure of the BES among both samples using confirmatory factor analysis. The internal consistency of the BES and its dimensions were tested using conventional measures (e.g., Cronbach's alpha, mean inter-item correlations, item-total correlations). Finally, we examined correlations for the BES total and its dimension with measures of theoretically relevant external criteria

Table 5
Cronbach's alphas, omega coefficients, mean inter-item correlations, and corrected item-total correlation range.

	Alpha	Omega	MIIC	CITCR
Forensic sample				
BES total	0.82	0.88	0.21	0.21–0.67
BES Affective	0.81	0.82	0.32	0.30–0.73
BES Cognitive	0.86	0.90	0.41	0.37–0.76
School sample				
BES total	0.90	0.93	0.31	0.22–0.66
BES Affective	0.85	0.86	0.34	0.30–0.69
BES Cognitive	0.93	0.94	0.59	0.63–0.80

Note. BES = Basic Empathy Scale; Alpha = Cronbach's alpha; Omega = Omega coefficient; MIIC = Mean inter-item correlation; CITCR = Corrected item-total correlation range.

(i.e., psychopathy, aggression, and conduct disorder symptoms) across the two distinct samples. The findings regarding these aims are discussed below within the context of prior research along with limitations of the current study and implications for future research.

To start, we examined the factor structure of the BES, comparing the one-factor model against the two-factor model in which the latter consisted of cognitive and affective aspects of empathy. In general, and consistent with prior research (e.g., Jolliffe & Farrington, 2006), we found considerable support for the two factor model of the BES. This was particularly true for the school sample. It is important to point out that the two-factor model did not achieve adequate fit without including error covariance between items 3 and 14, 1 and 6, as well as items 5 and 16. Alternatively, however, the two-factor model, despite being the best fitting model, failed to meet conventional cutoffs for model fit indices for the forensic sample. The lack of fit, however, could be attributed to the small sample size for the forensic sample. Thus, the BES seems best to capture empathy among female juveniles as a multidimensional construct measuring both the affective and cognitive dimensions. The multidimensional nature of the BES is also supported given the inter-correlations between the Affective and Cognitive dimensions. That is, the two scales did show some overlap for the school sample ($r = 0.36$) suggesting that they are in fact measuring distinct constructs. However, consistent with other recent work in a forensic sample (Kahn, Frick, Golmaryami, & Marsee, 2016) the two scales were unrelated in the forensic sample indicating there may be a greater divergence in these distinct components of empathy across detained samples.

Across the different indices for assessing internal reliability, the BES and its dimensions showed good to excellent reliability for both the school and forensic samples (Nunnally & Bernstein, 1994). These estimates of reliability are actually higher than those reported in previous research with mixed-gender samples (e.g., Albiero et al., 2009; Salas-Wright et al., 2012). It is also notable that the mean inter-item correlations and corrected item-total correlations for the Affective scale among the school sample were outside the recommended cutoffs (i.e., higher) suggesting that the items may be too homogeneous and not capturing a broad range of the construct (Clark & Watson, 1995). Future research, therefore, should conduct more item-level analyses (e.g., Item Response Theory analysis) in order to better determine the coverage of the construct given the items included on the BES, particularly for the Affective scale among females.

Finally, several interesting findings emerged with regard to the criterion validity (Kaplan & Saccuzzo, 2013) of the BES and its association with measures of psychopathic traits, aggression, social desirability and CD symptoms. Perhaps the most consistent finding was the negative association between the YPI Callous-Unemotional scale and the BES total scale and its dimensions (with the exception of the Cognitive scale among the forensic sample). This is not too surprising given that empathy tends to be most conceptually related to this aspect of psychopathy (e.g., lack of remorse, lack of guilt, and limited affective emotions). This association was bolstered by the consistent negative association between the ICU – a more comprehensive measure of the Callous-Unemotional aspects of psychopathy – and the BES among the school sample. This is largely consistent with prior research that has examined the association between the BES and the ICU in community or school samples (e.g., Muñoz, Qualter, & Padgett, 2011). It is possible that the lack of significant correlations for the forensic sample is due, in part, to low power among the forensic sample. On the other hand, the positive association between Cognitive and Uncaring scales found among the forensic sample is consistent with the notion that youth with CU traits may have the ability to understand and recognize (but not necessarily experience) others' emotions (Dadds et al., 2009). The BES did not present significant correlations with the SDRS-5 measure of social desirability, suggesting that the responses to the BES were not influenced by the adolescent's desire to appear more empathic than they actually were. This is consistent with previous research (e.g., Davis, 1983; Jolliffe & Farrington, 2006) and contributes to the

Table 6
Correlations with other measures and variables for the forensic and school samples.

	BES Forensic/school	BES affective Forensic/school	BES cognitive Forensic/school
YPI Total	$0.06^{ns}/-0.20^{**}$	$-0.17^{ns}/-0.23^{***}$	$0.27^{**}/-0.08^{ns}$
YPI Grandiose-Manipulative	$0.17^{ns}/-0.17^{**}$	$-0.07^{ns}/-0.16^{**}$	$0.33^{**}/-0.11^{ns}$
YPI Callous-Unemotional	$-0.23^{*}/-0.41^{***}$	$-0.38^{***}/-0.43^{***}$	$0.07^{ns}/-0.22^{***}$
YPI Impulsive-Irresponsible	$0.11^{ns}/0.04^{ns}$	$-0.07^{ns}/-.03^{ns}$	$0.22^{*}/0.11^{ns}$
ICU total	$0.02^{ns}/-0.41^{***}$	$0.03^{ns}/-0.32^{***}$	$0.01^{ns}/-0.35^{***}$
ICU Callousness	$0.04^{ns}/-0.22^{***}$	$0.09^{ns}/-0.22^{***}$	$-0.04^{ns}/-0.14^{*}$
ICU Uncaring	$0.04^{ns}/-0.43^{***}$	$-0.13^{ns}/-0.24^{***}$	$0.20^{*}/-0.49^{***}$
ICU Unemotional	$-0.04^{ns}/-0.31^{***}$	$0.07^{ns}/-0.30^{***}$	$-0.15^{ns}/-0.20^{**}$
RPQ total	$0.21^{*}/0.09^{ns}$	$0.15^{ns}/-0.08^{ns}$	$0.20^{*}/-0.08^{ns}$
RPQ Reactive	$0.21^{*}/-0.10^{ns}$	$0.20^{*}/-0.07^{ns}$	$0.17^{ns}/-0.10^{ns}$
RPQ Proactive	$0.17^{ns}/-0.05^{ns}$	$0.08^{ns}/-0.06^{ns}$	$0.18^{ns}/-0.03^{ns}$
SDRS-5	$0.13^{ns}/0.03^{ns}$	$0.15^{ns}/0.06^{ns}$	$0.03^{ns}/-0.02^{ns}$
CD symptoms	$0.30^{**}/0.04^{ns}$	$0.14^{ns}/0.02^{ns}$	$0.32^{**}/0.04^{ns}$

BES = Basic Empathy Scale; YPI = Youth Psychopathic Traits Inventory; ICU = Inventory of Callous-Unemotional Traits; RPQ = Reactive-Proactive Aggression Questionnaire; SDRS-5 = Social Desirability Response Set – 5; CD symptoms = Conduct Disorder symptoms score as a scale.
*** Significant at the 0.001 level.
** Significant at the 0.01 level.
* Significant at the 0.05 level; ns = non-significant.

literature that examines possible relationships between the BES and different measures of social desirability.

With regard to aggression and CD symptoms, only significant associations were found within the forensic sample. Specifically, positive associations between the BES and the RPQ as well as the BES and CD symptoms emerged. One possible interpretation of the positive correlations found in the current study is that girls with heightened levels of aggression (particularly reactive aggression) are also highly emotionally responsive and, in turn, more empathic. Despite this possibility, these findings are inconsistent with prior research that typically finds negative associations between measures of empathy and aggression among adolescents (e.g., Lovett & Sheffield, 2007; Vachon, Lynam, & Johnson, 2014). It is possible, considering that past research has focused mainly on samples of boys or mixed samples, that the association between aggressive behaviors and empathy is distinct for female adolescents, particularly those involved in the juvenile justice system. The prevalence of CD found in the current forensic sample (85.4%) was higher than those typically found among forensic samples composed of female youths (Sevecke & Kosson, 2010). Thus, future research is needed that more thoroughly examines the unique association between aggression and the different components of empathy among females.

Despite the unique contribution of the current study by examining the psychometric properties of the BES among a school-based and forensic sample of females, the findings should be considered in light of several limitations. First, the cross-sectional nature of the current study prevented the evaluation of test-retest reliability of the BES as well as evaluating its predictive utility. Second, the current study did not include other validated measures of empathy to assess its convergent validity because no other measures of empathy have been validated among Portuguese youth, nor did it include various other measures of constructs to assess criterion validity (e.g., sympathy, guilt, emotional intelligence, etc.). Third, due to the small size of the forensic sample, we were prevented from conducting formal tests of model fit between the two samples. Additionally, the small sample size may have resulted in reduced power which, in turn, may have attenuated many of the associations with external criterion for this sample. Finally, because most measures included in the current study were based on self-report, the current study suffers from the various limitations associated with such measures, including that of response distortion such as patterns of socially desirable responding.

These limitations aside, the current study provides a unique examination of the BES by evaluating its factor-structure, reliability, and criterion validity among two unique samples of females. In many ways the

current study suggests that the BES is a valid, multidimensional measure of empathy which can be used among samples of females, particularly school-based samples of females. However, the current study does raise some questions about its use among forensic samples of females, particularly with regard to its factor structure and criterion validity. Thus, further research examining the psychometric properties of the BES among female samples is needed.

References

Albiero, P., Matricardi, G., Speltri, D., & Toso, D. (2009). The assessment of empathy in adolescence: A contribution to the Italian validation of the Basic Empathy Scale. *Journal of Adolescence*, 32, 393–408. https://doi.org/10.1016/j.adolescence.2008.01.001.

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Anastacio, S., Vagos, P., Nobre-Lima, L., Rijo, D., & Jolliffe, D. (2016). The Portuguese version of the Basic Empathy Scale (BES): Dimensionality and measurement invariance in a community adolescent sample. *European Journal of Developmental Psychology*. https://doi.org/10.1080/17405629.2016.1167681 (Advanced Online Publication).

Andershed, H., Gustafson, S., Kerr, M., & Stattin, H. (2002). The usefulness of self-reported psychopathy-like traits in the study of antisocial behavior among non-referred adolescents. *European Journal of Personality*, 16, 383–402. https://doi.org/10.1002/per.455.

Andershed, H., Kerr, M., Stattin, H., & Levander, S. (2002). Psychopathic traits in non-referred youths: Initial test of a new assessment tool. In E. Blaauw, & L. Sheridan (Eds.), *Psychopaths: Current international perspectives* (pp. 131–158). Haag: Elsevier.

Ang, R., & Goh, D. (2010). Cyberbullying among adolescents: The role of affective and cognitive empathy, and gender. *Child Psychiatry & Human Development*, 41, 387–397. https://doi.org/10.1007/s10578-010-0176-3.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: an investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, 34, 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00.

Bentler, P., & Wu, E. (2015). *Supplement to EQS for Windows user's guide.* Temple City, CA: Multivariate Software, Inc.

Blair, R. J. R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition*, 14, 698–718. https://doi.org/10.1016/j.concog.2005.06.004.

Broidy, L., Cauffman, E., Espelage, D. L., Mazerolle, P., & Piquero, A. (2003). Sex differences in empathy and its relation to juvenile offending. *Violence and Victims*, 18, 503–516. https://doi.org/10.1891/vivi.2003.18.5.503.

Byrne, B. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming.* Mahwah, NJ: Lawrence Erlbaum Associates.

Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: a meta-analytic review. *Psychological Bulletin*, 139, 735–765. https://doi.org/10.1037/a0030737.

Clark, L., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309–319. https://doi.org/10.1037/1040-3590.7.3.309.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cooke, D., & Michie, C. (2001). Refining the construct of psychopathy: Towards a hierarchical model. *Psychological Assessment*, 13, 171–188. https://doi.org/10.1037/1040-3590.13.2.171.

Crowne, D., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24, 349–354. https://doi.org/10.1037/h0047358.

Dadds, M. R., Hawes, D. J., Frost, A. D., Vassallo, S., Bunn, P., Hunter, K., & Merz, S. (2009). Learning to 'talk the talk': the relationship of psychopathic traits to deficits in empathy across childhood. *Journal of Child Psychology and Psychiatry*, 50, 599–606. https://doi.org/10.1111/j.1469-7610.2008.02058.x.

D'Ambrosio, F., Olivier, M., Didon, D., & Besche, C. (2009). The Basic Empathy Scale: A French validation of a measure of empathy in youth. *Personality and Individual Differences*, 46, 160–165. https://doi.org/10.1016/j.paid.2008.09.020.

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, 10, 1–85 (Retrieved from http://www.uv.es/~friasnav/Davis_1980.pdf).

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, 44, 113–126. https://doi.org/10.1037/0022-3514.44.1.113.

Decety, J. (2010). The neurodevelopment of empathy in humans. *Developmental Neuroscience*, 32, 257–267. https://doi.org/10.1159/000317771.

Decety, J., & Jackson, P. L. (2004). The functional architecture of human empathy. *Behavioral and Cognitive Neuroscience Reviews*, 3, 71–100. https://doi.org/10.1177/1534582304267187.

Decety, J., & Michalska, K. (2010). Neurodevelopmental changes in the circuits underlying empathy and sympathy from childhood to adulthood. *Developmental Science*, 13, 886–899. https://doi.org/10.1111/j.1467-7687.2009.00940.x.

Eisenberg, N., & Eggum, N. D. (2009). Empathic responding: Sympathy and personal distress. In J. Decety, & W. Ickes (Eds.), *The social neuroscience of empathy* (pp. 71–83). Cambridge: MIT Press.

Eisenberg, N., Guthrie, I. K., Murphy, B. C., Shepard, S. A., Cumberland, A., & Carlo, G. (1999). Consistency and development of prosocial dispositions: A longitudinal study. *Child Development*, 70, 1360–1372. https://doi.org/10.1111/1467-8624.00100.

Eisenberg, N., Shea, C. L., Carlo, G., & Knight, G. P. (1991). Empathy-related responding and cognition: A chicken and the egg dilemma. In W. M. Kurtines (Ed.), *Handbook of moral behavior and development* (pp. 63–88). Hillsdale: Erlbaum.

Essau, C., Sasagawa, S., & Frick, P. (2006). Callous-unemotional traits in a community sample of adolescents. *Assessment*, 13, 454–469. https://doi.org/10.1177/1073191106287354.

Forouzan, E., & Cooke, D. J. (2005). Figuring out la femme fatale: Conceptual and assessment issues concerning psychopathy in females. *Behavioral Sciences and the Law*, 23, 765–778. https://doi.org/10.1002/bsl.669.

Frick, P., & Hare, R. (2001). *The Antisocial Process Screening Device (APSD): Technical manual.* Toronto: Multi-Health Systems.

Geng, Y., Xia, D., & Qin, B. (2012). The Basic Empathy Scale: A Chinese validation of a measure of empathy in adolescents. *Child Psychiatry & Human Development*, 43, 499–510. https://doi.org/10.1007/s10578-011-0278-6.

Hambleton, R., Merenda, P., & Spielberger, C. (2005). *Adapting educational and psychological tests for cross-cultural assessment.* Mahwah, NJ: Lawrence Erlbaum Associates.

Hays, R., Hayashi, T., & Stewart, A. (1989). A five-item measure of Socially Desirable Response Set. *Educational and Psychological Measurement*, 49, 629–636. https://doi.org/10.1177/001316448904900315.

Hoffman, M. L. (1987). The contribution of empathy to justice and moral judgment. In N. Eisenberg, & J. Strayer (Eds.), *Empathy and its development* (pp. 47–80). Cambridge: Cambridge University Press.

Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33, 307–316. https://doi.org/10.1037/h0027580.

IBM Corp (2016). *IBM SPSS Statistics for Windows (version 24).* Armonk, NY: Author.

Jolliffe, D., & Farrington, D. (2006). Development and validation of the Basic Empathy Scale. *Journal of Adolescence*, 29, 589–611. https://doi.org/10.1016/j.adolescence.2005.08.010.

Kahn, R. E., Frick, P. J., Golmaryami, F. N., & Marsee, M. A. (2016). The moderating role of anxiety in the associations of callous-unemotional traits with self-report and laboratory measures of affective and cognitive empathy. *Journal of Abnormal Child Psychology.* https://doi.org/10.1007/s10802-016-0179-z (Advance Online Publication).

Kaplan, R., & Saccuzzo, D. (2013). *Psychological testing: Principles, applications, and issues* (8th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Karnik, N. S., Soller, M., Redlich, A., Silverman, M., Kraemer, H. C., Haapanen, R., & Steiner, H. (2009). Prevalence of and gender differences in psychiatric disorders among juvenile delinquents incarcerated for nine months. *Psychiatric Services (Washington, D.C.)*, 60, 838–841. https://doi.org/10.1176/ps.2009.60.6.838.

Katsiyannis, A., Zhang, D., Barrett, D., & Flaska, T. (2004). Background and psychosocial variables associated with recidivism among adolescent males: A 3-year investigation. *Journal of Emotional and Behavioral Disorders*, 12(1), 23–29. https://doi.org/10.1177/10634266040120010301.

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Muñoz, L. C., & Morris, A. S. (2008). Assessing callous–unemotional traits in adolescent offenders: Validation of the Inventory of Callous–Unemotional Traits. *International Journal of Law and Psychiatry*, 31, 241–252. https://doi.org/10.1016/j.ijlp.2008.04.002.

Lahey, B. B., Schwab-Stone, M., Goodman, S. H., Waldman, I. D., Canino, G., Rathouz, P. J., & Jensen, P. S. (2000). Age and gender differences in oppositional behavior and conduct problems: A cross-sectional household study of middle childhood and adolescence. *Journal of Abnormal Psychology*, 109, 488–503. https://doi.org/10.1037//0021-843X.109.3.488.

Lam, C. B., Solmeyer, A. R., & McHale, S. M. (2012). Sibling relationships and empathy across the transition to adolescence. *Journal of Youth and Adolescence*, 41, 1657–1670. https://doi.org/10.1007/s10964-012-9781-8.

Leech, N., Barrett, K., & Morgan, G. (2015). *IBM SPSS for intermediate statistics: Use and interpretation* (5th ed.). New York: Routledge.

Lovett, B. J., & Sheffield, R. A. (2007). Affective empathy deficits in aggressive children and adolescents: A critical review. *Clinical Psychology Review*, 27, 1–13. https://doi.org/10.1016/j.cpr.2006.03.003.

Maroco, J. (2014). *Análise de Equações Estruturais: Fundamentos teóricos, software & aplicações [Structural equations analysis: Theoretical foundations, software and applications].* Pero Pinheiro: ReportNumber, Ltd.

McClure, E. B. (2000). A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychological Bulletin*, 126, 424–453. https://doi.org/10.1037/0033-2909.126.3.424.

Mehrabian, A., & Epstein, N. (1972). A measure of emotional empathy. *Journal of Personality*, 40, 525–543. https://doi.org/10.1111/j.1467-6494.1972.tb00078.x.

Michalska, K. J., Kinzler, K. D., & Decety, J. (2013). Age-related sex differences in explicit measures of empathy do not predict brain responses across childhood and adolescence. *Developmental Cognitive Neuroscience*, 3, 22–32. https://doi.org/10.1016/j.dcn.2012.08.001.

Moffitt, T. E., & Caspi, A. (2001). Childhood predictors differentiate life-course persistent and adolescence-limited antisocial pathways among males and females. *Development and Psychopathology*, 13, 355–375. https://doi.org/10.1017/S0954579401002097.

Moffitt, T. E., Caspi, A., Rutter, M., & Silva, P. A. (2001). *Sex differences in antisocial behavior.* New York: Cambridge University Press.

Moretti, M. M., Catchpole, R. E., & Odgers, C. (2005). The dark side of girlhood: Recent trends, risk factors and trajectories to aggression and violence. *The Canadian Child and Adolescent Psychiatry Review*, 14, 21–25.

Muñoz, L. C., Qualter, P., & Padgett, G. (2011). Empathy and bullying: Exploring the influence of callous-unemotional traits. *Child Psychiatry & Human Development*, 42, 183–196. https://doi.org/10.1007/s10578-010-0206-1.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Pajer, K. (1998). What happens to "bad" girls? A review of the adult outcomes of antisocial adolescent girls. *The American Journal of Psychiatry*, 155, 862–870. https://doi.org/10.1176/ajp.155.7.862.

Pechorro, P., Andershed, H., Ray, J., Maroco, J., & Gonçalves, R. (2015). Validation of the Youth Psychopathic Traits Inventory and Youth Psychopathic Traits Inventory – Short among incarcerated juvenile delinquents. *Journal of Psychopathology and Behavioral Assessment*, 37, 576–586. https://doi.org/10.1007/s10862-015-9490-1.

Pechorro, P., Ayala-Nunes, L., Nunes, C., Oliveira, J. P., & Gonçalves, R. (2016). Psychometric properties of the Socially Desirable Response Set-5 among incarcerated male and female

juvenile offenders. *International Journal of Law and Psychiatry*, 49, 17–21. https://doi.org/10.1016/j.ijlp.2016.05.003.

Pechorro, P., Hawes, S., Gonçalves, R., & Ray, J. (2017). Psychometric properties of the Inventory of Callous-Unemotional Traits short version (ICU-12) among detained female juvenile offenders and community youths. *Psychology, Crime & Law*, 23, 221–239. https://doi.org/10.1080/1068316X.2016.1239724.

Pechorro, P., Hidalgo, V., Nunes, C., & Jiménez, L. (2016). Confirmatory factor analysis of the Antisocial Process Screening Device: Self-Report among incarcerated male juvenile offenders. *International Journal of Offender Therapy and Comparative Criminology*, 60, 1856–1872. https://doi.org/10.1177/0306624X15588903.

Pechorro, P., Kahn, R., Ray, J., Raine, A., & Gonçalves, R. A. (2017). Psychometric properties of the Reactive-Proactive Aggression Questionnaire among detained female juvenile delinquents and community youths. *Criminal Justice and Behavior*, 44, 531–550. https://doi.org/10.1177/0093854816686395.

Pechorro, P., Ray, J., Barroso, R., Maroco, J., & Gonçalves, R. (2016). Validation of the Inventory of Callous-Unemotional Traits among a Portuguese sample of detained juvenile offenders. *International Journal of Offender Therapy and Comparative Criminology*, 60, 349–365. https://doi.org/10.1177/0306624X14551256.

Pechorro, P., Ray, J., Raine, A., Maroco, J., & Gonçalves, R. (2017). The Reactive-Proactive Aggression Questionnaire: Validation among a Portuguese sample of incarcerated juvenile delinquents. *Journal of Interpersonal Violence*, 32, 1995–2017. https://doi.org/10.1177/0886260515590784.

Pechorro, P., Ray, J., Salas-Wright, C., Maroco, J., & Gonçalves, R. (2015). Adaptation of the Basic Empathy Scale among a Portuguese sample of incarcerated juvenile offenders. *Psychology, Crime & Law*, 21, 699–714. https://doi.org/10.1080/1068316X.2015.1028546.

Pechorro, P., Ribeiro da Silva, D., Andershed, H., Rijo, D., & Gonçalves, R. A. (2016). The Youth Psychopathic Traits Inventory: Measurement invariance and psychometric properties among Portuguese youths. *International Journal of Environmental Research and Public Health*, 13, 852. https://doi.org/10.3390/ijerph13090852.

Raine, A., Dodge, K., Loeber, R., Gatzke-Kopp, L., Lynam, D., Reynolds, C., & Liu, J. (2006). The Reactive-Proactive Aggression Questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. *Aggressive Behavior*, 32, 159–171. https://doi.org/10.1002/ab.20115.

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

Reynolds, W. (1982). Development of reliable and valid short forms of the Marlowe–Crowne Social Desirability Scale. *Journal of Clinical Psychology*, 38(1), 119–125. https://doi.org/10.1002/1097-4679(198201)38:1%3C119::AIDJCLP2270380118%3E3.0.CO;2-I.

Robbins, P. C., Monahan, J., & Silver, E. (2003). Mental disorder, violence, and gender. *Law and Human Behavior*, 27, 561–571. https://doi.org/10.1023/B:LAHU.0000004886.13268.f2.

Rutter, M., Giller, H., & Hagell, A. (1998). *Antisocial Behaviour in Young People.* Cambridge, England: Cambridge University Press.

Salas-Wright, C., Olate, R., & Vaughn, M. (2012). Assessing empathy in Salvadoran high-risk and gang-involved adolescents and young adults: A Spanish validation of the Basic Empathy Scale. *International Journal of Offender Therapy and Comparative Criminology*, 57, 1393–1416. https://doi.org/10.1177/0306624X12455170.

Schnittker, J., & John, A. (2007). Enduring Stigma : The Long-Term Effects of Incarceration on Health. *Journal of Health and Social Behavior*, 48, 115–130. https://doi.org/10.1177/002214650704800202.

Sevecke, K., & Kosson, D. (2010). Relationships of child and adolescent psychopathy to other forms of psychopathology. In R. Salekin, & D. Lynam (Eds.), *Handbook of child and adolescent psychopathy* (pp. 284–314). New York: The Guilford Press.

Simões, M. (1994). *Investigação no âmbito da aferição nacional ao Teste das Matrizes Progressivas Coloridas de Raven* [Researches regarding the national validation of Raven's Coloured Progressive Matrices Test] (Unpublished doctoral dissertation). Faculdade de Psicologia e Ciências da Educação da Universidade de Coimbra, Coimbra, Portugal.

Singer, T., & Lamm, C. (2009). The social neuroscience of empathy. *The Year in Cognitive Neuroscience*, 1156, 81–96. https://doi.org/10.1111/j.1749-6632.2009.04418.x.

Skilling, T., Quinsey, V., & Craig, W. (2001). Evidence of a taxon underlying serious antisocial behavior in boys. *Criminal Justice and Behavior*, 28, 450–470. https://doi.org/10.1177/009385480102800404.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L, Dekker, J., et al. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60, 34–42. https://doi.org/10.1016/j.jclinepi.2006.03.012.

Tracy, P. E., Kempf-Leonard, K., & Abramoske-James, S. (2009). Gender differences in delinquency and juvenile justice processing: Evidence from national data. *Crime & Delinquency*, 55, 171–215. https://doi.org/10.1177/0011128708330628.

Tzoumakis, S., Lussier, P., & Corrado, R. (2012). Female juvenile delinquency, motherhood, and the intergenerational transmission of aggression and antisocial behavior. *Behavioral Sciences & the Law*, 30, 211–237. https://doi.org/10.1002/bsl.2010.

Vachon, D., Lynam, D., & Johnson, J. (2014). The (non)relation between empathy and aggression: Surprising results from a meta-analysis. *Psychological Bulletin*, 140, 751–773. https://doi.org/10.1037/a0035236.

van Langen, M. A. M., Wissink, I. B., van Vugt, E. S., van der Stouwe, T., & Stams, G. J. J. M. (2014). The relation between empathy and offending: A meta-analysis. *Aggression and Violent Behavior*, 19, 179–189. https://doi.org/10.1016/j.avb.2014.02.003.

West, S., Taylor, A., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York: The Guilford Press.