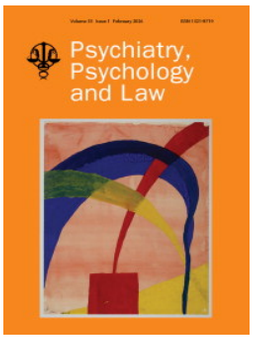

Psychiatry, Psychology and Law

ISSN: 1321-8719 (Print) 1934-1687 (Online) Journal homepage: www.tandfonline.com/journals/tppl20

Measuring the callous-unemotional traits in Spanish juvenile offenders: properties and functioning of the INCA questionnaire in this population

José M. Casas, Jorge-Manuel Dueñas, Pere J. Ferrando, Elena Castarlenas, Andreu Vigil-Colet, Juan C. Hernández-Navarro & Fabia Morales-Vives

To cite this article: José M. Casas, Jorge-Manuel Dueñas, Pere J. Ferrando, Elena Castarlenas, Andreu Vigil-Colet, Juan C. Hernández-Navarro & Fabia Morales-Vives (25 May 2025): Measuring the callous-unemotional traits in Spanish juvenile offenders: properties and functioning of the INCA questionnaire in this population, Psychiatry, Psychology and Law, DOI: 10.1080/13218719.2025.2497785

To link to this article: https://doi.org/10.1080/13218719.2025.2497785

Published online: 25 May 2025.

Psychiatry, Psychology and Law, 2025
Vol. 0, No. 0, 1–22, https://dx.doi.org/10.1080/13218719.2025.2497785

Measuring the callous-unemotional traits in Spanish juvenile offenders: properties and functioning of the INCA questionnaire in this population

José M. Casas[a], Jorge-Manuel Dueñas[a,b], Pere J. Ferrando[a,b], Elena Castarlenas[a,b], Andreu Vigil-Colet[a,b], Juan C. Hernández-Navarro[a] and Fabia Morales-Vives[a,b]

[a]*Psychology Department, Universitat Rovira i Virgili, Tarragona, Spain;* [b]*Research Center for Behavior Assessment (CRAMC), Tarragona, Spain*

The Inventory of Callous-unemotional Traits and Antisocial Behaviour (INCA) was developed to assess (a) traits precursors of psychopathy and (b) antisocial behaviour in adolescents, but its psychometric properties have only been assessed in community samples. Considering that juvenile offenders are an important target population for this kind of instrument, the current study aims to assess the INCA functioning in this population. Participants were 202 young Spanish offenders, who answered several questionnaires. Factor-analytic solutions, in which response biases were controlled, suggested that there were only three content factors underlying the data. As most of the antisocial behaviour items loaded in the callousness subscale, these items were removed, leading to adequate fit, scalability and scoring results. Callousness and uncaring were related to aggression and impulsivity, as expected. Results highlight the importance of assessing the psychometric properties of questionnaires in the specific populations where the instrument may be highly relevant, such as offenders.

Keywords: antisocial behaviour; callousness-unemotional traits; external validity; measurement properties; psychopathy; structural properties; young offenders.

Article History: Received 13 August 2024; Accepted 30 March 2025

Numerous studies have shown that certain characteristics associated with psychopathy can be observed at a very early age, even as young as 4 years old (e.g. Dadds et al., 2005; Frick et al., 1994; Loeber, 1982), and that these characteristics tend to remain relatively stable through life span (e.g. Frick et al., 2003; Lynam, 1996). An approach for assessing these characteristics in children and adolescents that has attracted a considerable interest in recent years is that based on the so-called 'callousness-unemotional' (CU) traits. According to Frick (2004), the CU traits are potential markers of a possible development of psychopathy, and involve characteristics such as poor emotional expression, lack of empathy, an irresponsible attitude towards self-performance, a lack of guilt and/or regret and manipulation of others. Some authors point to these traits as potential identifiers of a specific subgroup of adolescents with negative patterns of behaviour and psychological adjustment (e.g. López-Romero et al., 2012; Rose et al., 2020). In fact, the CU traits are related to dysfunctional behaviours such as drug-related crime (e.g. Brennan et al., 2022; Taylor & Lang, 2006;

Correspondence: Fabia Morales-Vives, Departamento de Psicología, Universitat Rovira i Virgili, Carretera de Valls s/n, 43007, Tarragona, Spain. E-mail: fabia.morales@urv.cat

Wymbs et al., 2012), aggression and antisocial behaviour (Frick et al., 2005; Frick & White, 2008; Pardini et al., 2003), bullying (Thornberg & Jungert, 2017) and instrumental violence (Fanti et al., 2009; Frick et al., 2005; Frick & White, 2008). Furthermore, children with high levels of CU traits tend to have lower levels of affective empathy, but not of cognitive empathy, which means that they are able to understand what others are experiencing, although this understanding does not evoke an emotional response (Waller & Hyde, 2018).

Considering the relevance of the psychopathic traits in children and adolescents, well-designed instruments aimed at assessing these characteristics in these age groups are clearly needed, and, in fact, several well-known instruments have been developed through the last years, focusing on different models and conceptualisations of psychopathy. For example, the Clinical Assessment of Prosocial Emotions: Version 1.1 (CAPE 1.1; Frick, 2013), based on clinical judgements, allows one to determine whether an individual meets the 'with limited prosocial emotions' specifier of conduct disorder, according to the 5th Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM–5; American Psychiatric Association, APA, 2013). This instrument involves the collection of multiple sources of information, using a structured professional judgement method, and therefore requiring training for a proper administration. Other well-known instruments are based on self-report data, which are useful for screening purposes, such as the Checklist–Youth Version (PCL:YV; Forth et al., 2003), the Youth Psychopathic Traits Inventory (Andershed et al., 2002) and the Antisocial Process Screening Device (APSD; Frick & Hare, 2001). However, as Ray and Wall Myers (2018) pointed out, many of these self-report instruments included the CU traits as facets of broader measures of global psychopathy, with few items specifically assessing these traits and with a limited range of response options. For this reason, Frick (2004) developed the well-known and widely used Inventory of Callous-Unemotional Traits (ICU), being at that time the only questionnaire specifically focused on these traits and the juvenile population. Because Frick (2004) did not provide evidence about the ICU dimensionality, Essau et al. (2006) carried out a study of this type, and proposed the following three-factor structure: callousness, unemotional and uncaring. However, some authors have found problems to replicate this structure, proposing alternative structures with two factors (Houghton et al., 2013) or five factors (Feilhauer et al., 2012). Others had to allow a high number of error terms to correlate to attain adequate goodness-of-fit values in confirmatory factor analytic solutions, and, furthermore, the post hoc freed residuals were based on the modification indices (e.g. Ciucci et al., 2014; Essau et al., 2006; Ezpeleta et al., 2013; Houghton et al., 2013). However, this is an undesirable approach, as it is more data driven than theoretically driven, and is likely to capitalise on chance, leading to unstable solutions that are difficult to replicate in other samples (Ferrando & Lorenzo-Seva, 2000). Morales-Vives et al. (2019) point out certain limitations of the instrument that may explain some of the discrepancies in the results of these studies, such as the fact that some items assigned to different subscales have very similar content.

Considering the limitations of the ICU questionnaire, which have led in some cases to the use of the overall scores instead of the subscales' scores (e.g. Stickle et al., 2009; White et al., 2009), Morales-Vives et al. (2019) developed a new questionnaire aimed at assessing the CU traits, named INventory of Callous-unemotional traits and Antisocial behaviour (INCA). This questionnaire includes the same three subscales as the ICU (callousness, unemotional and uncaring), as well as an additional subscale named antisocial behaviour. The items of this additional subscale refer

to misbehaviours such as violating norms and rules, being a rebel, drawing graffiti, engaging in illegal behaviour that affects others, and disrespecting authority. In addition, this questionnaire controls for both social desirability and acquiescence, which is a novelty compared to previously developed questionnaires. It should be considered that authors such as Campbell et al. (2009) have recommended controlling social desirability in self-reports of this field. More specifically, when developing the INCA, the authors implemented the procedure developed by Ferrando et al. (2009) and Lorenzo-Seva and Ferrando (2009) to control the response biases of acquiescence and social desirability. Therefore, this questionnaire provides scores 'cleaned' of these response biases, which is another difference compared to the ICU questionnaire. The authors split their sample of 719 Spanish participants in two halves and carried out an exploratory factor analysis (EFA) with the first sample and a confirmatory factor analysis (CFA) with the second sample. The parallel analysis (Timmerman & Lorenzo-Seva, 2011) carried out in the first subsample suggested that four common factors underlay the data. The congruence indices (Tucker, 1951) of all items were above .85, which suggests that the factor similarity between the rotated loading matrix and the ideal loading matrix was fair (Lorenzo-Seva & ten Berge, 2006). In fact, each item loads in the expected factor. Furthermore, the results obtained with Bentler's Simplicity index (S; Bentler, 1977) and the Loading Simplicity index (LS; Lorenzo-Seva, 2003) suggest that that the factor simplicity was high, with each item mainly related to only one dimension. Regarding the CFA results carried out with the second subsample, the fit indices results suggested an acceptable fit to the proposed four-factor solution. Equivalent results were obtained in both subsamples, which suggested a good across-sample stability of the factor structure. Similar results were obtained in the EFA carried out with the whole sample. The reliability estimates for the content factor scores ranged between $\rho_{\theta\theta} = .90$ and $\rho_{\theta\theta} = .96$. The unemotional, callous and uncaring subscales of the INCA had the expected correlations with the corresponding subscales of the ICU. Similarly, the INCA subscales most related to impulsivity were callousness and antisocial behaviour, although callousness had a positive, albeit small, correlation with motor impulsivity.

In spite of the good results achieved in Spain, the Mexican adaptation of the INCA raised some problems (Morales-Vives et al., 2020). More specifically, while the unemotional and callousness items loaded on their expected factors, most of the uncaring items and the antisocial behaviour items loaded on a common factor. In view of these results, the authors decided to remove the antisocial behaviour items, prioritising a three-factor structure that was similar to the one previously proposed for the ICU questionnaire. They named the resulting instrument as INventory of CAllous-unemotional traits–Mexico (INCA–M). According to the authors, cultural differences may explain the different results obtained in Spain and Mexico. Although a correct linguistic adaptation of the items was carried out, the fact that the antisocial behaviour items refer to behaviours with low social impact (being a rebel, disrespecting authority, etc.) could have led to them being associated with irresponsibility rather than antisocial behaviour in Mexican community samples, whereas in Spanish community samples these behaviours might have more serious implications. Furthermore, behaviours such as painting graffiti could be seen as a transgression in one community and as a form of social expression in another. If, for cultural reasons, in the Mexico case these behaviours are not perceived as significant transgressions, then it is understandable that this type of item will mostly load on the remaining factors.

After the study by Morales-Vives et al. (2019), no further assessments of the factor structure of the INCA have been carried out either in the general Spanish population or in more specific populations of interest in which

the instrument may be particularly useful. In particular, and given that the INCA assesses precursors of psychopathy, juvenile offenders are a target population of this type. In this respect, we would like to point out that instruments that have been validated in community samples are often used in the juvenile justice field on the assumption that they will work in the same way in this specific population, which may or may not be the case. For this reason, the present study first focuses on determining whether, in a sample of Spanish young offenders, the original four-factor structure of the INCA questionnaire is replicated, or whether the antisocial behaviour subscale presents problems, as was previously the case in the Mexican adaptation. In our view, the tentative explanation we provided for the Mexico results is also expected to operate here. Because most of the antisocial behaviour items do not refer to serious misbehaviours, especially when compared to the type of offence usually committed by young offenders, these items are expected again to overlap with the remaining factors, especially uncaring or callousness. In other words, the fact that these behaviours are not very problematic may mean that in certain countries, or in certain social contexts within the same country, they have fewer implications than in other countries or contexts. In fact, drawing graffiti or disrespecting authority in an educational context such as school may be a red flag to consider, when compared to the usual forms of behaviour in this context. But in a juvenile justice setting these behaviours are not so alarming when compared to the kind of behaviours that give rise to problems with the justice system (e.g. theft, bullying, assault, etc.). For this reason, our working hypothesis is that only three well-defined factors will underly the data in this type of population.

The interest in assessing how the INCA functions in the young offender population is enhanced by the fact that it is the only instrument that assesses the CU traits controlling for social desirability and acquiescence biases. This control has two advantages: (a) it provides scores on the content subscales that are more free from these biases; (b) it also provides a social desirability score and an acquiescence score, which can be very useful for professionals working in the field of juvenile justice. These response bias scores can help them to determine whether the person is responding in a less than credible way, so that they can relativise their responses and contextualise them with information from different sources and backgrounds. Furthermore, at the content level, this questionnaire allows professionals to obtain separate subscale scores for each of the CU traits (callousness, unemotional and uncaring), which is crucial in this area. It is not the same thing that a relatively high score is due to a lack of empathy and insufficient arousal to emotional cues, which some authors call primary psychopathy (e.g. Craig et al., 2021), than to a high level of irresponsibility. Knowing in which subscales a person scores higher on can help practitioners determine the most appropriate approach when implementing intervention programmes.

Besides determining the most appropriate structure, this study also aims to obtain evidence on other properties of the INCA that were not considered in the previous studies by Morales-Vives et al. (2019) and Morales-Vives et al. (2020), regarding the functioning of its items, and the scores derived from the determined structure. More specifically, we aim to assess the scalability of the items and the conditional and marginal reliability of the item response theory (IRT)-based expected a posteriori (EAP) scores, in order to determine at which levels they provide the most accurate measurement in this type of population. Finally, we would like to assess the extent to which the simple subscale sum scores can be considered to be good proxies for the factors they intend to measure. This objective is relevant because, as sum scores are easy to compute and they are familiar to them, many applied professionals prefer to report sum scores rather than more complex schemas

even when acknowledging that they are less accurate. Therefore, it should be considered that, in addition to being the first study that assesses the psychometric properties of the INCA among juvenile offenders, it is also the first study that includes the INCA item analysis results based not only on classical test theory but on item response theory as well.

In terms of external validity evidence, finally, we expect to find significant correlations between the uncaring and callousness factors and motor and non-planning impulsivity, as previous studies found (e.g. Maurer et al., 2021; Morales-Vives et al., 2019), and with direct and indirect aggression, as reported by Kimonis et al. (2008).

Method

Participants

Participants were 202 young offenders aged between 14 and 19 years old ($M = 17.68$; $SD = 1.24$) of which 82.2% were male. It should be considered that the INCA questionnaire was originally developed to assess the CU traits and antisocial behaviour in youth aged up to 19 years old.

According to the Hollingshead Four-Factor Index of Socioeconomic Status (SES, Hollingshead, 1975), 55.4% of the sample had SES levels between 8 and 19, 29.7% had levels between 20 and 29, 10.4% had levels between 30 and 39, 3.5% had levels between 40 and 54, and 1.0% had levels between 55 and 66. The SES is a widely used index to assess the socioeconomic status of an individual's family, taking into account the education, occupation, sex and marital status of the parents. For single-parent families, this index is calculated on the basis of the family member or reference person living with the minor. It should be considered that this information was collected through interviews with the participants (young offenders), so their relatives were not consulted directly.

This sample of young offenders was recruited from open intervention programmes in the province of Tarragona (Catalonia, Spain). These programmes are measures of criminal responsibility that are carried out in the social and family environment of the young person, so they do not involve deprivation of liberty. The sample included the 78.91% of the youth enrolled in these programmes during the time period in which the sample was collected. It was not possible to include 100% of the population for three reasons: (a) when we started the study, some people were about to finish the programme, so we did not have time to ask them to complete the questionnaires; (b) some people refused to participate; (c) language barriers: some people did not have sufficient knowledge of Spanish and had to be excluded from the study. Even so, the sample includes the majority of the population, and can be considered representative. In fact, our sample is predominantly male, as expected, considering the data provided by the Statistics Institute of Catalonia (IDESCAT, 2023), which shows that the 82.07% of young offenders in Catalonia are male. Furthermore, the team of the open intervention programmes in Tarragona has confirmed that the socio-demographic data of this sample corresponds to that of the population at the time it was collected.

With regard to the seriousness of the offence committed, 39.1% of the sample had committed misdemeanours: against people, against property, threats without aggression, and so on. The 39.1% had committed summary offences triable either way: assaults without injury or with minor injuries, minor threats, minor coercion, thefts of less than 400 euros. A total of 12.4% had committed misdemeanours offences: driving under the influence of drugs or alcohol (depending on the blood alcohol level and consequences), thefts above 400 euros, non-serious injuries (not requiring medical-surgical treatment). A total of 0.5% had committed felonies: manslaughter, murder, irreversible harm to another person, kidnapping, serious sexual offences. For 18 participants no information was provided about the type of offence for which they were convicted.

Instruments

The following measures were administered in this study:

Inventory of Callous-unemotional Traits and Antisocial Behaviour (INCA; Morales-Vives et al., 2019)

The questionnaire contains the following subscales: callousness (CA), uncaring (UC), unemotional (UE) and antisocial behaviour (AB). Callousness is defined as the lack of empathy and sensitivity to the needs and suffering of others. It also includes the tendency to manipulate others for personal benefit, without feeling guilty or remorse for the suffering caused. UC refers to the lack of responsibility and effort, with a tendency to laxity with regard to own duties and obligations. UE refers to a poor emotional affectivity, with a tendency to not show your own emotions. Finally, AB is defined as the tendency to violate the norms and rules of society, challenging authority and engaging in behaviours that are either illegal or violate social rules and that can harm others or the community as a whole. The CA subscale contains 11 items, the UC subscale contains 9 items, the UE subscale contains 9 items, and the AB subscale contains 8 items. A dummy item is also included as a first item of the questionnaire so that it can be used as a training item. For this reason, this first training item was not included in the analyses. Moreover, the instrument also contains four items that are markers of social desirability, to control for this response bias. In fact, the questionnaire was developed using the methods to control social desirability and acquiescence by Ferrando et al. (2009) and Lorenzo-Seva and Ferrando (2009), described below (see Data Analysis section), so it provides scores free from these response biases. The response categories of all the 43 items range from 1 (completely disagree) to 5 (completely agree). The marginal reliability of the content factor score estimates obtained in the original study were $\rho_{\theta\theta} = .96$ for UE, $\rho_{\theta\theta} = .91$ for CA, $\rho_{\theta\theta} = .93$ for UC, and $\rho_{\theta\theta} = .92$ for AB. The estimates obtained in the current study are shown below.

The Indirect–Direct Aggression Questionnaire (I–DAQ; Ruiz-Pamies et al., 2014)

This questionnaire is made of 27 items with response categories ranging from 1 (completely disagree) to 5 (completely agree). It contains three subscales: physical aggression (PA), verbal aggression (VA) and indirect aggression (IA). It also provides scores for overall aggression. PA refers to behaviours used to cause physical harm to another person (e.g. kicking or hitting). For example, the item 'Some people provoke me so much that we even hit each other' belongs to this subscale. VA refers to behaviour that is intended to harm another person through verbal expression, such as shouting or insulting them. For example, the item 'When I am angry with someone, I insult him/her' belongs to the VA subscale. While PA and VA are forms of direct aggression, IA refers to behaviours aimed to harm the victim without facing him/her directly, as, for example, spreading rumours, gossiping or excluding this person from the group. For example, the item 'When I get angry with someone, I spread unpleasant rumours about him/her' belongs to the IA subscale. The test was developed using the same methods to control social desirability and acquiescence as those in the INCA questionnaire, so the questionnaire provides scores free from these response biases. The marginal reliability of the content factor score estimates based on the 'a priori' three-dimensional structure was $\rho_{\theta\theta} = .83$ for PA, $\rho_{\theta\theta} = .77$ for VA, and $\rho_{\theta\theta} = .78$ for IA.

Barratt Impulsiveness Scale–11 for Children (BIS–11c; Chahin et al., 2010)

This instrument is a modified version of Barratt's BIS–11 (Barratt, 1985) adapted for Spanish children and adolescents. This questionnaire is made of 26 items with response categories ranging from 1 (never/almost never)

to 4 (always/almost always). It includes the following subscales: cognitive impulsiveness (CI), motor impulsiveness (MI) and non-planning Impulsiveness (N-PI). In this questionnaire, CI refers to the tendency to make quick decisions, which is related to functional impulsivity. For example, the item 'I am a quick thinker' belongs to the CI subscale. In contrast, MI and N-PI are related to dysfunctional impulsivity. MI refers to the tendency of acting without thinking, being carried away by the impulse of the moment. For example, the item 'I buy things without thinking' belongs to the MI subscale. N-PI refers to the lack of planning and interest in the future. For example, the reverse-keyed item 'I do projects for the future' belongs to this subscale. The reliability estimates of the raw subscale scores, assessed with Cronbach's alpha, were .70 for CI, .86 for MI and .81 for N-PI.

Procedure

This project was approved by the Research and Innovation Ethics Committee (CEIPSA) of the Universitat Rovira i Virgili (CEIPSA-2021-PR-0028). According to Spanish law, adolescents over 14 years old can consent by themselves to participate in a study of this type, so parental consent is not required for these ages (Organic Law 3/2018 of 5 December, article 7). For this reason, we obtained the written informed consent from all participants, in accordance with the Spanish law and the Declaration of Helsinki. Questionnaires were administered individually by two professional psychologists with training and experience in this field. Participants were informed that participation was voluntary. Questionnaires were anonymous, and confidentiality of participant's data was guaranteed.

Data analyses

Analyses proceeded sequentially in three general stages: (a) establishing the dimensionality and structure of the INCA, (b) assessing the scalability of the items and the properties of the factor/scale scores, and (c) obtaining validity evidence.

First general stage: dimensionality and structural assessment

Preliminary analyses were first carried out in order to assess: (a) the most appropriate modelling (linear or nonlinear; see below) and (b) the adequacy of the data for being factored. The second assessment was made using the Kaiser-Meye-Olkin (KMO) measure of overall adequacy; the normed measure of sampling adequacy (MSA) indices for each of the items, and the Expected REsidual correlation direct Change index for each pair of items (EREC; Ferrando et al., 2022), to determine possible correlated residuals or doublets.

Given that the antisocial behaviour factor had not been replicated in the previous study with a Mexican sample, and considering that the same result might happen in the present study, we decided to begin with a dimensional study to determine, first, the number and nature of the common factors underlying the data in this juvenile justice sample and, second, on which INCA factor the antisocial behaviour items load in the case of poor functioning. Dimensional assessment was based on the exploratory factor analysis (EFA) model and focused on the comparison between three-factor and four-factor solutions (see above). EFAs were performed using the FACTOR 12.06.08 (Lorenzo-Seva & Ferrando, 2006) and the *Psychological Test Toolbox* program (Navarro-González et al., 2019), which allows the procedures to control response bias that were already used by Morales-Vives et al. (2019) to be used. The solutions were fitted using robust unweighted least squares (ULS) estimation, and their adequacy was assessed by using fit indices that cover different facets of fit: (a) absolute fit (goodness of fit index, GFI, and root mean square residual, RMSR), (b) relative fit (root mean square error of approximation, RMSEA) and (c) comparative fit (CFI). The improvement of fit when going from three to four factors was assessed by

using the relative normed effect-size difference index by Ferrando, Navarro-González, & Lorenzo-Seva, (2024). Parallel analysis was also used as an auxiliary source of information. Next, EFA solutions were analytically rotated (oblique Promin rotation; Lorenzo-Seva, 1999), and their simplicity was assessed by using Bentler's simplicity index (Bentler, 1977) and the LS simplicity index (Lorenzo-Seva, 2003).

Provided that the final agreed EFA solution was simple enough, a confirmatory factor analysis (CFA) solution was next fitted to the data. The usual practice of fitting a CFA solution based on a preliminary EFA solution is to use a cross-validation schema in which the available sample is split into two halves and the CFA solution is fitted in the second 'new' validation subsample (e.g. Browne, 2000). This approach, however, was not feasible here due to the limited sample size (it should be stressed, however, that the sample size is small by structural modelling standards, but is quite large for such a specific population such a juvenile offenders). We acknowledge, indeed, that this is a (possibly unavoidable) limitation of the study.

When empirical cross-validation is not feasible, the best existing approach is to provide evidence of replicability based on the available single sample (Browne, 2000). In particular, we provided here two sources of evidence. First was via bootstrap resampling (e.g. Yung & Chan, 1999), which can be viewed as a multiple cross-validation assessment based on a large number of 'pseudo-samples' taken with replacement from the single available sample. Second, was via the expected cross-validation index (ECVI; e.g. Browne & Cudeck, 1993) which is an analytical estimate of the expected fit the CFA solution would have over all possible validation samples. In addition to the measures of replicability, the strength of the CFA solution in terms of power for detecting possible misspecifications was also assessed by undertaking a power analysis assessment (Lee et al., 2012). Overall, all the CFA part of the study was performed using Mplus 8.11 (Muthén & Muthén, 2017), and goodness-of-fit assessment used the same indices as the previous dimensional assessment above.

Second general stage: scalability and scoring

The functioning of the INCA items as measures was first assessed by using a graphical bivariate (location-discrimination) plot (e.g. Henrysson, 1971), in which the items were differentiated according to the factor they belonged to.

The basic scoring schema used here was Bayes-EAP (e.g. Bock & Mislevy, 1982), which uses all the information contained in the CFA solution from which the scores were derived. The accuracy of EAP scores varies at different trait (factor) levels (i.e. conditional reliability). So, it is possible to determine the trait ranges at which the scores provide more accurate measurement by using information curves (e.g. Brown & Croudace, 2015). The EAP-related outcomes in this study were: (a) the plot of the information curves corresponding to the estimated scores in each factor, and (b) the marginal reliabilities corresponding to each set of factor scores (e.g. Brown & Croudace, 2015). Because previous results at the first general stage suggested that there were three pairs of items with substantial correlated residuals, the marginal and conditional reliability estimates were corrected for this phenomenon.

Finally, the behaviour of the simple sum (subscale) scores as proxies for the factors they are intended to measure was also assessed (e.g. Comrey & Lee, 1992). Sum scores are a crude approach to the latent factor levels they attempt to estimate, but are very simple, familiar to practitioners, and can be very stable across samples. The outcomes associated with sum scores here were the reliability estimates and the coefficients of fidelity (COF). The reliability estimates were the unbiased omega estimates corrected for the presence of doublets ($\omega_{LD;}$ Ferrando, Navarro-González, &

Morales-Vives, 2024). The COF is an estimate of the correlation between the sum scores and the latent factor they intend to measure, and so it measures the extent to which sum scores are good proxies for the factor.

Third general stage: validity evidence

Finally, we obtained the correlations between the CU trait score estimates and the impulsiveness and aggressiveness subscale scores, using SPSS 29.0.1. Because the relations between both sets of variables are inferred from fallible measures that contain measurement error, we computed (a) the product–moment correlations between the corresponding scores (i.e. the empirical validity coefficients) and (b) the corresponding disattenuated correlations (corrected by using the marginal reliability estimates; i.e. theoretical validity coefficients).

Results

First general stage: dimensionality and structural assessment

Table 1 shows the means, standard deviations, skewness and kurtosis of all the content items. As the participants in this study were young offenders, who are expected to have higher trait levels associated with psychopathy and antisocial behaviour than the general population, we did not expect to obtain as many items with skewed distributions as in the previous study by Morales-Vives et al. (2019). However, six items had rather extreme skewness and kurtosis values (above 1 or below −1) which is the main reason for which the ordered-categorical non-linear FA was the chosen model.

The KMO index value was .75, which suggested that the correlation matrix was suitable for factor analysis. Only three pairs of items had substantial EREC values: pair 2–6, pair 32–37 and pair 7–43. The content of Items 2 and 6 is very similar, each item being a reformulation of the other, which explains why

Table 1. Means, standard deviations, skewness and kurtosis on the INCA items.

	Mean	Standard deviation	Skewness	Kurtosis
Item 2	3.59	1.29	−0.57	−0.77
Item 3	2.19	1.22	0.80	−0.21
Item 4	3.07	1.30	−0.11	−1.03
Item 5	3.46	1.56	−0.43	−1.38
Item 6	3.67	1.28	−0.78	−0.44
Item 7	1.50	0.79	1.80	3.57
Item 8	3.92	1.07	−0.82	0.01
Item 9	2.13	1.33	0.96	−0.29
Item 10	2.85	1.34	0.15	−1.12
Item 11	3.99	1.14	−1.23	0.98
Item 12	3.86	1.09	−0.89	0.26
Item 13	3.94	0.96	−0.87	0.66
Item 14	3.19	1.33	−0.24	−1.04
Item 15	4.20	1.05	−1.40	1.56
Item 16	2.28	1.16	0.59	−0.49
Item 17	2.42	1.17	0.42	−0.68
Item 18	3.49	1.40	−0.43	−1.17
Item 19	2.65	1.28	0.31	−0.89
Item 20	3.88	1.15	−0.98	0.27
Item 21	2.65	1.18	0.31	−0.77
Item 22	3.68	1.64	−0.77	−1.12
Item 23	3.12	1.24	−0.13	−0.95
Item 24	2.67	1.45	0.30	−1.31
Item 25	3.98	1.07	−0.97	0.37
Item 26	3.36	1.29	−0.44	−0.83
Item 27	3.84	1.22	−0.79	−0.37
Item 28	2.94	1.24	0.00	−0.99
Item 29	4.08	0.84	−0.72	0.54
Item 30	2.56	1.32	0.51	−0.83
Item 31	3.93	1.08	−0.96	0.53
Item 32	3.33	1.28	−0.33	−0.93
Item 33	3.05	1.41	−0.14	−1.24
Item 34	1.81	1.17	1.40	0.97
Item 35	3.72	1.21	−0.86	−0.07
Item 36	1.71	1.09	1.52	1.52
Item 37	3.19	1.31	−0.27	−1.05
Item 38	3.86	1.11	−0.92	0.28
Item 39	3.62	1.28	−0.68	−0.57
Item 40	1.92	1.20	1.26	0.58
Item 41	3.85	1.13	−0.94	0.20
Item 42	3.03	1.39	−0.13	−1.23
Item 43	2.07	1.24	0.92	−0.17

Note: INCA = Inventory of Callous-unemotional Traits and Antisocial Behaviour.

they share specificity (2. I find it very difficult to show my feelings; 6. I keep my feelings to myself). The same happens with the pair 32–37 (32. People can tell how I am feeling just by looking at me; 37. Other people can tell at once if I am sad or angry) and the pair 7–43 (7. I take advantage of others; 43. I sometimes use others to get what I want). None of the initial items had normed MSA indices below 0.50 (they ranged between 0.51 and 0.84), which suggests that each item is consistent with respect to the entire item set.

With regard to the comparisons between the three-content-factor solution and the initial solution proposed by Morales-Vives et al. (2019), to start with, parallel analysis suggested three content factors as the most plausible dimensionality for this data. As for the improvement in fit when going from three to four content factors, the value of relative normed effect-size difference index (0.05) was below the threshold value (0.06), which suggests that extracting a fourth factor does not noticeably improve the amount of explained common variance beyond that explained with three factors.

Turning now to the rotated solutions, although all the items of the UE and UC subscales under the four-factor solution loaded in the expected factors, many items of the CA and AB subscales were complex, loading in both factors. In contrast, when only three content factors were extracted, all but two items (20 and 24) from the CA and AB subscales loaded in a common factor. As the AB items (5, 9, 13, 17, 22, 26, 31 and 36) did not define a factor on their own, we decided to remove these items and fitting again a three-content-factors EFA solution. As all the items except the 24th loaded in the expected factor, we removed this item and fitted a final EFA solution. In this solution, all the items substantially loaded in the expected factors, with loadings ranging from .33 to .73, although Item 38 had a secondary loading of .28 in UC, and Item 40 had a secondary loading of .27 on UE. The S simplicity index was .99 (99th percentile) and the LS simplicity index was .52 (99th percentile), which suggest that, in general, the items mainly load on a single factor or, in other words, that the solution closely approaches simple structure. As for the impact of response biases, eight items loaded in the social desirability factor, and 14 items loaded in the acquiescence factor, which suggests that the need to control for these biases is warranted, as Morales-Vives et al. (2019) pointed out. The goodness-of-fit results for this final EFA solution were: GFI = .97; RMSR = 0.053; RMSEA = 0.057 and CFI = 0.95. Therefore, the fit can be considered very good for each of the facets that were considered (overall, relative and comparative fit).

Given the simplicity of the final EFA structure, a CFA solution was next specified and fitted to the total sample, as described above, and its stability and tenability were assessed by using bootstrap resampling, the ECVI and power analysis. The proposed solution was based on that obtained in the previous EFAs and consisted of a three-content correlated-factor solution, in which each item had only a non-zero loading on one factor, except for Items 38 and 40, which were specified to have secondary loadings in UC and UE, respectively. Three pairs of correlated residuals (2–6, 32–37, 7–43) were also specified. CFA results are in Table 2. As can be seen, all the items have salient loadings in their corresponding factors. Regarding the estimated inter-factor correlations, the UE factor had a correlation of .22 with UC and .26 with CA, and the UC factor had a correlation of .61 with the CA factor.

Goodness of fit results for the solution in Table 2 were: GFI = .96; RMSR = 0.060; RMSEA = 0.060 and CFI = 0.90. Furthermore, bootstrap resampling based on 2000 draws suggested that the solution was quite replicable, in terms of both parameter estimates and goodness-of fit results (confidence intervals are available from the authors). In the same line, the ECVI results predicted that the fit would slightly deteriorate across

Table 2. Loading matrix obtained from the confirmatory factor analysis.

		Loading
FACTOR 1: Unemotional	Item 2	.68
	Item 6	.68
	Item 10	−.81
	Item 14	.71
	Item 19	.47
	Item 23	−.55
	Item 28	−.48
	Item 32	−.47
	Item 37	−.52
	Item 40	−.34
	Item 41	.49
FACTOR 2: Uncaring	Item 4	.57
	Item 8	−.64
	Item 12	−.58
	Item 16	.60
	Item 21	.69
	Item 25	−.31
	Item 30	.49
	Item 35	−.72
	Item 38	.36
	Item 39	−.35
FACTOR 3: Callousness	Item 3	.52
	Item 7	.67
	Item 11	−.52
	Item 15	−.55
	Item 20	−.30
	Item 29	−.65
	Item 34	.45
	Item 38	−.48
	Item 40	.69
	Item 43	.61

different equivalent samples (which is only to be expected) but that goodness of fit would continue to be acceptable. For example, the estimated RMSEA across equivalent samples would be 0.066. Finally, power analysis results based on the standard settings by Lee et al. (2012) gave a beta estimate of 0.99 for detecting a moderate model misspecification. So, given the size of the solution (369 degrees of freedom), there is ample power for detecting model misspecification. Overall, the solution in Table 2 can be taken as a strong basis for determine individual scores and related outcomes in the next section.

Second general stage: scalability information and scoring

Figure 1 shows the location-discrimination plot in which the items are differentiated according to the factor they belong to. With regard to the overall item discriminating power, it was assessed by using the square root of the item communality, and the results are both clear and satisfactory. The discriminations are in the 0.3–0.8 range, which is considered very good for measurement purposes (Henrysson, 1971). With regard to location, the chosen index is the item mean (after reversing if necessary) scaled in the 0–1 range (i.e. proportion metric; see Lord, 1980). In this scaling, values close to zero indicate 'difficulty' (low proportion of agreement) and values close to 1 'easiness'. Overall, the range of locations (0.10–0.80) can be considered optimal for scalability purposes (Henrysson, 1971). However, this statement should be qualified when items are differentiated by factors. In fact, as can be seen in Figure 1, the item locations for the callousness factor are considerably lower than those for the items belonging to the unemotional and uncaring factors. Given that callousness is related to the interpersonal and affective facets of psychopathy, being particularly maladaptive, it is understandable that its items are more 'difficult' than the others, even in a sample of juvenile offenders. In fact, there are different profiles of offenders (Pederson et al., 2023), and not all of them have a callous profile, being in some cases more characterised by impulsivity than by a callous personality (Frick & White, 2008).

We turn now to the functioning and properties of the IRT-based EAP scores. Figure 2 shows the information curves (in conditional reliability metric) corresponding to the EAP scores within each factor. The information profiles are quite clear and agree with the item locations in Figure 1. The unemotional and

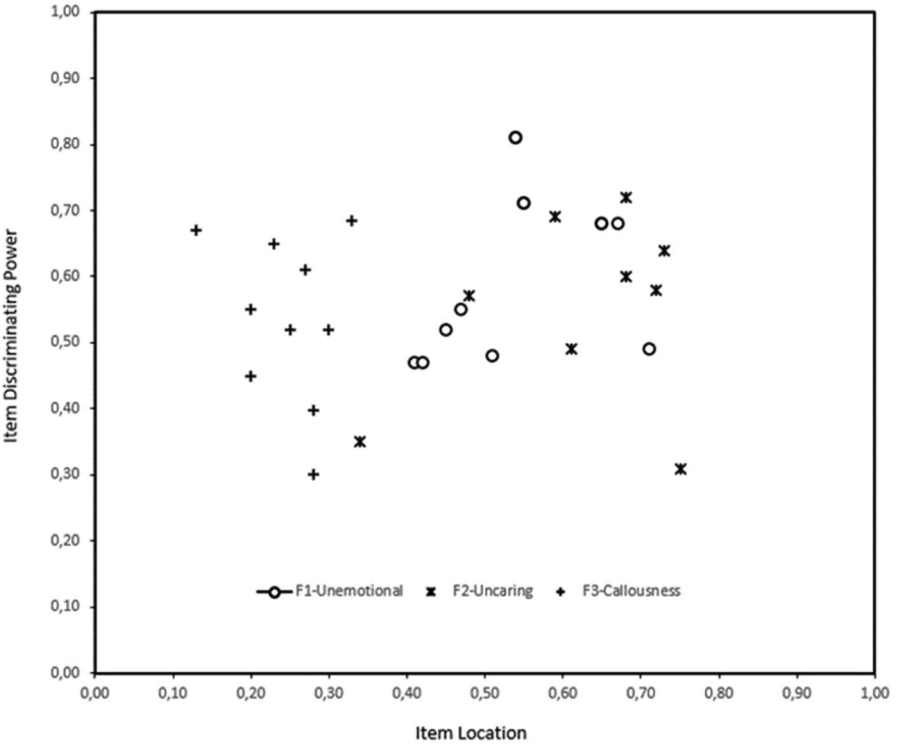

Figure 1. Graphical display of item characteristics indicating the factor they belong to.

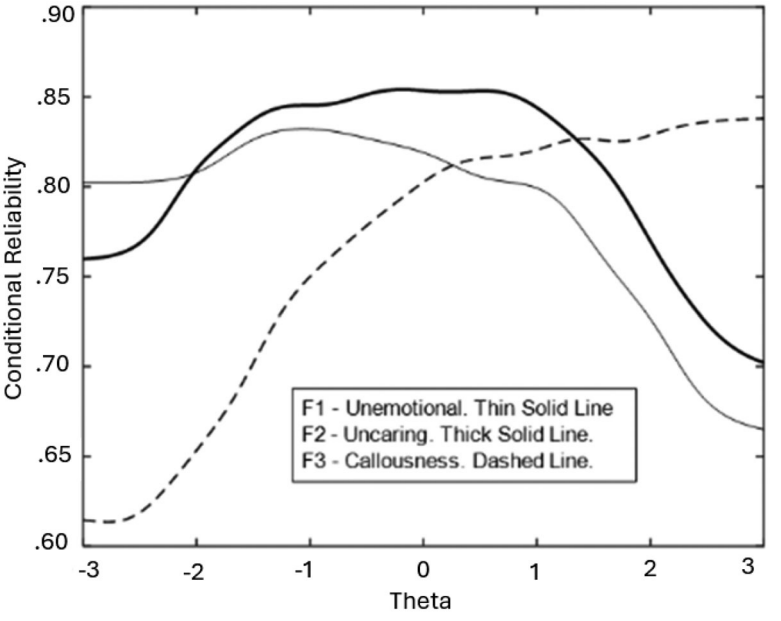

Figure 2. Information (conditional reliability) curves corresponding to the expected a posteriori (EAP) score estimates in each factor.

uncaring scores have both typical bell-shaped curves but they provide maximal accuracy at different points: the unemotional scores are most reliable about one standard deviation below the mean, whereas the uncaring scores attain maximal reliability at about one standard deviation above the mean. As for the callousness scores, the curve is monotonically increasing, suggesting that accuracy improves with increasing trait levels.

The marginal reliabilities of the EAP scores can be viewed as the averaged conditional reliabilities across the curves in Figure 2. The estimates were: .86 for UE, .84 for UC and .83 for CA. Although these values can be considered as adequate, they are lower than those found in the study by Morales-Vives et al. (2019). It should be considered, however, that the current sample is much more homogeneous than the one used in the previous study. The fact of having a sample of juvenile offenders, instead of a community sample like in the previous study, may have given rise to an attenuation of these estimations. In other words, these lower values probably reflect an attenuation-due-to selection effect because of the type of population to which the INCA was applied. Therefore, these values would be expected to be higher in more heterogeneous populations.

We turn now to the properties of the simple sum scores for the UE, UC and CA subscales. The omega reliability estimates corrected for doublets (ω_{LD}) are .79 for UE, .77 for UC and .74 for CA. These values are unbiased and more correct than the uncorrected omega estimates, but also 'deflated' with respect to the latter (Ferrando, Navarro-González, & Morales-Vives, 2024). However, the marginal reliabilities for the EAP scores reported above are also corrected for the residual correlations, and are systematically higher than the ω_{LD}·s. Note also that the rank of the reliability estimates is the same in both types of score. So, the differences can be plausibly interpreted as: (a) reflecting the loss of accuracy of the cruder sum scores with respect to the EAP scores and (b) the fact that the EAP scores have been 'cleaned' from the impact of response biases. Finally, the COF estimates were .89 for UE, .91 for UE and .85 for CA, which suggest that the sum scores are fairly good proxies for the 'true' factor scores.

As a summary, if the scores are to be used for screening or descriptive purposes, such as knowing approximately the levels of the respondents, the simple sum scores will we expected to suffice, and their use may be justified. In scenarios in which accurate individual measurement is required, however, the EAP factor scores provide more accurate estimates, as well as the profile of score functioning across subject levels. This profile, in turn, allows 'tailored' confidence intervals to be obtained for each individual estimate. So, in this case, the use of these scores is the most appropriate choice. Finally, regarding the marginal reliability estimates obtained with both scoring schemas, it should be taken into account that the sample (and the target population) are very specific and homogeneous regarding the constructs that are measured. So, the estimates can be viewed as appropriate. In community samples, larger values, more similar to the ones reported in the original study by Morales-Vives et al. (2019), would be expected.

External validity

Table 3 shows the Pearson correlations and the disattenauted correlations between the INCA, the BIS–11 and the I–DAQ scores estimations. As can be seen, UE was not correlated with impulsivity and aggression subscales. In contrast, CA was positively correlated with motor impulsiveness and non-planning impulsiveness. Although the effect sizes of these correlations were small, the disattenuated correlations, as expected, were relatively higher, showing in a clearer way the relationships between these variables. Thus, UC was positively correlated with the same impulsivity subscales, with medium effect sizes for the

Table 3. Pearson correlation coefficients and disattenuated correlations between the INCA scores and the BIS–11c and I–DAQ scores.

		UE	CA	UC
BIS–11c	MI	−.09 (−.13)	.22** (.34)	.44** (.67)
	N-PI	−.02 (−.03)	.19** (.31)	.48** (.77)
	CI	−.01 (−.02)	.03 (.05)	−.11 (−.20)
I–DAQ	PA	−.05 (−.08)	.31** (.50)	.27** (.43)
	VA	.02 (.03)	.23** (.40)	.28** (.46)
	IA	.02 (.03)	.39** (.68)	.23** (.39)

Note: Disattenuated correlations are shown in brackets. INCA = Inventory of Callous-unemotional Traits and Antisocial Behaviour; BIS–11c = Barratt Impulsiveness Scale–11 for Children; I–DAQ = Indirect–Direct Aggression Questionnaire; UE = unemotional; CA = callousness; UC = uncaring; MI = motor impulsivity; N-PI = non-planning impulsivity; CI = cognitive impulsivity; PA = physical aggression; VA = verbal aggression; IA = indirect aggression.
**$p < .01$.

Pearson correlations and large effect sizes for the disattenuated correlations.

Regarding the correlations with I–DAQ subscales scores, CA was positively correlated with the three subscales. Although the effect size for the Pearson correlation with verbal aggression is small, the disattenuated correlation is considerably higher, suggesting a medium effect size. UC was positively correlated with these subscales, with small effect sizes for the Pearson correlations and medium effect sizes for the disattenuated correlations.

Discussion

This study focuses on assessing the structural and measurement properties of the INCA questionnaire in a sample of Spanish juvenile offenders. Because previous studies were carried out in the community population, there was no evidence of the suitability of this instrument for the juvenile justice population, even though the assessment of these traits is particularly relevant in this population. Furthermore, the previous studies were mostly concerned with the structural properties, but not on measurement issues such as the properties of different types of scoring or their marginal and conditional reliabilities. These measurement properties have been assessed in the current study, as they are thought to be of great interest for professionals working in the field of juvenile justice.

Regarding the structural results, in the current study, as it occurred in the research carried out in Mexico (Morales-Vives et al., 2020), the antisocial behaviour items did not define a separate factor on its own. However, while in the Mexican study these items loaded on the uncaring factor, in the present study they loaded on the same factor as callousness. Taken together, these results suggest that these items function differentially and so have different implications in each type of population. It might be conjectured that in the Mexican community sample these antisocial behaviours are mostly indicators of lack of responsibility, whereas in the Spanish juvenile justice sample they mostly indicate lack of empathy, lack of remorse and indifference to the suffering of others. Continuing in this line, in the Mexican community sample, the reason might be cultural as, in this culture, the behaviours described in these items are perhaps not perceived as such relevant transgressions, which means that irresponsible people might engage in them without this engagement really implying a lack of empathy or sensitivity towards others. However, the case of Spanish young offenders seems to be different. According to the present results, the lack of empathy, remorse, and so on would possibly be behind

these behaviours, giving rise to a common, broader factor that includes consistently the antisocial-behaviour and callousness items. Furthermore, we could also hypothesise that, should the behaviours described by the antisocial-behaviour items had been more severe, these items could have been able to define a consistent factor on their own. However, given that they mostly describe lenient behaviours that are likely to be on a far smaller scale than those that got the respondents into trouble with the justice system, these items are well integrated in the callousness factor, a result that, as expected, did not occur in the Spanish community sample. If our reasoning is correct, a separate factor would be very difficult to identify in this type of population. In other words, these small acts of transgression, which work well to assess antisocial behaviour in Spanish community samples, are insignificant or too basic to work properly with juvenile offenders. In view of these results and those previously obtained in Mexico, the antisocial behaviour factor of the INCA questionnaire cannot be considered universal, since, unlike the three CU traits, it operates differently according to the type of population.

Considering the results so far discussed, we decided to eliminate the antisocial behaviour items, as it was previously done in the Mexican study. We also removed an item that did not load in the expected factor. As in the case of the Mexican study, eliminating the antisocial behaviour items results in adequate fit indices, a very simple factor structure, in which each item mainly loads on a single factor, and adequate congruence indices for all items. According to these results, the trimmed resulting instrument has adequate structural properties in this sample of juvenile justice: the fit to the three-factor model is acceptable, and the solution closely approaches a simple structure.

We turn now to the measurement properties of the INCA questionnaire in this sample of juvenile offenders, starting with the range of locations of the items. The range can be considered optimal for scalability purposes (Henrysson, 1971), especially for the unemotional and uncaring items. The callousness items are more 'difficult' than the others, which is understandable, as callousness refers to characteristics associated to primary psychopathy (Del Gaizo & Falkenbach, 2008). In fact, these results are congruent with the remarks of Eysenck and Eysenck (1976) and Eaves and Eysenck (1977), who already pointed out that few people state things such as not feeling discomfort with another's suffering, leading to strongly skewed distributions in this kind of measure. This property of the callousness subscale should be even more evident in community samples, where the proportion of callous people is expected to be lower.

Likewise, the information curve (in conditional reliability metric) corresponding to the EAP scores for the callousness factor is also considerably different from the curves of the other two factors. The uncaring and unemotional scores have 'standard' information profiles that reach their peak about one standard deviation above (uncaring) or below (unemotional) the mean. However, the curve for callousness is monotonically increasing, suggesting that accuracy improves with increasing trait levels. Similar results, also based on callousness scores, have been obtained recently by Morales-Vives et al. (in press), and these authors suggested that they might be explained by the unipolar nature of callousness, meaning that this construct would be more meaningful and interpretable at the upper pole of the continuum (different degrees of callousness) than at the lower pole (which would mostly reflect absence of trait manifestations). Exploring the possibility of calibrating and scoring callousness using a unipolar model deserves future research, as this would have implications for important decisions based on these scores (for example, determining cut-off values for identifying individuals susceptible to having high levels of this trait; Morales-Vives et al., 2023, in press). However, at present unipolar response models

to fit multidimensional measures such as the INCA have not yet been developed, and even less with measures in which not all the subscales would be unipolar.

We have also compared the properties of the simple sum scores with respect to the EAP scores, concluding that the former, in spite of the loss of accuracy they involve and that response biases are not corrected, are fairly good proxies for the 'true' factor scores. As many practitioners are more accustomed to using sum scores than EAP scores, it is important that they be good proxies, as is the case here. However, we encourage the use of EAP scores, as they are more accurate, allow tailored confidence intervals to be obtained for each respondent, and (in this case) are corrected from response biases. For this reason, in future studies we will develop a user-friendly and freely available programme for scoring INCA (both EAP scores and simple sum scores).

Regarding the external validity of INCA in this specific population, as expected, we found significant correlations between the uncaring and callousness factors and the two subscales of the BIS–11c that are linked to dysfunctional impulsivity (motor and non-planning impulsivity; Chahin et al., 2010). This is congruent with the study by Maurer et al. (2021) with male young offenders, which found that these two kinds of impulsivity are related to several psychopathic traits. But it should be considered that there are different profiles of juvenile offenders (Frick, 2006; Frick & White, 2008) and that not all adolescents with high CU levels are impulsive. However, according to the study on young offenders by Roşan & Costea-Bărluţiu (2013), the CU traits amplify the effect of impulsivity on aggressive behaviour. We did not find a significant correlation between the unemotional factor and impulsivity, as it was expected, considering the previous results found by Morales-Vives et al. (2019) with the INCA and Morales-Vives et al. (2020) with the INCA–M. Likewise, we obtained the expected results with the aggression subscales, callousness and uncaring being correlated with both direct and indirect forms of aggression and unemotional not being correlated with any kind of aggression. These results are congruent with the study by Kimonis et al. (2008) with adolescent offenders, which also showed that callous and irresponsible adolescents tend to display higher levels of aggressiveness, while the unemotional feature does not seem to play a relevant role, especially in boys. The meta-analytic review carried out by Cardinale and Marsh (2020) also suggests that unemotional is not correlated to reactive and proactive aggression. Therefore, according to these results, the INCA scores have good convergent and discriminant validity with impulsivity and aggression in juvenile offenders.

To sum up, the results of the three studies carried out in different populations (Spanish community sample, Mexican community sample and Spanish juvenile offenders sample) confirm that the CU traits assessed by the INCA questionnaire are stable through different populations and resist cross-cultural validation. However, this is not the case with the antisocial behaviour factor, which has not been replicated in the current study or in the previous Mexican study. This factor was only obtained in the original research carried out with a Spanish community sample, and no further studies have been done with this population that confirm its replicability in other samples. For this reason, further studies in which both Spanish community samples and Spanish juvenile offender samples are compared need to be done if a version of this instrument that works equally well in both populations and that has an equivalent factor structure is to be developed. In our opinion, it is not advisable to maintain a subscale that is not universal and that does not allow meaningful comparisons between different types of populations to be done. Moreover, further studies are needed with Spanish community samples that provide evidence on the psychometric properties of the instrument in this population beyond its factorial structure and

external validity. In fact, the present study shows the interest of these additional analyses, which provide a more global insight into the score properties for this kind of population.

A further limitation of this study is that most of the participants are male, which is characteristic of the population studied, but does not allow for gender-specific analyses. Future studies should be conducted with larger samples that include a higher number of females, as gender is thought to be a relevant variable in relation to both psychopathic traits and criminal behaviour (e.g. Gower et al., 2024; Joseph, 2022).

The present study makes important contributions that we would like to emphasise. The use of the INCA questionnaire in the population of young offenders may allow those with high levels in the CU traits to be identified. And this identification, in turn, is expected to be quite useful for the implementation of intervention programmes specifically targeted at this profile of offenders, to help them develop empathy and appropriate social skills, thereby reducing the risk of more serious future antisocial behaviour, and to improve their social integration. In addition, the study by Pechorro et al. (2022) showed that the CU traits can predict general recidivism, so the fact that the INCA scores are able to adequately assesses these traits may be helpful in implementing more effective and personalised rehabilitation programmes to reduce recidivism among juvenile offenders. For these reasons, we believe that this instrument may be of great interest to psychologists and other social professionals working in the justice system.

Regarding the implications of this study, we believe that it provides evidence that the same instrument does not necessarily function with the same properties in the community and in the juvenile justice populations. This possibility is not always considered in these domains, in which certain instruments are used 'by fiat' without having been specifically validated in the juvenile justice field. This lack of supporting validation evidence may lead to inadequate psychological assessments when instruments affected by this differential functioning are used with this purpose. Given that these young people may find themselves in complicated personal, family and social situations, and that they may engage in behaviours that affect themselves and others, the consequences of inadequate assessment may be problematic. It is, therefore, important to pay attention to this issue and to ensure adequate assessment of those involved in juvenile justice interventions.

Ethical standards

Declaration of conflicts of interest

The authors have declared that any financial benefit or interest has arisen from the direct applications of this research. In particular, each author has declared the following:

José M. Casas has declared no conflicts of interest.

Jorge-Manuel Dueñas has declared no conflicts of interest.

Pere J. Ferrando has declared no conflicts of interest.

Elena Castarlenas has declared no conflicts of interest.

Andreu Vigil-Colet has declared no conflicts of interest.

Juan C. Hernández-Navarro has declared no conflicts of interest.

Fabia Morales-Vives has declared no conflicts of interest.

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and national research committee *Research and Innovation Ethics Committee* (CEIPSA) of the Universitat Rovira i Virgili (CEIPSA-2021-PR-0028) and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Informed consent

Written informed consent was obtained from all individual participants included in the study. All participants acknowledge that they cannot be identified via the paper and that we have fully anonymised them.

Funding

This work is part of the project I + D + I under grant PID2023-148374NB-I00, funded by MICIU/AEI/10.13039/501100011033, and by the Catalan Ministry of Universities, Research and the Information Society under grant 2021 SGR 00036. The funding source was not involved in any step of the research process, nor in the writing and publication process.

Data availability statement

The data that support the findings of this study are openly available in the Research Data Repository at https://doi.org/10.34810/data2206.

ORCID

José M. Casas http://orcid.org/0000-0002-8547-6336
Jorge-Manuel Dueñas http://orcid.org/0000-0002-8954-7947
Pere J. Ferrando http://orcid.org/0000-0002-3133-5466
Elena Castarlenas http://orcid.org/0000-0003-0383-2526
Andreu Vigil-Colet http://orcid.org/0000-0003-3818-4514
Fabia Morales-Vives http://orcid.org/0000-0002-2095-0244

References

American Psychiatric Association (APA). (2013). *The diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association Publishing Textbooks.

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

Barratt, E. S. (1985). Impulsiveness subtraits: Arousal and information processing. In J. T. Spence & C. E. Izard (Eds.), *Motivation, emotion and personality* (pp. 137–145). Elsevier Science.

Bentler, P. M. (1977). Factor simplicity index and transformations. *Psychometrika*, *42*(2), 277–295. https://doi.org/10.1007/BF02294054

Bock, R. D., & Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*(4), 431–444. https://doi.org/10.1177/014662168200600405

Brennan, G. M., Stuppy-Sullivan, A. M., & Baskin-Sommers, A. R. (2022). Psychopathy and substance use disorders. In J. E. Vitale (Ed.), *The complexity of psychopathy. Dangerous behavior in clinical and forensic psychology*. Springer. https://doi.org/10.1007/978-3-030-83156-1_11

Brown, A., & Croudace, T. J. (2015). Scoring and estimating score precision using multidimensional IRT models. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of Item Response Theory Modeling: Applications to typical performance assessment* (pp. 307–333). Routledge/Taylor & Francis Group.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*(1), 108–132. https://doi.org/10.1006/jmps.1999.1279

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Campbell, M. A., Doucette, N. L., & French, S. A. (2009). Validity and stability of the youth psychopathic traits inventory in a non-forensic sample of young adults. *Journal of Personality Assessment*, *91*(6), 584–592. https://doi.org/10.1080/00223890903228679

Cardinale, E., & Marsh, A. (2020). The reliability and validity of the inventory of callous unemotional traits: A meta-analytic review. *Assessment*, *27*(1), 57–71. https://doi.org/10.1177/1073191117747392

Chahin, N., Cosi, S., Lorenzo-Seva, U., & Vigil-Colet, A. (2010). Stability of the factor structure of Barrat's Impulsivity Scales for children across cultures: A comparison of Spain and Colombia. *Psicothema*, *22*(4), 983–989.

Ciucci, E., Baroncelli, A., Franchi, M., Golmaryami, F. N., & Frick, P. J. (2014). The association between callous-unemotional

traits and behavioral and academic adjustment in children: further validation of the inventory of callous-unemotional traits. *Journal of Psychopathology and Behavioral Assessment*, *36*(2), 189–200. https://doi.org/10.1007/s10862-013-9384-z

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Psychology Press. https://doi.org/10.4324/9781315827506

Craig, S. G., Goulter, N., & Moretti, M. M. (2021). A systematic review of primary and secondary callous-unemotional traits and psychopathy variants in youth. *Clinical Child and Family Psychology Review*, *24*(1), 65–91. https://doi.org/10.1007/s10567-020-00329-x

Dadds, M. R., Fraser, J., Frost, A., & Hawes, D. J. (2005). Disentangling the underlying dimensions of psychopathy and conduct problems in childhood: a community study. *Journal of Consulting and Clinical Psychology*, *73*(3), 400–410. https://doi.org/10.1037/0022-006x.73.3.400

Del Gaizo, A. L., & Falkenbach, D. M. (2008). Primary and secondary psychopathic-traits and their relationship to perception and experience of emotion. *Personality and Individual Differences*, *45*(3), 206–212. https://doi.org/10.1016/j.paid.2008.03.019

Eaves, L. J., & Eysenck, H. J. (1977). A genotype-environmental model for psychoticism. *Advances in Behaviour Research and Therapy*, *1*(1), 5–26. https://doi.org/10.1016/0146-6402(77)90002-9

Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Callous-unemotional traits in a community sample of adolescents. *Assessment*, *13*(4), 454–469. https://doi.org/10.1177/1073191106287354

Eysenck, H. J., & Eysenck, S. B. G. (1976). *Psychoticism as a dimension of personality*. Hodder and Stoughton.

Ezpeleta, L., de la Osa, N., Granero, R., Penelo, E., & Domènech, J. M. (2013). Inventory of callous-unemotional traits in a community sample of preschoolers. *Journal of Clinical Child and Adolescent Psychology: The Official Journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53*, *42*(1), 91–105. https://doi.org/10.1080/15374416.2012.734221

Fanti, K. A., Frick, P. J., & Georgiou, S. (2009). Linking Callous-unemotional traits to instrumental and non-instrumental forms of aggression. *Journal of Psychopathology and Behavioral Assessment*, *31*(4), 285–298. https://doi.org/10.1007/s10862-008-9111-3

Feilhauer, J., Cima, M., & Arntz, A. (2012). Assessing callous-unemotional traits across different groups of youths: Further cross-cultural validation of the Inventory of callous-unemotional traits. *International Journal of Law and Psychiatry*, *35*(4), 251–262. https://doi.org/10.1016/j.ijlp.2012.04.002

Ferrando, P. J., Hernandez-Dorado, A., & Lorenzo-Seva, U. (2022). Detecting correlated residuals in exploratory factor analysis: New proposals and a comparison of procedures. *Structural Equation Modeling: A Multidisciplinary Journal*, *29*(4), 630–638. https://doi.org/10.1080/10705511.2021.2004543

Ferrando, P. J., & Lorenzo-Seva, U. (2000). Unrestricted versus restricted factor analysis of multidimensional test items: Some aspects of the problem and some suggestions. *Psicológica*, *21*(2), 301–323. https://www.redalyc.org/articulo.oa?id=16921206

Ferrando, P. J., Lorenzo-Seva, U., & Chico, E. (2009). A general factor-analytic procedure for assessing response bias in questionnaire measures. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(2), 364–381. https://doi.org/10.1080/10705510902751374

Ferrando, P. J., Navarro-González, D., & Lorenzo-Seva, U. (2024). A relative normed effect-size difference index for determining the number of common factors in exploratory solutions. *Educational and Psychological Measurement*, *84*(4), 736–752. https://doi.org/10.1177/00131644231196482

Ferrando, P. J., Navarro-González, D., & Morales-Vives, F. (2024). Linear and nonlinear indices of score accuracy and item effectiveness for measures that contain locally dependent items. *Educational and Psychological Measurement*, *85*(1), 60–81. https://doi.org/10.1177/00131644241257602

Forth, A. E., Kosson, D., & Hare, D. (2003). *The Hare PCL: Youth version*. Multi-Health Systems.

Frick, P. J. (2004). *The inventory of callous-unemotional traits*. University of New Orleans.

Frick, P. J. (2006). Developmental pathways to conduct disorder. *Child and Adolescent Psychiatric Clinics of North America*, *15*(2), 311–331, vii. https://doi.org/10.1016/j.chc.2005.11.003

Frick, P. J. (2013). *Clinical assessment of prosocial emotions (cape)*. University of New Orleans. https://faculty.lsu.edu/pfricklab/pdfs/cape-manual.pdf

Frick, P. J., & Hare, R. D. (2001). *The antisocial process screening device*. Multi-Health Systems.

Frick, P. J., Kimonis, E. R., Dandreaux, D. M., & Farell, J. M. (2003). The 4 year stability of psychopathic traits in non-referred youth. *Behavioral Sciences & the Law*, *21*(6), 713–736. https://doi.org/10.1002/bsl.568

Frick, P. J., O'Brien, B. S., Wootton, J. M., & McBurnett, K. (1994). Psychopathy and conduct problems in children. *Journal of Abnormal Psychology*, *103*(4), 700–707. https://doi.org/10.1037/0021-843X.103.4.700

Frick, P. J., Stickle, T. R., Dandreaux, D. M., Farrell, J. M., & Kimonis, E. R. (2005). Callous-unemotional traits in predicting the severity and stability of conduct problems and delinquency. *Journal of Abnormal Child Psychology*, *33*(4), 471–487. https://doi.org/10.1007/s10648-005-5728-9

Frick, P. J., & White, S. F. (2008). Research Review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *49*(4), 359–375. https://doi.org/10.1111/j.1469-7610.2007.01862.x

Gower, M., Morgan, F., & Saunders, J. (2024). Gender responsivity in the assessment and treatment of offenders. *Psychiatry, Psychology, and Law: An Interdisciplinary Journal of the Australian and New Zealand Association of Psychiatry, Psychology and Law*, *31*(4), 587–611. https://doi.org/10.1080/13218719.2023.2206855

Henrysson, S. (1971). Gathering, analyzing and using data on test items. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 130–115). America Council on Education.

Hollingshead, A. B. (1975). *Four factor index of social status*. Yale University.

Houghton, S., Hunter, S. C., & Crow, J. (2013). Assessing callous unemotional traits in children aged 7- to 12-years: a confirmatory factor analysis of the inventory of callous unemotional traits. *Journal of Psychopathology and Behavioral Assessment*, *35*(2), 215–222. https://doi.org/10.1007/s10862-012-9324-3

IDESCAT. (2023). *Juvenile justice. Minors and youths attended to. By sex, ages and nationality*. Statistical Institute of Catalonia. https://www.idescat.cat/indicadors/?id=aec&n=15873&tema=justi&lang=en

Joseph, J. J. (2022). Exploring sex differences between dimensions of psychopathy, executive functioning and youth gang membership. *Psychology, Crime & Law*, *30*(8), 891–916. https://doi.org/10.1080/1068316X.2022.2131785

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., Aucoin, K. J., & 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*(3), 241–252. https://doi.org/10.1016/j.ijlp.2008.04.002

Lee, T., Cai, L., & MacCallum, R. (2012). Power analysis for test of structural equation models. In R. H. Hoyle (Ed.), *Handbook of structural equation modelling* (pp. 181–194). Guilford Press.

Loeber, R. (1982). The stability of antisocial and delinquent child behavior: A review. *Child Development*, *53*(6), 1431–1446. https://doi.org/10.1111/j.1467-8624.1982.tb03465.x

López-Romero, L., Romero, E., & Luengo, M. A. (2012). Disentangling the role of psychopathic traits and externalizing behavior in predicting conduct problems from childhood to adolescence. *Journal of Youth and Adolescence*, *41*(11), 1397–1408. https://doi.org/10.1007/s10964-012-9800-9

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Erlbaum.

Lorenzo-Seva, U. (1999). Promin: A method for oblique factor rotation. *Multivariate Behavioral Research*, *34*(3), 347–365. https://doi.org/10.1207/S15327906MBR3403_3

Lorenzo-Seva, U. (2003). A factor simplicity index. *Psychometrika*, *68*(1), 49–60. https://doi.org/10.1007/BF02296652

Lorenzo-Seva, U., & Ferrando, P. J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods*, *38*(1), 88–91. https://doi.org/10.3758/bf03192753

Lorenzo-Seva, U., & Ferrando, P. J. (2009). Acquiescent responding in partially balanced multidimensional scales. *British Journal of Mathematical and Statistical Psychology*, *62*(2), 319–326. https://doi.org/10.1348/000711007X265164

Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*(2), 57–64. https://doi.org/10.1027/1614-2241.2.2.57

Lynam, D. R. (1996). Early identification of chronic offenders: Who is the fledgling psychopath? *Psychological Bulletin*, *120*(2), 209–234. https://doi.org/10.1037/0033-2909.120.2.209

Maurer, J. M., Tirrell, P. S., Anderson, N. E., Rodriguez, S. N., Caldwell, M. F., Van Rybroek, G. J., & Kiehl, K. A. (2021). Dimensions of impulsivity related to psychopathic traits and homicidal behavior among

incarcerated male youth offenders. *Psychiatry Research*, *303*, 114094. https://doi.org/10.1016/j.psychres.2021.114094

Morales-Vives, F., Cosi, S., Lorenzo-Seva, U., & Vigil-Colet, U. (2019). The Inventory of Callous-unemotional traits and Antisocial behaviour (INCA) for young people: Development and validation in a community sample. *Frontiers in Psychology*, *10*, 713. https://doi.org/10.3389/fpsyg.2019.00713

Morales-Vives, F., Ferrando, P. J., & Dueñas, J.-M. (2023). Should suicidal ideation be regarded as a dimension, a unipolar trait or a mixture? A model-based analysis at the score level. *Current Psychology*, *42*(25), 21397–21411. https://doi.org/10.1007/s12144-022-03224-6

Morales-Vives, F., Ferrando, P. J., & Hernández-Dorado, A. (in press). Modelling maladaptive personality traits with unipolar item response theory: The case of Callousness. *The Journal of General Psychology*, 1–28. https://doi.org/10.1080/00221309.2024.2404398

Morales-Vives, F., Gómez-Herrera, M., & Vigil-Colet, A. (2020). INCA-M: Mexican adaptation of the inventory of callous-unemotional traits and antisocial behavior. *Frontiers in Psychology*, *11*, 753. https://doi.org/10.3389/fpsyg.2020.00753

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén. https://www.statmodel.com/HTML_UG/introV8.htm

Navarro-Gonzalez, D., Vigil-Colet, A., Ferrando, P. J., & Lorenzo-Seva, U. (2019). Psychological Test Toolbox: A new tool to compute factor analysis controlling response bias. *Journal of Statistical Software*, *91*(6), 1–21. https://doi.org/10.18637/jss.v091.i06

Pardini, D. A., Lochman, J. E., & Frick, P. J. (2003). Callous/unemotional traits and social-cognitive processes in adjudicated youths. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*(3), 364–371. https://doi.org/10.1097/00004583-200303000-00018

Pechorro, P., Marsee, M., DeLisi, M., & Gonçalves, R. A. (2022). Utility of the inventory of callous-unemotional traits in predicting criminal recidivism among detained youth. *Journal of Psychopathology and Behavioral Assessment*, *44*(4), 1090–1098. https://doi.org/10.1007/s10862-021-09904-0

Pederson, C. A., Griffith, R. L., Nowalis, S., & Fite, P. J. (2023). Creating profiles of juvenile offenders using functions of aggression and callous-unemotional traits: Relations to crime type. *Psychiatry, Psychology, and Law: An Interdisciplinary Journal of the Australian and New Zealand Association of Psychiatry, Psychology and Law*, *30*(5), 713–736. https://doi.org/10.1080/13218719.2022.2116609

Ray, J. V., & Wall Myers, T. D. (2018). Callous–unemotional traits. Relevance and implications for juvenile justice. In M. DeLisi (Ed.), *Routledge international handbook of psychopathy and crime* (pp. 80–94). Routledge Taylor & Francis Group.

Roşan, A. M., & Costea-Bărluţiu, C. (2013). Associations between callous-unemotional traits, aggression and psychopathology in detained adolescent males. *Journal of Cognitive & Behavioral Psychotherapies*, *13*(2), 397–407.

Rose, K., Woodworth, M., & Minton, J. (2020). An exploration of individual differences in a sample of youth charged with violent sexual and non-sexual crimes. *Psychiatry, Psychology, and Law: An Interdisciplinary Journal of the Australian and New Zealand Association of Psychiatry, Psychology and Law*, *27*(1), 44–60. https://doi.org/10.1080/13218719.2019.1687043

Ruiz-Pamies, M., Lorenzo-Seva, U., Morales-Vives, F., Cosi, S., & Vigil-Colet, A. (2014). I-DAQ: a new test to assess direct and indirect aggression free of response bias. *The Spanish Journal of Psychology*, *17*(E41), E41. https://doi.org/10.1017/sjp.2014.43

Stickle, T. R., Kirkpatrick, N. M., & Brush, L. N. (2009). Callous-unemotional traits and social information processing: Multiple risk-factor models for understanding aggressive behavior in antisocial youth. *Law and Human Behavior*, *33*(6), 515–529. https://doi.org/10.1007/s10979-008-9171-7

Taylor, J., & Lang, A. R. (2006). Psychopathy and substance use disorders. In C. J. Patrick (Ed.), *Handbook of psychopathy* (pp. 495–511). The Guilford Press.

Thornberg, R., & Jungert, T. (2017). Callous-unemotional traits, harm-effect moral reasoning, and bullying among Swedish children. *Child & Youth Care Forum*, *46*(4), 559–575. https://doi.org/10.1007/s10566-017-9395-0

Timmerman, M. E., & Lorenzo-Seva, U. (2011). Dimensionality assessment of ordered polytomous items with Parallel Analysis. *Psychological Methods*, *16*(2), 209–220. https://doi.org/10.1037/a0023353

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies (No. PRS-984)*. Educational Testing Services.

Waller, R., & Hyde, L. W. (2018). Callous-unemotional behaviors in early childhood: The development of empathy and prosociality gone awry. *Current Opinion in Psychology*, *20*, 11–16. https://doi.org/10.1016/j.copsyc.2017.07.037

White, S. F., Cruise, K. R., & Frick, P. J. (2009). Differential correlates to self-report and parent-report of callous–unemotional traits in a sample of juvenile sexual offenders. *Behavioral Sciences & the Law*, *27*(6), 910–928. https://doi.org/10.1002/bsl.911

Wymbs, B. T., McCarty, C. A., King, K. M., McCauley, E., Vander Stoep, A., Baer, J. S., & Waschbusch, D. A. (2012). Callous-unemotional traits as unique prospective risk factors for substance use in early adolescent boys and girls. *Journal of Abnormal Child Psychology*, *40*(7), 1099–1110. https://doi.org/10.1007/s10802-012-9628-5

Yung, Y. F., & Chan, W. (1999). Statistical analyses using bootstrapping: Concepts and implementation. In R. H. Hoyle (Ed.). *Statistical strategies for small sample research* (pp. 82–105). Sage.