



A Multi-Informant Study of the Validity of the Inventory of Callous-Unemotional Traits in a Sample of Norwegian Adolescents with Behavior Problems

Dagfinn Mørkrid Thøgersen^{1,2} · Mette Elmoose Andersen² · Gunnar Bjørnebekk³

Published online: 15 February 2020

© The Author(s) 2020

Abstract

The Inventory of Callous-Unemotional traits (ICU) is a widely used measure of the affective aspects of psychopathy in children and adolescents. Although a 3-bifactor model of the ICU has often been supported, method-variance effects and mixed results for the Unemotional subscale raise concerns. This study applied a multi-informant design to investigate the structural and concurrent validity of the self-, parent- and teacher-versions of the ICU in a Norwegian at-risk adolescent sample ($N = 160$, female = 46.3%, mean age = 14.73 years, $SD = 1.47$). Confirmatory Factor Analysis demonstrated method-variance effects of the standard- and reverse-scored items. The best fitting model was the 12-item 2-factor ICU, comprising a Callousness and an Uncaring factor. The scale reliabilities were acceptable with Cronbach alphas ranging from .810 to .906 across respondent groups. Cross-rater reliability was poor, with Intra Class Correlations ranging from .170 to .226. The concurrent validity of the ICU12 was supported by within-rater associations to externalizing problems, aggression, and problematic alcohol use. Cross-rater associations of ICU12-scores to self-reported delinquency and lower levels of anxiety, provided additional support for the concurrent validity. The unidimensional ICU12 total score was associated both to delinquency and rule-breaking behavior as observed for the Callousness factor and to lack of anxiety and reduced fear sensitivity as observed for the Uncaring factor. The ICU items aimed at capturing lack of prosocial emotions, seem to need revision. Future research should assess the predictive validity and clinical relevance of unidimensional versus factor models of the ICU.

Keywords Callous-unemotional traits · Adolescents · Behavior problems · Confirmatory factor analysis · Validation · Multi-informant

Introduction

The Inventory of Callous-Unemotional traits (ICU; Frick 2004) is the most commonly used measure of callous-

unemotional traits (CU) in children and adolescents. The ICU has been applied in multiple studies of both community and at-risk samples in several countries and languages (Byrd et al. 2017; Essau et al. 2006; Hawes et al. 2014a; Hawes et al. 2014b; Kimonis et al. 2008; Roose et al. 2010). Research on the factor structure of the ICU is not consistent in regard to the dimensionality of the measure (e.g. Ciucci et al. 2014; Pechorro et al. 2016b; Hawes et al. 2014b). A recent meta-analytic paper argues that there are no strong theoretical background nor empirical correlates to subordinate subfactors on the ICU (Ray and Frick 2018). Three studies applying Item Response Theory (IRT) analysis have suggested that the empirically derived factor structures of the ICU are due to methodological variance caused by the use of standard- and reverse-scored items (Lin et al. 2018; Paiva-Salisbury et al. 2017; Ray et al. 2016). Typically, studies find two main factors related to item-scoring procedures: a Callousness factor consisting of standard-scored

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10862-020-09788-6>) contains supplementary material, which is available to authorized users.

✉ Dagfinn Mørkrid Thøgersen
d.m.thogersen@nubu.no

¹ Norwegian Center for Child Behavioral Development, Essendrops gate 3, 0368 Oslo, Norway

² Department of Psychology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark

³ Department of Special Needs Education, University of Oslo, Sem Sælands vei 7, 0371 Oslo, Norway

items like “I do not care who I hurt to get what I want”, and an Uncaring factor consisting of reverse-scored items like “I apologize to people I hurt”. IRT-analyses suggest that the standard-scored items comprised of socially demeaned statements are harder to endorse (i.e. more difficult in IRT terminology) compared to *not* endorsing the socially encouraged statements of the reverse-scored items.

Another possible explanation for the variations observed in the ICU factor structure could be differential item functioning (DIF) across raters and age groups. In previous studies, the parent-report version has primarily been studied with children younger than 12 years, while the self-report version is most widely used in adolescent samples. The DIF between self- and parent-report on the standard- and reverse-scored ICU-items observed in a large sample of delinquent youth (Lin et al. 2018), could give rise to different factor structures of the measure across raters. Additionally, a meta-analysis of studies that supported bi-factor models of the ICU, found that the unique variance explained by the general CU-factor exceeded that of the specific bi-factors (Ray and Frick 2018). These findings suggest that despite the observed factor structures created by method variance effects, DIF and other sample-based sources of variance, the ICU primarily measures a core unidimensional CU construct. Additional research across samples and responder groups is needed to assess the empirical support for such a unidimensional conceptualization of the ICU.

Study Aims and Hypotheses

The primary aim of this study was to further investigate the structural validity of the ICU. Both the parent-, self- and teacher-report versions of the ICU were tested in a Norwegian mixed gender sample of adolescents with behavior problems. We hypothesized that all three versions of the Norwegian ICU represented a unidimensional construct of Callous-Unemotional Traits (Paiva-Salisbury et al. 2017; Ray et al. 2016; Ray and Frick 2018). In line with previous IRT-based research, we assumed that any factor structure found in the data would primarily be due to method variance resulting from the increased difficulty of agreeing with standard-scored items compared to disagreeing with reverse-scored items. Consequently, we expected that the addition of a method-variance element to the tested structural models of the ICU would increase model fit.

The secondary aim of this study was to assess the concurrent validity of the Norwegian translations of the ICU in the same sample. Although ICU has been studied in several other languages, psychological assessment tools should be re-assessed when translated and applied in other countries and contexts (de Vet et al. 2011). In particular, an instrument used to assess anti-social traits should be particularly closely assessed, as the consequences of measurement errors could potentially lead to wrongful labeling, stigma and

flawed treatment interventions (Edens et al. 2001). The Norwegian ICU has not yet been subject to solid psychometric validation, and this study aimed to address this problem by using a multi-informant design to assess both intra-, inter- and cross-rater validity. CU has been linked to increased and particularly recalcitrant behavior problems (Frick 2012), which in adolescents manifests itself as more pronounced aggressive, violent and criminal behavior (Frick et al. 2014; Frick and White 2008; Lawing et al. 2010). We therefore hypothesized the ICU to correlate positively with moderate strength to measures of externalizing problems, aggression, self-reported delinquency and problematic alcohol use. The increased level of anti-social acts in adolescents with CU is linked to a lack of prosocial emotions, guilt, fear and behavioral inhibition (Bjørnebekk 2007; Viding and Kimonis 2018), and we therefore hypothesized the ICU to show negative correlations of moderate strength to measures of anxiety and behavioral inhibition (Ciucci et al. 2014). As the association between CU and anxiety is potentially moderated and thereby masked by co-occurring conduct problems (Frick 2012), we controlled for the concurrent level of behavior problems in this analysis.

To assess the discriminative validity of the ICU we hypothesized negligible correlations of the ICU to self-ratings of subjective well-being, and to parent- and teacher-ratings of withdrawn/depressive symptoms and somatic complaints.

Methods

Participants

Data for this study came from a randomized controlled trial of Functional Family Therapy (FFT) in Norway (Ogden 2013). Families referred to Child Welfare Services for FFT-treatment were asked to participate in the trial. The inclusion criteria were adolescents aged 11–19 years who displayed or were at risk for one or more of the following behavior problems: delinquency, aggressive or violent behavior, verbal aggression or threats, truancy, school-related problem behavior and/or drug use in relation to problem behaviors mentioned above. The following exclusion criteria also applied: adolescents living by themselves, autism, acute psychotic episode, imminent risk of suicide, home environments that pose a threat to therapist life or safety, ongoing investigations by the local child welfare service, and concurrent services that were incompatible with commencing FFT-treatment.

We recruited 160 families to the study comprising 160 adolescents, 152 mothers (including eight step- and nine foster-mothers) and 90 fathers (including 15 step- and five foster-fathers). The participants were all informed of their right to later revoke their consent to the study. The adolescents in the

sample were 86 (53.7%) boys and 74 (46.3%) girls with a collective mean age of 14.7 years ($SD = 1.47$). The majority (77.5%) of the adolescents were on or above the culture-, age- and gender-appropriate clinical cut-off score on the externalizing scale of the Child Behavior Check List (CBCL; Achenbach and Rescorla 2007). A small fraction (8.8%) of the adolescents were in the borderline range, while the remaining (12.5%) scored in the normal range. CBCL-externalizing scores were missing for 2 (1.2%) adolescents as their parent(s) only partially completed the CBCL. Among the adolescents, 38.8% resided primarily with one biological parent, 25.5% with both biological parents, 24.8% with one biological parent and a step-parent, 6.4% in adoption- or foster-homes and 4.5% had joint physical custody arrangements. The mothers had a mean age of 43.6 years ($SD = 6.80$) and the average paternal age was 46.4 years ($SD = 7.35$). Most of the mothers (76.2%) and fathers (61.9%) were working full- or part-time, and 43.0% of the mothers and 34.4% of the fathers had a university or college degree.

Procedures

The data used in this study was collected prior to randomization to treatment condition in the trial. All questionnaires for parents and adolescents were programmed in the Ci3 software (Sawtooth Software n.d.). The family met with a research assistant in their home or at a municipality office to complete the questionnaires on portable computers provided by the research assistant. The research assistant gave general instructions on how to use the Ci3 system and was available for assistance during questionnaire completion. The family received a light snack and a minor monetary compensation (approximately 50 US Dollars) for taking the time to complete the questionnaires. When consenting to the trial, parents and adolescents were asked to give consent for the research assistant to contact the adolescent's teacher. If consent was obtained, teachers were contacted and asked to complete and mail back questionnaires on paper.

Measures

Callous-Unemotional Traits

The ICU is a 24-item questionnaire consisting of three standard- and three reverse-scored items derived from each of the four CU-items on the Antisocial Process Screening Devise (Kimonis et al. 2008). Examples of standard-scored items are “I seem very cold and uncaring to others” and “I do not care who I hurt to get what I want”. Examples of reverse-scored items are “I feel bad or guilty when I do something wrong” and “I try not to hurt others' feelings”. Items are rated on a 4-point scale: 0 (*not true at all*), 1 (*somewhat true*), 2 (*very true*), and 3 (*definitively true*). A total score is

calculated by adding all item scores together. The self-, parent- and teacher-report versions of the ICU were translated to Norwegian following general guidelines for questionnaire translation (de Vet et al. 2011). In this sample, the ICU means were 33.80 ($SD = 12.48$) for the parent-version, 29.40 ($SD = 10.70$) for the self-report version and 35.66 ($SD = 12.88$) for the teacher-version.

Externalizing Behavior

Parents and teachers in this study completed the widely used CBCL and Teacher Report Form (TRF; Achenbach and Rescorla 2001), respectively. Both questionnaires contain 120 items describing various child behaviors that are rated on a 3-point scale: 0 (*not true*), 1 (*somewhat or sometimes true*), and 2 (*very true or often true*). Raw scores were used to ensure analyses of the full range of variability on the scale and avoid loss of data resulting from the truncating process entailed in conversion to T-scores (Achenbach and Rescorla 2001). In this study the rule-breaking behavior and the aggressive behavior subscales showed acceptable reliability, $\alpha_s = .813-.921$, and were used as parent- and teacher-reported measures of externalizing behavior. In addition, parents were asked to report at what age the behavior problems of their child started, $M = 9.32$ years, $SD = 4.57$.

Self-Reported Delinquency

The Self-Reported Delinquency scale (SRD; Elliott et al. 1983) consists of items related to offences with a base rate above 1% as reported in the Uniform Crime Report (Elliott and Ageton 1980). In the present study, the adolescents reported how many times (from 0 to 9) in the past month they had engaged in each of the delinquent behaviors listed on the SRD. The total sum of all items on the SRD was used as a measure of the *amount* of delinquency, and this measure showed good reliability ($\alpha = .904$). A sum score of dichotomized answers (yes/no) on each of the SRD questions was calculated as a measure of the *diversity* of delinquency. Both the SRD-total and the SRD-diversity scores were truncated due to floor effects and skewed distributions. Therefore, both scores underwent Box-Cox transformation to allow for parametric analysis within a range of .15–.95 and .15–.98 of the z-scores, respectively.

Aggression

The Norwegian version of the Angry Aggression Scales (AAS; Bjørnebekk and Howard 2012) was used to assess four motivationally distinct types of aggression. The scales are based on Howard's (2011) quadripartite model of aggression (QVT) designating four types of aggression: explosive/reactive, vengeful/ruminative, thrill seeking and coercive. Items from

the explosive/reactive type of aggression include “Sometimes I get so angry that I don’t know what I’m doing”; items from the vengefulness type include “When I feel angry with somebody, I work out ways to get my own back”; items from the thrill seeking type include “When I make someone suffer I get ‘turned on’ and lose control”; and items from the coercive type include “Sometimes I use aggression to control others”. The four AAS subscales had acceptable reliability, with α s ranging from .835 (vengeful/ruminative) to .941 (thrill seeking). Floor effects and skewed distributions were observed for the subscales, which suggested the use of non-parametric tests.

Problematic Alcohol Use

The Alcohol Use Disorder Identification Test (AUDIT; Saunders et al. 1993) is a 10-item self-report measure of alcohol use. Each item is rated on a scale from 0 (*never*) to 4, where higher scores indicate more frequent alcohol use. Item scores are summarized to a total score ranging from 0 to 40. If the first AUDIT-question (“How often do you have a drink containing alcohol?”) was answered with a 0, all remaining items on the AUDIT were scored 0. This applied to 61.9% of our sample. A dichotomous variable indicating problematic alcohol use, was computed based on a cut-off score of 5 or above on the total AUDIT score (Liskola et al. 2018). This applied for 18.7% of our sample, with the highest observed AUDIT score being 24.

Anxiety

Anxious-depressed symptoms were measured by both self-, parent- and teacher-report. Adolescents completed the Symptom Checklist 10-item-version (SCL-10; Strand et al. 2003) where 4 items related to anxiety (e.g. “Feeling fearful”) and 6 items related to depression (e.g. “Feeling blue”) are rated on a 4-point scale from 1 (*not at all*) to 4 (*extremely*). The full scale SCL-10 score showed good reliability ($\alpha = .919$) and was used as self-report measure of anxiety. A cut-off score of 18.5 is proposed as representing symptoms of psychological problems (Strand et al. 2003), and 39.4% in our sample scored above this cut-off. Parent- and teacher-ratings of anxiety were obtained from the anxious-depressed subscale on the CBCL/TRF, which showed reliabilities of $\alpha = .865$ and $\alpha = .825$, respectively. When comparing the CBCL anxious-depressed scale scores to the multicultural, age and gender dependent cut-off scores, 37.5% of the sample were in the clinical range and 16.3% were in the borderline range.

Sensitivity to Punishment

The youth version of Carver and White’s BAS/BIS scales (Carver and White 1994) was used to assess sensitivity in the behavioral inhibition system (BIS). The BIS scale consists of 7 items that refer to the anticipation of punishment (e.g. “I worry

about making mistakes”). Items are rated on a 4-point scale: 4 (*very true for me*), 3 (*fairly true for me*), 2 (*partly true for me*), and 1 (*not at all true for me*). Empirical studies conducted by Bjørnebekk (2009) on normal populations of middle school children, and Bjørnebekk and Howard (2012) on a sample of adolescent offenders, have demonstrated the reliability and validity of the Norwegian youth version of these scales. The reliability of the BIS-scale in the present study was acceptable ($\alpha = .798$).

Measures Related to Discriminative Validity

Three measures were included to assess the discriminative validity of the ICU. Firstly, self-reported well-being was measured by the WHO-5 Well-being index (World Health Organization, Regional Office for Europe 1998). The WHO-5 scale consists of 5 items on well-being that are rated on a 6-point scale, ranging from 0 (*Never*) to 5 (*All of the time*). Within an adolescent sample a cutoff score of 9 or lower should advise further investigation of the patient (Allgaier et al. 2012). The scale showed acceptable reliability ($\alpha = .844$) in our sample, and 47.7% of the adolescents reported scores of 9 or below. Secondly, the withdrawn-depressed scales on the CBCL/TRF were used as discriminative validity measures towards withdrawn-depressed symptoms. Thirdly, the somatic complaints scale on the CBCL/TRF was also included to test for discriminative validity. The reliabilities of these four CBCL/TRF scales were acceptable (α s = .781–.795) apart from the TRF Somatic complaints scale that showed poor reliability ($\alpha = .596$).

Analyses

Missing Data

Maternal data was the primary source of parental data, and if missing, paternal data was used when available and complete. Multiple Imputation (MI) was applied within each responder group¹ to estimate missing values based on age, gender and the available scale scores. We ran 20 imputations, resulting in imputed data sets with 159 parental responses, 157 adolescent responses and 95 teacher responses² available for the analyses. Given the uneven number of participants in each respondent group, cross-informant analyses created missing data points, which were handled by pairwise deletion.

¹ Some missing data was observed in 14 (8.8%) parent-, 9 (5.6%) adolescent- and in 69 (43.1%) teacher-respondents. Data was completely missing for one parent-, three adolescent- and 62 teacher-respondents, for which multiple imputation was not used to estimate data. The lack of teacher-respondents was primarily due to lack of family consent or adolescent not attending school.

² To increase sample size when assessing the structural validity of the teacher-based ICU, we used post-treatment or follow-up scores, if baseline ICU was missing. This increased the sample size for this analysis to 114 (92 from baseline, 18 from post-treatment, 4 from follow-up).

Data Analysis Plan

We applied Confirmatory Factor Analysis (CFA) to assess the structural validity of the ICU. All CFA analyses were run in Mplus Version 8 (Muthén and Muthén 2017), where we tested the fit of different models of the ICU to our data. We chose weighted least squares estimation (WLSMV) for categorical indicators as the ICU uses a 4-point scale and the item response distributions in our sample rarely resembled a normal distribution. The fit of each model was assessed using the chi-square (X^2) fit statistic, the comparative fit index (CFI; Bentler 1990), the Tucker-Lewis index (TLI; McDonald and Marsh 1990), the root mean square error of approximation (RMSEA) and the weighted root mean square residual (WRMR). The CFI and TLI both indicate acceptable fit with values above .90 and good fit with values above .95 (Hu and Bentler 1999). RMSEA values between .05 and .08 indicate acceptable fit, whereas values below .05 are considered to represent good fit. The WRMR is considered acceptable when lower than 1.0 for both continuous and categorical variables (Yu 2002). The DIFFTEST function in Mplus was applied to assess the statistical significance of X^2 differences between nested models.

All remaining statistical analyses were conducted in SPSS version 25 (IBM Corporation 2017). In the identified best fitting model of the ICU, Cronbach's α s were calculated separately for each respondent group to analyze the reliabilities of the scales. Corrected item-total correlations (CITCs) were used to identify poorly functioning items, with the limit value set to .30 (Nunnally and Bernstein 1994). Inter- and cross-rater reliabilities were assessed by estimating a single measure, consistency, 2-way mixed effects model for intraclass correlations (ICC) and their 99% confidence intervals. Convergent validity of the ICU was investigated by its Pearson product-moment correlations to the hypothesized measures of relevance. Spearman's rank-order correlation was used for non-normally distributed variables. Partial correlation was applied when analyzing ICU subfactors to control for the effect of correlating factors. Independent sample t-tests were used to assess gender differences. To control for age and gender effects, logistic regression was applied to assess the influence of CU on the dichotomous variable of problematic alcohol use. The level of statistical significance was set to $\alpha = .01$, to adjust for multiple testing without being too restrictive when testing correlated hypotheses (Bender and Lange 2001). In the assessment of the concurrent validity, however, the strengths of the relationships were regarded as more relevant than the statistical significance of the results (de Vet et al. 2011). For partial correlation analyses, pooled p values are not computed in SPSS25 for MI-datasets, and we chose to restrictively report on the largest observed p value among the original and imputed datasets.

Results

Confirmatory Factor Analysis

Based on previous research on the parental- and self-report-versions of the ICU, four main models of the ICU were tested in a CFA: a 3-bifactor model (Essau et al. 2006; Roose et al. 2010; Waller et al. 2015), an IRT-shortened 2-factor model (S. W. Hawes et al. 2014b), a unidimensional model (Ray and Frick 2018) and an IRT-shortened unidimensional model (Ray et al. 2016). To account for any method-variance effect of standard- and reverse-scored items, the Unidimensional and 3-bifactor models were also tested with the addition of a method-variance bi-factor loading on all reverse-scored items (Paiva-Salisbury et al. 2017). This addition was not relevant for the IRT-shortened 2-factor model, as the Uncaring factor comprised all reversed-scored items. In the IRT-shortened Unidimensional model, eight of the ten items are reverse-scored, and it was therefore more parsimonious to test for method variance in this model by specifying shared method variance between the two standard-scored items on this scale.

The DIFFTEST function showed significant reductions of the X^2 test statistics for the models when method-variance was accounted for. This applied in both parent-reported data: 3-bifactor model ($\Delta X^2 = 53.15$, $df = 12$, $p < .001$), unidimensional model ($\Delta X^2 = 100.956$, $df = 12$, $p < .001$), and unidimensional short model ($\Delta X^2 = 5.194$, $df = 1$, $p = .023$, marginally significant); in self-reported data: 3-bifactor model ($\Delta X^2 = 182.669$, $df = 12$, $p < .001$), unidimensional model ($\Delta X^2 = 163.030$, $df = 12$, $p < .001$) and unidimensional short model ($\Delta X^2 = 60.269$, $df = 1$, $p < .001$); and in two of three models in the teacher-reported data: the 3-bifactor model ($\Delta X^2 = 63.900$, $df = 12$, $p < .001$) and the unidimensional model ($\Delta X^2 = 152.841$, $df = 12$, $p < .001$), but not the unidimensional short model ($\Delta X^2 = 1.793$, $df = 1$, $p = .181$). All other fit indices were also improved when a method-variance element was included in the model, again except for the unidimensional short model in the teacher-data. Subsequently, all analysis and model comparisons were made between models accounting for method-variance. Table 1, 2 and 3 show the fit-indices for these models when analyzing the ICU by parent-, self- and teacher-report, respectively. The fit-indices of the models without a method-variance element can be found in Online Resource 1.

Across all responder groups, the full unidimensional model displayed overall poor model fit. The IRT-shortened unidimensional model showed adequate fit in relation to the CFI, the TLI and the WRMR indices in both parent- and self-data, while these indices showed marginal fit in the teacher data. In all datasets however, the RMSEA indicated inadequate fit of this model. The 2-factor model showed good fit in parent, self and teacher data based on the CFI, the TLI and the WRMR indices. The RMSEA indicated marginal fit for this model

Table 1 Fit-indices in the parent-reported data of ICU models including a method variance element

	X^2 / df	CFI	TLI	RMSEA [90%CI]	WRMR
Unidimensional	695.671 / 240	.808	.779	.110 [.100–.119]	1.39
Unidimensional short	104.454 / 34	.932	.909	.115 [.090–.104]	0.97
2-factor	116.770 / 53	.956	.946	.088 [.066–.109]	0.80
3-bifactor	387.455 / 216	.928	.908	.071 [.059–.082]	0.91

Table 2 Fit indices in the self-reported data of ICU models including a method variance element

	X^2 / df	CFI	TLI	RMSEA [90% CI]	WRMR
Unidimensional	519.807 / 240	.842	.818	.087 [.077–.097]	1.17
Unidimensional short	116.791 / 34	.918	.891	.126 [.101–.151]	0.94
2-factor	109.768 / 53	.952	.940	.083 [.061–.105]	0.86
3-bifactor	408.400 / 216	.891	.861	.076 [.065–.087]	0.98

Table 3 Fit-indices in the teacher-reported data of ICU models including a method variance element

	X^2 / df	CFI	TLI	RMSEA [90%CI]	WRMR
Unidimensional	767.219 / 240	.817	.790	.139 [.128–.150]	1.52
Unidimensional short	189.327 / 34	.864	.820	.200 [.173–.228]	1.36
2-factor	134.161 / 53	.957	.947	.116 [.092–.140]	0.87
3-bifactor	481.843 / 217 ^a	.908	.883	.103 [.091–.116]	1.06

^a Due to convergence issues in model estimation of the 3-bifactor model using teacher data, item 15 on the Uncaring factor was assigned a specific starting value to enable convergence, hence the additional degree of freedom compared to the model tested in the parent- and youth-data.

across datasets. The 3-bifactor model showed adequate fit in the parent data and marginal fit in the self and teacher data.

The RMSEA indicator favors models with higher degrees of freedom, and in our study the 3-bifactor model with a method-variance factor by far has the highest degrees of freedom. This could explain why this model has the lowest RMSEA scores. Typically, it would require a larger sample size to have enough power to detect significant findings related to RMSEA scores on models such as the 2-factor model with 53 degrees of freedom (MacCallum et al. 1996). The CFI and TFI indices tend to be less sensitive to sample size. The 2-factor model was therefore assessed as the overall better fitting model for ICU across datasets, despite the slightly lower RMSEA values observed for the 3-bifactor model.

The interpretation of the 2-factor model in relation to method variance effects is not clear cut. The model embeds a method variance effect within the 2-factor structure with standard and reverse-scored items separately constituting each factor. Arguably, the 2-factor structure could be a method-variance artefact and could alternatively be modeled as a unidimensional CU construct constituting the 12 items, and a method variance bi-factor related to the five reverse-scored “Uncaring”-items. Post-hoc CFA-analysis of this alternative model yielded very similar model fit indices as for the 2-factor model in both parent- ($X^2 = 125.655$, $df = 51$, $CFI = .949$, $TLI = .934$,

$RMSEA = .097$, $90\% CI = [.075–.118]$, $WRMR = 0.87$), self- ($X^2 = 101.730$, $df = 51$, $CFI = .957$, $TLI = .944$, $RMSEA = .080$, $90\% CI = [.057–.103]$, $WRMR = 0.82$), and teacher-reported data ($X^2 = 130.869$, $df = 51$, $CFI = .958$, $TFI = .946$, $RMSEA = .117$, $90\% CI = [.093–.142]$, $WRMR = 0.87$). Thus, the results of our confirmatory factor analyses alone cannot make the distinction on whether the 2-factor model represents two interrelated factors, or a unidimensional construct with method-variance effects between items. Therefore, all subsequent analysis of convergent validity used both full scale and factor scores to enable a comparison of the scores’ correlational patterns.

Reliability and Subscale Correlations

Reliability analyses of the ICU12 were run separately for each respondent group. For the ICU12-P (parent-report), Cronbach’s α was .869 for the full scale, .816 for the 7-item Callousness subscale, and .791 for the 5-item Uncaring subscale. For the ICU12-S (self-report), the Cronbach’s α s were .810 (total scale), .848 (Callousness-subscale), and .811 (Uncaring-subscale). For the ICU12-T (teacher-report), the Cronbach’s α s were .901 (total scale), .855 (Callousness-subscale), and .839 (Uncaring-subscale). For all ICU12 versions, all items showed corrected item-total correlations (CITCs) > .30 to both the total

Table 4 Convergent validity of the ICU-scales to measures of behavior problems

	Aggression		Rule breaking		Age of Onset	Delinquency	
	CBCL	TRF	CBCL	TRF		Total	Diversity
ICU12-Parent	.475**	.196	.433**	.118	-.203	.270*	.212
-callousness	.290**	.080	.356**	.124	-.186	.117	.123
-uncaring	.163	.108	.032	-.021	.015	.125	.062
ICU12-Self	.007	.084	.162	.163	.064	.284*	.334**
-callousness	-.003	.182	.103	.149	.132	.303*	.270*
-uncaring	.013	-.106	.113	.046	-.074	.045	.156
ICU12-Teacher	.105	.392*	.219	.539**	.044	.248	.358*

ICU12 = 12-item Inventory of Callous Unemotional traits; CBCL = Child Behavior Checklist; TRF = Teacher Report Form. Italics are partial correlations controlling for the influence of the remaining factor of the respective ICU12 -questionnaire

* $p < .01$, ** $p < .001$

scale and its respective subscale. While the two subscales correlated strongly on the ICU12-P, $r(155) = .623, p < .001$, and the ICU12-T, $r(88) = .707, p < .001$, the correlation was weak and not statistically significant on the ICU12-S, $r(151) = .174, p = .031$. Due to the observed inter-correlation between the ICU12 subscales, partial correlation was applied when subsequently assessing the convergent validity of the subscale scores.

Cross- and Inter-Rater Reliability

The cross-rater reliability of the ICU12 was weak, both between parent- and self-report, ICC of .170, 99% CI [-.04, .37], $F(149, 149) = 1.41, p = .018$, and between teacher- and self-report, ICC of .226, 99% CI [-.05, .47], $F(85, 85) = 1.59, p = .018$. The cross-rater correlations between the ICU12-P and -S subscales were similarly weak for both the Callousness, ICC = .067, 99% CI [-.15, .27], $F(149, 149) = 1.14, p = .209$, and the Uncaring subscales, ICC = .141, 99% CI [-.07, .34], $F(150, 150) = 1.33, p = .042$. The inter-rater reliability of the ICU12 between parents and teachers was also weak, ICC = .204, 99% CI [-.07, .45], $F(87, 87) = 1.51, p = .028$.

Age and Gender Effects

While self-reported CU showed a small negative correlation to age, $r(155) = -.192, p = .017$, this was not observed for parent-report, $r(157) = .061, p = .446$, nor teacher-report, $r(93) = .045, p = .665$. The 99% CI for the gender difference in ICU12 scores between boys and girls was [-3.70, 2.43], $p = .592$ for parent-report, [0.73, 6.25], $p = .001$ for self-report and [-.50, 7.93], $p = .023$ for teacher-report. On average boys obtained scores that were approximately 0.5 SD higher on the ICU12-S and the ICU12-T when compared to girls, while they were 0.08 SD higher for boys on the ICU12-P.

Convergent, Divergent and Discriminative Validity

Externalizing Problems and Aggression

The results of the correlation analyses between parent, self- and teacher-reported CU and the main measures of externalizing problems are shown in Table 4. By parent- and teacher-report we found strong within-rater correlations between ICU and measures of aggression and rule breaking behavior. The cross-rater correlations were weak and for self- and teacher-reported CU only observed for the rule breaking scale. Age of onset had a weak negative correlation to the ICU-P only. Moderate correlations were seen between self-reported delinquency and the various ICU scores. On a subscale level, the associations appeared more marked for the Callousness subscale.

Spearman rank-order correlations showed the expected convergent validity of self-reported CU to coercive and thrill-seeking aggression, $r(155) = .286, p < .001$ and $r(155) = .285, p = .001$, respectively. The ICU12-S also had a small to moderate positive correlation to the Vengeful aggression scale, $r(155) = .224, p = .005$, while the correlation to the Explosive aggression scale was smaller, $r(155) = .137, p = .088$. Parent-reported CU showed small to negligible correlations to the aggression scales ($r_s = .035-.149, p_s = .065-.675$), while teacher-reported CU correlated somewhat with thrill-seeking aggression only, $r(91) = .226, p = .035$.

Logistic regression controlling for age and gender effects was applied to study the relationship between the ICU12-scores and indicated problematic alcohol use (AUDIT-score ≥ 5). Self- and teacher-reported, but not parent-reported, CU showed a significant association to problematic alcohol use: OR = 1.09, $p = .008$; OR = 1.11, $p = .008$, and OR = 1.04, $p = .173$, respectively. When both subscales were entered in the model instead of the total ICU12-S score, the Uncaring subscale had a higher odds ratio, OR: 1.13, $p = .045$, compared to the Callousness scale, OR: 1.07, $p = .101$.

Table 5 Relationship of the ICU-scales to measures of anxiety and behavioral inhibition

	SCL10 (Self) ^a	CBCL-A/D (Parent) ^a	TRF-A/D (Teacher) ^a	Behavioral Inhibition
ICU12-Parent	.001	-.187	-.038	.009
-callousness	.033	.034	.041	.100
-uncaring	-.035	-.228*	-.086	-.099
ICU12-Self	-.206	-.188	-.230	-.368**
-callousness	-.066	-.049	-.101	-.179
-uncaring	-.228	-.224	-.227	-.346**
ICU12-Teacher	-.234	-.140	-.208	-.329*

ICU12 = 12-item Inventory of Callous Unemotional traits; SCL-10 = 10-item Symptom Checklist; CBCL = Child Behavior Checklist; TRF = Teacher Report Form. A/D = Anxious-depressed subscale. a) partial correlations controlling for the CBCL externalizing scores. Numbers in italics are partial correlations controlling for the effects of the remaining factor of the respective ICU12-questionnaire

* $p < .01$, ** $p < .001$

Anxiety and Punishment Sensitivity

The partial correlations of the ICU-scores to measures of anxiety and punishment sensitivity when controlling for concurrent levels of externalizing problems are seen in table 5. While the ICU12-S and the ICU12-T showed some small to moderate negative correlations to both within- and cross-rater measures of anxiety and punishment sensitivity, the ICU12-P only demonstrated a within-rater negative relationship to anxiety. On a subscale level, the associations were observed mostly for the Uncaring subscale.

Discriminant Validity

Overall, the ICU12-S demonstrated discriminative validity to the WHO-5 Well-being index and the withdrawn-depressed and somatic complaints scales on the CBCL/TRF ($r_s = -.090-.034$, $p_s = .336-.918$). Both the ICU12-P and ICU12-T showed discriminative validity to subjective well-being and parent- and teacher-reports of somatic complaints ($r_s = -.146-.091$, $p_s = .088-.996$), but moderate strength positive correlations to within-rater withdrawn-depressed scales were observed, $r(157) = .323$, $p < .001$, and $r(93) = .415$, $p < .001$, respectively.

Discussion

This study sought to assess the structural validity of the ICU and the concurrent validity of the Norwegian ICU translation in a sample of adolescents with behavior problems. In line with previous research, our results indicated a method variance effect of the standard- and reverse-scored items of the ICU (Paiva-Salisbury et al. 2017; Ray et al. 2016). However, our test of models including this method variance, failed to yield strong support for the ICU as a unidimensional measure

of CU. Neither the full nor the IRT-shortened unidimensional model provided the best model fit in our data. The overall best fit was observed for the 12 item 2-factor model of the ICU across all responder groups. This is in line with some previous studies of the ICU self-report version in adolescent samples (Carvalho et al. 2018; Colins et al. 2015; Pechorro et al. 2016a). These findings do not fully undermine the idea of CU as unidimensional as the ICU12 2-factor model can be remodeled as a unidimensional construct with a method variance bi-factor related to reverse-scoring items. Our post-hoc analyses showed that this would result in similar levels of model fit. Therefore, we will discuss the concurrent validity of the ICU total score across responder groups, before subsequently discussing whether the two factors can be meaningfully justified by any distinct pattern of correlations to the concurrent measures.

Cross-Rater Agreement and Concurrent Validity

The low levels of cross-rater agreement observed between parents, adolescents and teachers on the ICU12, were similar to or slightly weaker than those reported in previous studies (Berg et al. 2013; Gao and Zhang 2015; R. D. Lutzman et al. 2013; Levy et al. 2017; Roose et al. 2010; White et al. 2009). Low levels of parent-child agreement on mental health measures could reflect different responder perspectives (Hemmingsson et al. 2016), as observed by differential item functioning across observer- (parent/teacher) and self-report on the ICU (Lin et al. 2018). These findings indicate that the ICU12 might not measure the same construct across respondents, and caution against relying on single-informant assessment of CU.

This study found support for convergent and divergent concurrent validity of the ICU12 when using within-rater measures, as commonly used in previous research on the ICU (Ciucci et al. 2014; Fanti et al. 2009; Kimonis et al.

2008; Pechorro et al. 2016a; Pihet et al. 2014). The ICU12-P showed convergent validity to parent-reported aggression, rule-breaking behavior and age of onset and divergent validity to parent-reported anxiety, with similar strength of the associations as seen in previous studies (Berg et al. 2013; Levy et al. 2017). The ICU12-S showed convergent validity to self-reported aggression, delinquency and problematic alcohol use, and divergent validity to self-reported anxiety and punishment sensitivity. Similarly, the concurrent validity of ICU12-T was supported by expected associations to teacher-reported externalizing behaviors and anxiety. The ICU12 versions also showed appropriate discriminant validity, apart from the within-rater correlations to the CBCL/TRF withdrawn-depressed scale, possibly caused by shared method variance (Gao and Zhang 2015). These intra-rater results for the hypothesized relationships and the divergent validity support the construct validity of the ICU12 within each respondent group.

Consistent with previous multi-informant validation studies of the ICU, we found fewer and weaker cross- and inter-rater correlations between the ICU12 scores and hypothesized convergent constructs (Berg et al. 2013; Docherty et al. 2016; Lin et al. 2018; Roose et al. 2010). This might be related to the relatively weak inter- and cross-rater reliability of the ICU12 (Docherty et al. 2016; Gao and Zhang 2015; Lin et al. 2018), and the lack of shared method-variance that might inflate intra-rater correlations (Gao and Zhang 2015). Demonstrating cross-rater correlations for convergent validity could therefore point to particularly salient relationships between constructs.

In this study, cross-rater construct validity was most evident between self-reported delinquency and parent- and teacher-reported CU. This applied for both the total delinquency score and the diversity of delinquency score. These cross-rater relationships indicate how CU specifically relates to a higher likelihood of criminal behavior among adolescents with behavior problems (Frick et al. 2014; Frick and White 2008). The strength of the cross-rater relationships was similar to the within-rater correlations, indicating that parent-, teacher- and self-reported CU can account equally well for retrospectively self-reported delinquency. A prospectively designed study found parent-reported CU to be better at predicting youth detained status (Docherty et al. 2016), which also provides an argument for assessing CU in youth not only based on self-report.

For the remaining hypothesized correlates only some cross-rater support for the concurrent validity of the ICU was observed. The ICU12-S had a small correlation to parent- and teacher-ratings of rule-breaking behavior, but none to parents' and teachers' aggression ratings. This contrasts previous studies using these CBCL scales where CU has been more strongly associated to aggression than to rule-breaking (Benesch et al. 2014; Berg et al. 2013), or the associations have been

equal in strength (Pihet et al. 2014; Waller et al. 2015). The ICU12-S also showed the hypothesized negative relationship to both parent- and teacher-ratings of adolescent anxiety. Overall, the concurrent validity of the ICU12-S was supported by cross-rater relationships.

When comparing the cross-informant concurrent validity of the ICU12-P and the ICU12-T to self-reported aggression, alcohol use, anxiety and fear sensitivity, the ICU12-T displayed small correlations in the hypothesized direction, while the ICU12-P did not. A previous study found a combined teacher-parent CU-measure to have cross-rater relationships to self-reported fear sensitivity (Roose et al. 2010), but it is not known if this result would have been found by using parental CU-scores only. The lower ability of parent-reported CU to gain cross-rater support for its validity compared to teacher-reported CU, could indicate that teachers might have a more objective and less biased appraisal when assessing a potentially stigmatizing construct as CU. Another possible reason for the ICU12-P to not be related to self-reported anxiety and punishment insensitivity, is the weak cross-rater agreement on the Uncaring subscale, which is the subscale of the ICU12-S most strongly related to anxiety and fear sensitivity. If the effects of differential item functioning on the ICU lead parents to assess a CU construct containing less of the Uncaring aspects experienced by the adolescent, this might then result in a parent-based CU measure with a less clear relationship to self-reported anxiety and fear sensitivity.

The fact that the concurrent validity of the ICU12 was only partially supported by cross-rater measures, could indicate biases in single rater assessment of CU. While adolescents might underreport, be unaware of or strategically report on a negative character trait like CU (Levy et al. 2017), parents and teachers can only infer about the internal states of their child/student and might thus be unaware or biased in their assessments. In addition, differential item functioning between parent- and self-report suggests that some ICU items are read, interpreted and rated differently by parent and adolescents (Lin et al. 2018). Although general recommendations on mental health assessments, supporting the use of comprehensive multi-informant data, might limit the risk of biased single source scores in CU-assessments (Gao and Zhang 2015; Hemmingsson et al. 2016), there could still be a concern for the degree to which the ICU12 versions, individually or combined, appropriately measure CU.

Unidimensional vs Two Factors

Across raters, this study found the subscales of the ICU12 to show distinctive patterns of correlation to the measures of convergent validity. The Callousness subscale generally showed stronger associations to aggression, rule breaking behavior and delinquency in our study, in line with studies linking this subscale to rule-breaking behavior (Colins et al.

2015), criminal behavior (Colins et al. 2015; Lin et al. 2018; Pechorro et al. 2016b) and proactive aggression (Fanti et al. 2009; Pihet et al. 2014). The increased item difficulty of this scale (i.e. the higher thresholds for endorsing these standard-scored items), could make it more informative within a group of at-risk adolescents. The Uncaring subscale, on the other hand, was more strongly related to lower levels of anxiety and punishment sensitivity as seen elsewhere (Colins et al. 2015). These findings support a two factors representation of the ICU12.

However, findings from other studies are often more mixed, as some studies link the Callousness scale to lower punishment sensitivity, and the Uncaring scale aggression and externalizing problems (Ciucci et al. 2014; Colins et al. 2015; Fanti et al. 2009). Our results also showed that while the subscales did show distinct correlation patterns, they rarely indicated correlations not observed for the full scale ICU12. So, while a factor structure can be observed in the ICU12 due to varying item difficulty and item wording (standard- or reverse-scored), and a distinct correlational pattern of the factors can be seen, the full scale ICU12 unidimensional score might in and of itself be an appropriate measurement of the CU construct.

Assessing Lack of Prosocial Emotions

Concern could be raised about whether the ICU12 appropriately assesses a lack of prosocial emotions. The ICU12 contains only one item from the Unemotional subscale which has shown negative associations to emotional empathy and emotional responses to distress (Berg et al. 2013; Kimonis et al. 2008). Post-hoc analysis of our data showed that the Unemotional subscale was unrelated to self-reported delinquency, punishment-sensitivity and aggression and showed some *negative* correlations to parent- and teacher-reported externalizing problems, see Online Resource 2. Additionally, the self-reported Unemotional subscale showed significant *positive* correlations to both self- and parent-reported levels of anxious-depressed symptoms.

These findings suggest that the Unemotional subscale is more related to negative emotions in general, than to lack of prosocial emotionality specifically (Berg et al. 2013). Several other studies point to weaknesses of this subscale (Carvalho et al. 2018; Ciucci et al. 2014; R. D. Latzman et al. 2013; Waller et al. 2015), and a meta-analytic study found it to lack appropriate reliability and construct validity (Cardinale and Marsh 2020). While the current findings indicate that the *Uncaring* factor might cover some aspects of lowered emotionality related to CU, the call made for rewriting the Unemotional items to specifically capture lack of prosocial

emotionality is warranted, as this could improve the ICU as an assessment of CU (Waller et al. 2015).

Furthermore, the value of assessing psychopathy in children and adolescents more broadly than CU, should not be forgotten (Andershed et al. 2018; Colins et al. 2018; Salekin 2017). Recent studies have shown that the inclusion of measures from other domains of psychopathy leads to better predictions (Andershed et al. 2018; Colins et al. 2018). Within a framework of cumulative risk factors, this equals the notion of improved risk assessment based on the number of risk factors measured. This does not necessarily call for an abandonment of further use and development of CU-measures when assessing adolescents with serious behavior problems. It does, however, suggest that more detailed and multi-faceted assessments within the broader psychopathic domain might increase our understanding of the multiple factors that contribute to persistent anti-sociality. Similarly, in line with work on psychopathy in adults, it could be relevant to include the multiple dimensions of psychopathy as specifiers for conduct disorder in the DSM-V and ICD-11 diagnostic systems (American Psychiatric Association 2013; Salekin 2016; World Health Organization 2018). Future research should also prioritize to study whether multiple-informant measures of the various psychopathy dimensions can assist in selecting optimal treatment interventions for children and adolescents with severe behavior problems.

Limitations

Although this study has strengths in evaluating the ICU using a multi-informant design with several validated measures in an at-risk adolescent population, it has some limitations worth mentioning. Firstly, the sample size of this study is somewhat small, as study recruitment did not reach the goal of 250 participants but was terminated at 160 participants after four years of recruiting. The sample size creates a risk of lacking power to gain statistical significance for observed relationships and is a barrier to applying Item Response Theory to study item difficulty, item discriminative values and differential item functioning. Secondly, the data used in this study was cross-sectional and does not relay information regarding directionality of the associations found. Thirdly, the data in this study is based on questionnaire responses and did not include observational, neurological or biological data nor a criterion related measure of CU or the Conduct Disorder diagnosis. Lastly, in relation to the Unemotional subscale, this study is limited by not including a more focused measure of shallow and flattened affect or a measure of emotional empathy that might have been more relevant for the concurrent validity of this subscale.

Acknowledgements Open Access funding provided by NORCE Norwegian Research Centre AS. We also want to acknowledge the families and teachers participating in this study.

Funding Information This study was funded by TrygFonden (grant ID-number 120499) and the Norwegian Center for Child Behavioral Development.

Compliance with Ethical Standards

Conflict of Interest Dagfinn Mørkrød Thøgersen, Mette Elmose and Gunnar Bjørnebekk declare that they have no conflict of interest.

Ethical Approval Ethical approval for the study was given by The National Committee for Medical and Health Research Ethics (NEM) on 02.11.2010 and 02.10.2012 (Reference number 2010/497).

Informed Consent Informed consent was obtained from all individual participants included in the study.

Experiment Participants All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

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

References

- Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment*. Burlington: ASEBA.
- Achenbach, T. M., & Rescorla, L. (2007). Multicultural supplement to the manual for the ASEBA School-age Forms & Profiles. Aseba.
- Allgaier, A.-K., Pietsch, K., Frühe, B., Prast, E., Sigl-Glöckner, J., & Schulte-Körne, G. (2012). Depression in pediatric care: Is the WHO-five well-being index a valid screening instrument for children and adolescents? *General Hospital Psychiatry, 34*(3), 234–241. <https://doi.org/10.1016/j.genhosppsych.2012.01.007>.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed.). Arlington: American Psychiatric Association.
- Andershed, H., Colins, O. F., Salekin, R. T., Lordos, A., Kyranides, M. N., & Fanti, K. A. (2018). Callous-unemotional traits only versus the multidimensional psychopathy construct as predictors of various antisocial outcomes during early adolescence. *Journal of Psychopathology and Behavioral Assessment, 40*(1), 16–25. <https://doi.org/10.1007/s10862-018-9659-5>.
- Bender, R., & Lange, S. (2001). Adjusting for multiple testing—When and how? *Journal of Clinical Epidemiology, 54*(4), 343–349.
- Benesch, C., Gortz-Dorten, A., Breuer, D., & Dopfner, M. (2014). Assessment of callous-unemotional traits in 6 to 12 year-old children with oppositional defiant disorder/conduct disorder by parent ratings. *Journal of Psychopathology and Behavioral Assessment, 36*(4), 519–529. <https://doi.org/10.1007/s10862-014-9420-7>.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246.
- Berg, J. M., Lilienfeld, S. O., Reddy, S. D., Latzman, R. D., Roose, A., Craighead, L. W., et al. (2013). The inventory of callous and unemotional traits. *Assessment, 20*(5), 532–544. <https://doi.org/10.1177/1073191112474338>.
- Bjørnebekk, G. (2007). Dispositions related to sensitivity in the neurological basis for activation of approach-avoidance motivation, antisocial attributes and individual differences in aggressive behavior. *Social Behavior and Personality: An International Journal, 35*(9), 1251–1264. <https://doi.org/10.2224/sbp.2007.35.9.1251>.
- Bjørnebekk, G. (2009). Psychometric properties of the scores on the behavioral inhibition and activation scales in a sample of Norwegian children. *Educational and Psychological Measurement, 69*(4), 636–654. <https://doi.org/10.1177/0013164408323239>.
- Bjørnebekk, G., & Howard, R. (2012). Validation of a motivation-based typology of angry aggression among antisocial youths in Norway. *Behavioral Sciences & the Law, 30*(2), 167–180. <https://doi.org/10.1002/bsl.2007>.
- Byrd, A. L., Hawes, S. W., Burke, J. D., Loeber, R., & Pardini, D. A. (2017). Boys with conduct problems and callous-unemotional traits: Neural response to reward and punishment and associations with treatment response. *Developmental Cognitive Neuroscience, 30*, 51–59. <https://doi.org/10.1016/j.dcn.2017.12.004>.
- Cardinale, E. M., & Marsh, A. 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>.
- Carvalho, M., Faria, M., Conceição, A., de Matos, M. G., & Essau, C. A. (2018). Callous-unemotional traits in children and adolescents. *European Journal of Psychological Assessment, 34*(2), 101–110. <https://doi.org/10.1027/1015-5759/a000449>.
- Carver, C. S., & White, T. L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality and Social Psychology, 67*(2), 319–333. <https://doi.org/10.1037/0022-3514.67.2.319>.
- 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>.
- Colins, O. F., Andershed, H., Hawes, S. W., Bijttebier, P., & Pardini, D. A. (2015). Psychometric properties of the original and short form of the inventory of callous-unemotional traits in detained female adolescents. *Child Psychiatry & Human Development, 47*(5), 679–690. <https://doi.org/10.1007/s10578-015-0601-8>.
- Colins, O. F., Andershed, H., Salekin, R. T., & Fanti, K. A. (2018). Comparing different approaches for subtyping children with conduct problems: Callous-unemotional traits only versus the

- multidimensional psychopathy construct. *Journal of Psychopathology and Behavioral Assessment*, 40(1), 6–15. <https://doi.org/10.1007/s10862-018-9653-y>.
- de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in medicine*. Cambridge University Press.
- Docherty, M., Boxer, P., Huesmann, L. R., O'Brien, M., & Bushman, B. (2016). Assessing callous-unemotional traits in adolescents: Determining cutoff scores for the inventory of callous and unemotional traits. *Journal of Clinical Psychology*, 73(3), 257–278. <https://doi.org/10.1002/jclp.22313>.
- Edens, J. F., Skeem, J. L., Cruise, K. R., & Cauffman, E. (2001). Assessment of “juvenile psychopathy” and its association with violence: A critical review. *Behavioral Sciences & the Law*, 19(1), 53–80. <https://doi.org/10.1002/bsl.425.abs>.
- Elliott, D. S., & Ageton, S. S. (1980). Reconciling race and class differences in self-reported and official estimates of delinquency. *American Sociological Review*, 45(1), 95–110. <https://doi.org/10.2307/2095245>.
- Elliott, D., Ageton, S., Huizinga, D., Knowles, B., & Canter, R. (1983). *The prevalence and incidence of delinquent behavior: 1976–1980 the National Youth Survey Project Report no. 26*. Boulder: Behavioral Research Institute.
- 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>.
- 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>.
- Frick, P. J. (2004). *Inventory of callous–unemotional traits*. Unpublished rating scale.
- Frick, P. J. (2012). Developmental pathways to conduct disorder: Implications for future directions in research, assessment, and treatment. *Journal of Clinical Child & Adolescent Psychology*, 41(3), 378–389.
- 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*, 49(4), 359–375. <https://doi.org/10.1111/j.1469-7610.2007.01862.x>.
- Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychological Bulletin*, 140(1), 1–57.
- Gao, Y., & Zhang, W. (2015). Confirmatory factor analyses of self- and parent- report inventory of callous-unemotional traits in 8- to 10-year-olds. *Journal of Psychopathology and Behavioral Assessment*, 38(3), 331–340. <https://doi.org/10.1007/s10862-015-9527-5>.
- Hawes, D. J., Price, M. J., & Dadds, M. R. (2014a). Callous-unemotional traits and the treatment of conduct problems in childhood and adolescence: A comprehensive review. *Clinical Child and Family Psychology Review*, 17(3), 248–267. <https://doi.org/10.1007/s10567-014-0167-1>.
- Hawes, S. W., Byrd, A. L., Henderson, C. E., Gazda, R. L., Burke, J. D., Loeber, R., & Pardini, D. A. (2014b). Refining the parent-reported inventory of callous–unemotional traits in boys with conduct problems. *Psychological Assessment*, 26(1), 256–266. <https://doi.org/10.1037/a0034718>.
- Hemmingson, H., Ólafsdóttir, L. B., & Egilson, S. T. (2016). Agreements and disagreements between children and their parents in health-related assessments. *Disability and Rehabilitation*, 39(11), 1059–1072. <https://doi.org/10.1080/09638288.2016.1189603>.
- Howard, R. C. (2011). The quest for excitement: A missing link between personality disorder and violence? *Journal of Forensic Psychiatry & Psychology*, 22(5), 692–705.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>.
- IBM Corporation. (2017). *IBM SPSS statistics for Macintosh, version 25.0*. Armonk: IBM Corp.
- Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., et al. (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>.
- Latzman, R. D., Lilienfeld, S. O., Latzman, N. E., & Clark, L. A. (2013). Exploring callous and unemotional traits in youth via general personality traits: An eye toward DSM-5. *Personality Disorders*, 4(3), 191–202. <https://doi.org/10.1037/a0000001>.
- Lawing, K., Frick, P. J., & Cruise, K. R. (2010). Differences in offending patterns between adolescent sex offenders high or low in callous-unemotional traits. *Psychological Assessment*, 22(2), 298–305. <https://doi.org/10.1037/a0018707>.
- Levy, T., Apter, A., Djalovski, A., Peskin, M., Fennig, S., Gat-Yablonski, G., et al. (2017). The reliability, concurrent validity and association with salivary oxytocin of the self-report version of the inventory of callous-unemotional traits in adolescents with conduct disorder. *Psychiatry Research*, 256(supplement C), 124–129. <https://doi.org/10.1016/j.psychres.2017.06.028>.
- Lin, B., Kerig, P. K., & Adkins, D. E. (2018). Cross-validation of the inventory of callous-unemotional traits across reporters and genders in a sample of detained youth. *Psychological Assessment*, 31(1), 15–26. <https://doi.org/10.1037/pas0000636>.
- Liskola, J., Haravuori, H., Lindberg, N., Niemelä, S., Karlsson, L., Kiviruusu, O., & Marttunen, M. (2018). AUDIT and AUDIT-C as screening instruments for alcohol problem use in adolescents. *Drug and Alcohol Dependence*, 188, 266–273. <https://doi.org/10.1016/j.drugalcdep.2018.04.015>.
- MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1(2), 130–149. <https://doi.org/10.1037/1082-989X.1.2.130>.
- McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, 107, 247–255.
- Muthén, L. K. & Muthén, B.O. (2017). *Mplus User's Guide*. Eighth Edition. Los Angeles: Muthén & Muthén. Retrieved from: http://www.statmodel.com/download/usersguide/MplusUserGuideVer_8.pdf
- Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory*, (3rd ed). New York: McGraw-Hill.
- Ogden, T. (2013). Evaluation of functional family therapy (FFT) in Norway. *BMC / Springer Nature*. <https://doi.org/10.1186/ISRCTN58861782>.
- Paiva-Salisbury, M. L., Gill, A. D., & Stickle, T. R. (2017). Isolating trait and method variance in the measurement of callous and unemotional traits. *Assessment*, 24(6), 763–771.
- Pechorro, P., Hawes, S. W., Gonçalves, R. A., & Ray, J. V. (2016a). 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(3), 221–239. <https://doi.org/10.1080/1068316X.2016.1239724>.
- Pechorro, P., Ray, J. V., Barroso, R., Maroco, J., & Abrunhosa Gonçalves, R. (2016b). 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(3), 349–365.
- Pihet, S., Etter, S., Schmid, M., & Kimonis, E. R. (2014). Assessing callous-unemotional traits in adolescents: Validity of the inventory of callous-unemotional traits across gender, age, and community/

- institutionalized status. *Journal of Psychopathology and Behavioral Assessment*, 37(3), 407–421. <https://doi.org/10.1007/s10862-014-9472-8>.
- Ray, J. V., & Frick, P. J. (2018). Assessing callous-unemotional traits using the Total score from the inventory of callous-unemotional traits: A meta-analysis. *Journal of Clinical Child and Adolescent Psychology*, 1–10. <https://doi.org/10.1080/15374416.2018.1504297>.
- Ray, J. V., Frick, P. J., Thornton, L. C., Steinberg, L., & Cauffman, E. (2016). Positive and negative item wording and its influence on the assessment of callous-unemotional traits. *Psychological Assessment*, 28(4), 394–404.
- Roose, A., Bijttebier, P., Decoene, S., Claes, L., & Frick, P. J. (2010). Assessing the affective features of psychopathy in adolescence: A further validation of the inventory of callous and unemotional traits. *Assessment*, 17(1), 44–57. <https://doi.org/10.1177/1073191109344153>.
- Salekin, R. T. (2016). Psychopathy in childhood: Why should we care about grandiose-manipulative and daring-impulsive traits? *The British Journal of Psychiatry*, 209(3), 189–191. <https://doi.org/10.1192/bjp.bp.115.179051>.
- Salekin, R. T. (2017). Research review: What do we know about psychopathic traits in children? *Journal of Child Psychology and Psychiatry*, 58(11), 1180–1200. <https://doi.org/10.1111/jcpp.12738>.
- Saunders, J. B., Aasland, O. G., Babor, T. F., Grant, M., & Fuente, J. R. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption-II. <http://www.who.int/iris/handle/10665/51167>.
- Sawtooth Software (n.d.). Ci3 for windows. Sequim, WA, USA. Retrieved from <http://www.sawtoothsoftware.com/support/>.
- Strand, B. H., Dalgard, O. S., Tambs, K., & Rognerud, M. (2003). Measuring the mental health status of the Norwegian population: A comparison of the instruments SCL-25, SCL-10, SCL-5 and MHI-5 (SF-36). *Nordic Journal of Psychiatry*, 57(2), 113–118. <https://doi.org/10.1080/08039480310000932>.
- Viding, E., & Kimonis, E. R. (2018). Callous-unemotional traits. In C. J. Patrick (Ed.), *Handbook of psychopathy, second edition (2nd ed.)*. New York: Guilford Publications.
- Waller, R., Wright, A. G. C., Shaw, D. S., Gardner, F., Dishion, T. J., Wilson, M. N., & Hyde, L. W. (2015). Factor structure and construct validity of the parent-reported inventory of callous-unemotional traits among high-risk 9-year-olds. *Assessment*, 22(5), 561–580. <https://doi.org/10.1177/1073191114556101>.
- 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>.
- World Health Organization. (2018). *International Classification of Diseases for Mortality and Morbidity Statistics (11th Revision)*. Retrieved from <https://icd.who.int/browse11/l-m/en>.
- World Health Organization, Regional Office for Europe (1998). *Use of Well-being Measures in Primary Health Care - the DepCare Project Health for All. Target 12*. E60246. Geneva: WHO.
- Yu, C.-Y. (2002). *Evaluating Cutoff Criteria of Model Fit Indices for Latent Variable Models with Binary and Continuous Outcomes*. Retrieved from <http://www.statmodel.com/download/Yudissertation.pdf>.

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