

Confirmatory Factor Analyses of Self- and Parent- Report Inventory of Callous-Unemotional Traits in 8- to 10-Year-Olds

Yu Gao^{1,2} · Wei Zhang²

Published online: 14 November 2015
© Springer Science+Business Media New York 2015

Abstract The factorial structure of the Inventory of Callous-Unemotional Traits (ICU; Frick 2004) has been inconsistent across various studies using different rater versions and samples. In this study the factorial structure of the self- and parent-report versions of the ICU was examined in a community sample of 340 boys and girls aged 8- to 10-years. A series of confirmatory factor analyses revealed that (1) for self-report ICU, a two-factor model comprising one Uncaring and one Callousness factor fit the data best; the factorial structure was invariant across sex, although boys showed higher scores than girls on the Uncaring factor; (2) for the parent-report ICU, a three-factor structure was supported, and boys and girls did not differ on factor structure or levels of the factors. For both versions, CU traits were positively associated with the count of Oppositional Defiant Disorder/Conduct Disorder and internalizing behavior symptoms, demonstrating ICU's good convergent albeit unsatisfactory discriminant validity. Findings suggest further refinement of this instrument for the self-report and parent versions separately.

Keywords Callousness · Uncaring · Unemotional · Child · Parent · Factor structure · Validity

In recent years interests in understanding the early precursor of adult psychopathy have been increasing. Among the several dimensions emerged in the construct of adult psychopathy, Callous/Unemotional traits (CU), referring to specific deficiencies in affective experience and interpersonal style, are believed to be the core features of childhood and adolescent psychopathy (Frick and White 2008). These traits are important for designating a distinct subgroup of antisocial and delinquent children and adolescents who show more severe, stable, and instrumental aggressive pattern of behavior (e.g., Essau et al. 2006). These individuals are at higher risk for early onset delinquency and later antisocial behavior (Frick and White 2008), and also show relatively poorer response to treatment (e.g., Frick and Dickens 2006). Finally, longitudinal studies have shown that the CU traits are relatively stable from late childhood to early adolescence (e.g., Munoz and Frick 2007). Therefore it is important to assess these traits reliably and validly.

The Inventory of Callous and Unemotional Traits (ICU) was developed to specifically assess the CU traits in youth (Frick 2004). It includes 24 items coded on a 4-point Likert scale, ranging from 0 (not at all true) to 3 (definitely true). There are Youth Self-Report, Parent Report, Teacher Report, Parent Report (Preschool), and Teacher Report (Preschool) versions. Studies have generally indicated that there are three aspects of CU traits: Callousness, capturing behavior that includes a lack of empathy, guilt, and remorse; Uncaring, indicating a lack of caring about one's performance in tasks and for others' feelings; and Unemotional, representing a lack of emotional expression (e.g., Essau et al. 2006).

These three dimensions have generally been supported by different samples using different translations (e.g., Byrd et al. 2013; Ezpeleta et al. 2013; Kimonis et al. 2013). For example, Essau et al. (2006) using an exploratory factor analysis in a sample of 1443 13–18-year-old Germans from general

✉ Yu Gao
yugao@brooklyn.cuny.edu

Wei Zhang
wzhang2@gradcenter.cuny.edu

¹ Department of Psychology, Brooklyn College, 5401 James Hall, 2900 Bedford Avenue, Brooklyn, NY 11210, USA

² Graduate Center of the City University of New York, New York, NY, USA

population for the first time identified these three factors, plus a common general dimension comprising all the items (i.e., bifactor model) of the self-report version, and the factor structure was invariant for boys and girls. Kimonis et al. (2008) confirmed this factor structure with the self-report version in a sample of 248 12- to 20-year-old American juvenile offenders. However, in their model two Callousness items (item #2 and #10) were deleted due to low item-total correlations, and the internal consistency for the Unemotional factor was low (0.53). Ciucci et al. (2014) in 540 Italian children (mean age = 12.7 years) confirmed the bifactor model with self-report version and the factor structure was invariant across sex and grade.

Recently a few studies raised questions about the bifactor model and validity of scales (Benesch et al. 2014; Hawes et al. 2014; Willoughby et al. 2015). For example, Willoughby et al. (2015) examined the parent ratings of the ICU in 1078 American first-graders (mean age = 7.3 years) and found a two-factor model that distinguished “callous-unemotional” (including nine Callousness and two Unemotional items from the original scale) and “empathic-prosocial” (including eight Uncaring, two Callousness, and three Unemotional items) dimensions to be the best fit, and none of the items were cross-loaded on more than one factor. Hawes et al. (2014) in a sample of 250 American boys with significant conduct problems (ages 6–12 years) obtained a two-factor structure (Callousness and Uncaring) using 12 of the original 24 items of the parent-report version. Houghton et al. (2013a) examined the factor structure of self-report ICU in 268 Australian children aged 7- to 12- years and found that a two-factor model was superior to the three-factor model. In this two-factor model, all Unemotional items were deleted due to poor model fit and poor internal reliability, and the Callousness and Uncaring factors each consisted of 8 items with internal reliabilities of 0.77 and 0.85, suggesting satisfactory reliability. In addition, eight pairs of error terms were correlated to improve model fit.

In summary, the bifactor model (with three intercorrelated factors and an overarching general dimension) has obtained the strongest support from studies focusing on adolescent and adult samples (above 12 years), whereas in children aged 7–12 years a two-factor structure (without the overarching general dimension) seems to be a better fit. However, the factor structure of the ICU varies for different raters and various samples, and in most studies either the self-report or parent-report was investigated (although Roose et al. (2010) examined both). In Roose et al.’s study, the same bifactor structure was supported for self, parent, and teacher versions of the ICU in a community sample of adolescents. However, for the convergent and criterion validity analyses, the authors only reported results for a teacher/parent composite rating, utilizing the highest score on each item from these different raters. It is generally believed that different informants are likely to report different aspects of behavior, and children are less capable

than adults as informants for disruptive behavior problems (Loeber et al. 1991). However, it can also be argued that children are more aware of their own behaviors across all settings and are also most intimately familiar with their own motivations and feelings (particularly important for assessing the characteristics such as lack of empathy and remorse). To shed further light on this issue, the present study included both self-report and parent ratings in order to provide a more comprehensive account of CU traits.

Another limitation in this research is that only two studies have explicitly tested the invariance of the factor structure of the scale across sex, with both reporting that the factor structure of the ICU was invariant for male and female adolescents (Essau et al. 2006) and children at ages 11 to 14 (Ciucci et al. 2014). Thus, the factor structure of the ICU appears to be invariant across sex but further testing among children under 11 is lacking.

In terms of the validity of the ICU, many studies have demonstrated the significant associations between CU traits and antisocial, aggressive, and delinquent behavior (e.g., Fanti et al. 2009; Kimonis et al. 2008; Roose et al. 2010). However, most studies have mainly relied on self-report for the validation measures of the ICU, and more importantly, less is known about the relationships between CU traits and internalizing behavior problems. Although some studies have shown that CU traits are unrelated or negatively related to internalizing symptoms including anxiety and depression (Frick et al. 1999; Pardini et al. 2012), others have linked CU traits to higher levels of internalizing behavior problems (Berg et al. 2013; Essau et al. 2006). These inconsistent findings highlight the need for further examination of the associations between CU traits and internalizing symptoms.

The main aim of the study was to investigate the psychometric properties of the self-report and parent versions of the ICU in boys and girls drawn from the community. Specifically, we expect for each version of the ICU to (1) compare the bifactor, three-factor, and two-factor models reported in previous studies by conducting a series of confirmatory factor analyses (CFA), (2) investigate the item-total correlation to identify poorly discriminating items, (3) directly test whether the factor structure and levels of scales were invariant across sex, and (4) investigate convergent and discriminant validity of the refined measure’s factor and total scores by relating them to the number of Conduct Disorder (CD) and Oppositional Defiant Disorder (ODD) symptoms and internalizing symptoms. Based on prior research (Roose et al. 2010), we predicted that the two-factor structure (Callousness and Uncaring) would be replicated in this sample for both self-report and parent versions, and that the structure of ICU would be invariant across sex. In addition, we predicted that the symptoms of ODD and CD would be associated with CU traits (Feilhauer et al. 2012). Finally, we also explored the relationships between ICU subscale scores and internalizing symptoms.

Method

Participants

The sample was drawn from participants in the Healthy Childhood Study, an ongoing longitudinal study aiming to examine the interplay of biological and social factors on the development of antisocial behavior in middle childhood. It consisted of 8- to 10-year old boys and girls (Mean age = 9.06, $SD = 0.60$) living in Brooklyn, New York. Within the study area, fliers soliciting enrollment were placed in public areas and targeted mailings were also sent to parents of 8- to 10-year-old children living in the geographic catchment area. Children with a diagnosed psychiatric disorder, mental retardation, or a pervasive developmental disorder were excluded. The sample included 164 male (48.2 %), 11 % Hispanic ($n = 38$), 21 % Caucasian ($n = 71$), 52 % African-American ($n = 176$), 2 % Asian ($n = 8$), and the remaining 14 % of mixed/other ($n = 48$). Compared to the Kings County or New York population, our sample consisted of slightly more African-Americans, and had lower median family income (\$43,200 compared to \$45,215). Participants and their main caregivers were invited to the university for a 2-h laboratory assessment including behavioral interviews, neurocognitive testing, psychophysiological recording, and social risk factor assessment. Incentives were provided to the participating families at the end of the assessment. All procedures were approved by the university Institutional Review Board, and both parental consent and child assent were obtained.

Measures

Callous-Unemotional Traits Both parent and child filled out the 24-item parent- and self-report version of the ICU (Frick 2004). A trained research assistant accompanied the child during the whole testing session and answered any question the child might have had. Parent filled out the ICU in a different room. The items are rated on a 4-point scale from 0 (not true at all) to 3 (very true) and are summed up for scorings for each subscale. There are 11 items for the callousness subscale (e.g., Does not seem to know “right” from “wrong”), 8 items for the uncaring subscale (e.g., Feeling bad or guilty when he/she has done something wrong), and 5 items for the unemotional subscale (e.g., Expresses his/her feeling openly). In this sample with original ICU items, means were 18.89 ($SD = 7.03$, $\alpha = 0.64$) and 17.71 ($SD = 8.81$, $\alpha = 0.85$) for the self- and parent-reported ICU total scores, respectively. For the subscales, means were 5.87 ($SD = 3.94$, $\alpha = 0.56$), 5.68 ($SD = 3.92$, $\alpha = 0.70$), and 7.34 ($SD = 2.80$, $\alpha = 0.39$) for the self-report version and 4.29 ($SD = 3.46$, $\alpha = 0.66$), 9.37 ($SD = 4.97$, $\alpha = 0.83$) and 4.04 ($SD = 2.54$, $\alpha = 0.64$) for the parent version of the Callousness, Uncaring, and Unemotional subscales, respectively.

ODD/CD Symptoms Children’s ODD/CD symptoms were assessed using the Diagnostic Interview Schedule for Children (DISC-IV; Shaffer et al. 2000). The reliability and validity of the DISC-IV are well established (Shaffer et al. 2004). Caregivers were administered the DISC-IV to assess the lifetime number of CD and ODD symptoms. For CD, 44 % of the boys had at least one symptom, and 22 % of boys had at least 2 symptoms in their lifetime. In contrast, only 28 % of girls had one or more CD symptoms. For ODD, 84 % of the boys had at least one symptom, and 75 % of boys had at least 2 symptoms in their lifetime. In girls, 81 % had one or more ODD symptoms. None of the participants met diagnostic criteria for CD or ODD.

Internalizing Problems Information on internalizing problems was collected using the Child Behavior Checklist (CBCL; Achenbach 1991) and Youth Self-Report (YSR; Achenbach 1991). The CBCL is a caregiver rating scale composed of 112 items concerning a child’s behavior within the past 12 months. Items are rated on a 3-point scale (0 = *not true*, 1 = *sometimes or somewhat true*, 2 = *very true or often true*). Children completed the YSR, a 3-point scale containing a list of 118 specific problems in children and adolescents (Achenbach 1991). For the purposes of the present article, only the Anxious/Depressed, Withdrawal/Depressed, and Somatic Complaints subscales of the CBCL and YSR were used in our analyses.

Statistical Analyses

Firstly, the raw scores of the 24 items of ICU would be subjected to CFA with maximum likelihood estimation using AMOS 21.0. A unidimensional model with all items loading on a single ICU factor (Model 1) would be tested. Then a three-factor model in which items are loaded on three distinct but correlated latent factors, Callousness, Uncaring, and Unemotional (Model 2) would be examined. The third model (Model 3) would be a three-factor bifactor model in which there is an overarching general ICU dimension as well as three ICU factors. Five commonly used fit indices would be used to assess the goodness of fit of all measurement models: the χ^2 , the comparative fit index (CFI; Bentler 1990), the root mean square error of approximation (RMSEA; Steiger 1990), the chi-square/ df ratio (CMIN/DF; Carmines and McIver 1981), and the goodness-of-fit index (GFI; Jöreskog and Sörbom 1989; Tanaka and Huba 1985). Non-significant χ^2 indicates a good fit. An adequate model fit is indicated by CFI > 0.90, RMSEA < 0.08, CMIN/DF < 3, and GFI > 0.90, and a good model fit is indicated by CFI > 0.95, RMSEA < 0.05, CMIN/DF < 2, and GFI > 0.95 (Hu and Bentler 1995; Schermelleh-Engel et al. 2003). Internal reliability would be assessed using Cronbach’s α (< 0.60 indicates insufficient fit, 0.60 to 0.69 indicates marginal fit, 0.70 to 0.79 indicates

acceptable fit, 0.80 to 0.89 indicates good fit, and >0.90 indicates excellent fit; Barker et al. 1994). Secondly, items with item-total correlation below 0.30 would be considered to provide poor discrimination and be eliminated (Nunnally and Bernstein 1994). Prior models would again be subjected to CFAs with remaining items. Thirdly, multiple groups analyses would be conducted to evaluate if the factor structure of the revised ICU are equivalent across sex. The factor loadings, factor variances, factor covariance, and error variances and covariance would be assessed incrementally. The change of CFI (Δ CFI) would be used to evaluate the merits of the competing model, and Δ CFI >0.01 indicates that the unconstrained model should be accepted, meaning the existence of sex differences (Cheung and Rensvold 2002). Then independent samples *t*-tests would be conducted to determine the effects of sex on factor scores and total score. Finally, validity of the ICU total score and its subscales would be assessed by examining their correlations with the number of ODD/CD symptoms and internalizing problems. Partial correlation would also be conducted to examine these associations after controlling for sex. Separate analyses would be conducted for the self-report and parent versions of ICU.

Results

Self-Report ICU

A unidimensional model (Model 1) with all items loading on a single ICU factor showed an unsatisfactory fit. The three-factor model (Model 2) fit significantly better ($\Delta df = 3$, $\Delta\chi^2 = 228.938$, $p < 0.001$), although several fit indices were unacceptable (see Table 1 for model indices). Further, the three-factor bifactor model (Model 3) showed a better fit than the three-factor model ($\Delta df = 21$, $\Delta\chi^2 = 143.651$, $p < 0.001$), yet continued to provide inadequate fit to the data. In addition, the internal consistencies for the original three subscales were inadequate. The item–total correlations across the items were examined to identify poorly discriminating items (item-total correlation $r < 0.30$). Item 2 ($r = 0.07$), item 4 ($r = 0.08$), item 8 ($r = 0.01$), and item 10 ($r = 0.11$) for the Callousness factor; item 5 ($r = 0.24$) and item 13 ($r = 0.22$) for the Uncaring factor; and item 6 ($r = 0.15$) and item 22 ($r = 0.11$) for the Unemotional factors were removed in a stepwise fashion.¹ With these items eliminated, CFAs for this

¹ The four Callousness items were: item 2 (*What I think is “right” and “wrong” is different from what other people think*); item 4 (*I do not care who I hurt to get what I want*); item 8 (*I am concerned about the feelings of others*); item 10 (*I do not let my feelings control me*). The two Uncaring items were: item 5 (*I feel bad or guilty when I do something wrong*); item 13 (*I easily admit to being wrong*). The two Unemotional items were: item 6 (*I do not show my emotions to others*); item 22 (*I hide my feelings from others*).

modified three-factor model (Model 4) and the three-factor bifactor model (Model 5) were performed. The modified three-factor model (Model 4) was significantly better than the original three-factor model (Model 2, $\Delta df = 148$, $\Delta\chi^2 = 365.1$, $p < 0.001$), and fit adequately to the data. Similarly, the modified three-factor bifactor model (Model 5) also showed a better fit than the prior bifactor model ($\Delta df = 128$, $\Delta\chi^2 = 209.284$, $p < 0.001$), and fit satisfactorily to the data. Given that Model 5 did not fit better than Model 4 ($\Delta df = 1$, $\Delta\chi^2 = -12.165$, $p = 1.000$), and that compared to Model 4, Model 5 had poorer values of all fit indices, lower factor loadings items (below 0.35), and an opposite sign to the content of the item, the revised three-factor model without a hierarchical factor (Model 4) was considered superior and retained for further analyses (Ezpeleta et al. 2013).

For each factor of the revised three-factor model (Model 4), the internal consistency was insufficient for the Unemotional factor ($\alpha = 0.53$) (Callousness factor, $\alpha = 0.65$; Uncaring factor, $\alpha = 0.77$). Following Houghton et al. (2013a, b), we clustered the Callousness and Unemotional items together, so a model where Unemotional items were loaded on the Callousness factor (Model 6) was tested. However, this model did not provide a sufficient fit, and factor loadings for the three Unemotional items on the Callousness factor were low (between -0.11 and 0.01). As a result, the entire Unemotional factor was deleted and a two-factor model (Model 7) showed an adequate fit. To further improve the fit, modification indices for Model 7 were reviewed to evaluate if certain pairs of error terms with similar item content could be correlated. As a result, two Uncaring items were correlated in Model 8.² This model fit data significantly better than Model 7, $\Delta\chi^2(1) = 23.881$, $p < 0.001$, and this final model (Model 8) fit data well (see Table 1 for model indices). Table 2 shows factor loadings and reliabilities for the final model.

Effect of Sex A series of model comparisons indicated that the factor structures were invariant across sex (all Δ CFIs <0.008 ; not shown for brevity, but available on request). Means for the revised Callousness factor, Uncaring factor, and total ICU were 2.37 ($SD = 3.11$), 3.40 ($SD = 3.50$), and 5.77 ($SD = 4.70$) for boys, and 2.66 ($SD = 3.31$), 2.414 ($SD = 2.71$), and 5.08 ($SD = 4.84$) for girls, respectively. Independent samples *t*-tests showed that boys and girls did not differ significantly on the Callousness factor ($t(321) = -0.81$, $p = 0.419$, $d = -0.09$) and the total ICU ($t(321) = 1.31$, $p = 0.190$, $d = 0.15$). However, boys showed significantly higher Uncaring factor score than girls, $t(321) = 2.84$, $p < 0.05$, $d = 0.32$.

² The two Uncaring items were: item 16 (*I apologize (“say I am sorry”) to persons I hurt*); item 17 (*I try not to hurt others’ feelings*).

Table 1 Fit indices comparing the measurement models for self-report (Top) and parent-report (Bottom) ICU

	χ^2	<i>df</i>	RMSEA (90 % CI)	CFI	CMIN/DF	GFI
Self-Report ICU						
Model 1 (1 factor)	757.920*	252	0.068 (0.072–0.085)	0.520	3.008	0.810
Model 2 (3-factor)	528.982*	249	0.058 (0.052–0.065)	0.734	2.124	0.878
Model 3 (3-factor bifactor)	385.331*	228	0.046 (0.038–0.054)	0.851	1.690	0.910
Model 4 (modified 3-factor)	163.882*	101	0.044 (0.031–0.055)	0.921	1.623	0.940
Model 5 (modified 3-factor bifactor)	176.047*	100	0.048 (0.036–0.060)	0.905	1.760	0.938
Model 6 (modified 2-factor with “Unemotional” items clustered)	251.085*	136	0.066 (0.056–0.077)	0.815	2.438	0.907
Model 7 (modified 2-factor without “Unemotional” items)	134.713*	64	0.058 (0.044–0.072)	0.902	2.105	0.939
Model 8 (modified 2-factor without “Unemotional” items, and with 1 pair of correlated errors)	110.832*	63	0.048 (0.033–0.063)	0.934	1.759	0.950
Parent-Report ICU						
Model 1 (1 factor)	1083.890*	252	0.100 (0.094–0.106)	0.617	4.301	0.732
Model 2 (3-factor)	917.019*	249	0.096 (0.084–0.096)	0.692	3.683	0.795
Model 3 (3-factor bifactor)	598.448*	228	0.070 (0.063–0.077)	0.829	2.625	0.868
Model 4 (modified 3-factor)	462.961*	149	0.080 (0.071–0.088)	0.818	3.107	0.864
Model 5 (modified 3-factor bifactor)	519.947*	147	0.087 (0.07–0.096)	0.784	3.537	0.862
Model 6 (modified 3-factor with 8 pairs of correlated errors)	303.940*	141	0.059 (0.050–0.070)	0.906	2.156	0.918

* $p < 0.001$

Parent-Report ICU

The unidimensional model (Model 1), the three-factor model (Model 2), and the three-factor bifactor model (Model 3) did not provide insufficient fit to the data (see Table 1 for model indices). Item 2 ($r = 0.26$), item 8 ($r = 0.16$), item 10 ($r = 0.08$), and item 21 ($r = 0.25$) for the Callousness factor and item 6 ($r = 0.26$) for the Unemotional factor had low item-total correlation and factor loadings, and thus were removed in a step-wise fashion. With deletions of these items, the modified three-factor model (Model 4) was superior to the original three-factor ($\Delta df = 100$, $\Delta \chi^2 = 454.058$, $p < 0.001$) and the original three-factor bifactor model ($\Delta df = 79$, $\Delta \chi^2 = 135.487$, $p < 0.001$) but did not reach an acceptable fit. The modified three-factor bifactor model (Model 5) did not provide better fit than the original bifactor model ($\Delta df = 81$, $\Delta \chi^2 = 78.501$, $p = 0.558$) and fit inadequately to the data. In comparing the two modified models (Model 4 and Model 5), analyses suggested that some of the parameters for the Model 5 were unsatisfactory ($\Delta \chi^2$ was statistically non-significant, factor loadings were below 0.35 or with an opposite sign to the

content of the item), thus the modified three-factor model without a hierarchical factor (Model 4) was considered superior. Internal consistency was sufficient for all factors in Model 4 (Callousness factor, $\alpha = 0.71$; Uncaring factor, $\alpha = 0.83$; Unemotional factor, $\alpha = 0.63$). To further improve the fit, modification indices for Model 4 were reviewed, and 8 pairs of error variables were correlated.³ This model with correlated error terms (Model 6) achieved an acceptable fit, and fit significantly better than Model 4, $\Delta \chi^2 (8) = 159.021$, $p < 0.001$. See Table 2 for factor loadings.

³ The correlated items were: item 20 (*Does not like to put the time into doing things well*) was correlated with item 15 (*Always tries his/her best*) and item 23 (*Works hard on everything*); item 4 (*Does not care who he/she hurts to get what he/she wants*) was correlated with item 16 (*Apologizes (“says he/she is sorry”) to persons he/she has hurt*); item 23 (*Works hard on everything*) was correlated with item 15 (*Always tries his/her best*), item 3 (*Is concerned about schoolwork*) and item 5 (*Feels bad or guilty when he/she has done something wrong*); item 15 (*Always tries his/her best*) was correlated with item 5 (*Feels bad or guilty when he/she has done something wrong*); and item 16 (*Apologizes (“says he/she is sorry”) to persons he/she has hurt*) was correlated with item 17 (*Tries not to hurt others’ feelings*).

Table 2 Factor loadings for final self-report (top) and parent-report (bottom) ICU models

		Items		Factor loadings
Self-Report ICU	(13 items; $\alpha = 0.70$)			
	Callousness factor (7 items; $\alpha = 0.65$)	Item 7	I do not care about being on time.	0.49
		Item 9	I do not care if I get into trouble.	0.36
		Item 11	I do not care about doing things well.	0.58
		Item 12	I seem very cold and uncaring to others.	0.36
		Item 18	I do not feel remorseful when I do something wrong.	0.46
		Item 20	I do not like to put the time into doing things well.	0.53
		Item 21	The feelings of others are unimportant to me.	0.45
	Uncaring factor (6 items; $\alpha = 0.77$)	Item 3	I care about how well I do at school or work.*	0.44
		Item 15	I always try my best.*	0.80
		Item 16	I apologize ('say I am sorry') to persons I hurt.*	0.51
		Item 17	I try not to hurt others' feelings.*	0.44
		Item 23	I work hard on everything I do.*	0.76
Item 24		I do things to make others feel good.*	0.60	
Parent-Report ICU	(19 items; $\alpha = 0.85$)			
	Callousness factor (7 items; $\alpha = 0.71$)	Item 4	Does not care who he/she hurts to get what he/she wants.	0.37
		Item 7	Does not care about being on time.	0.49
		Item 9	Does not care if he/she is in trouble.	0.55
		Item 11	Does not care about doing things well.	0.75
		Item 12	Seems very cold and uncaring.	0.37
		Item 18	Shows no remorse when he/she has done something wrong.	0.40
		Item 20	Does not like to put the time into doing things well.	0.64
	Uncaring factor (8 items; $\alpha = 0.83$)	Item 3	Is concerned about schoolwork.*	0.37
		Item 5	Feels bad or guilty when he/she has done something wrong.*	0.71
		Item 13	Easily admits to being wrong.*	0.55
		Item 15	Always tries his/her best.*	0.68
		Item 16	Apologizes ("says he/she is sorry") to persons he/she has hurt.*	0.67
		Item 17	Tries not to hurt others' feelings.*	0.66
		Item 23	Works hard on everything.*	0.62
		Item 24	Does things to make others feel good.*	0.64
	Unemotional factor (4 items; $\alpha = 0.63$)	Item 1	Expresses his/her feelings openly.*	0.59
		Item 14	It is easy to tell how he/she is feeling.*	0.76
		Item 19	Is very expressive and emotional.*	0.47
Item 22		Hides his/her feelings from others.	0.36	

* Reverse scored items. All factor loadings are significant at $p < 0.01$

Effect of Sex The factor structures are invariant across sex (all Δ CFIs < 0.005), although there was sex difference in error variances (Δ CFI = 0.032). Specifically, sex differences were found for the error variance of Callousness item 12 (*Seems very cold and uncaring*), suggesting that the reliability of this item was lower in boys. Means for the revised Callousness, Uncaring, Unemotional factor, and total ICU were 2.05 ($SD = 2.36$), 9.77 ($SD = 4.81$), 4.04 ($SD = 2.22$), and 15.85 ($SD = 7.57$) for boys, and 1.82 ($SD = 2.52$), 8.92 ($SD = 5.01$), 3.64 ($SD = 2.38$), and 14.38 ($SD = 7.81$) for girls, respectively. Boys and girls did not differ on total ICU scores or any of the subscale scores, $t < 1.75$, $p > 0.08$.

Associations with Behavioral Problems

The associations between the revised ICU scales scores (self- and parent-report), internalizing behaviors, and ODD/CD symptoms are presented in Table 3. In general, higher levels of CU traits were related to more internalizing and ODD/CD symptoms, and these correlations largely remain unchanged after sex was controlled for. The cross-informant correlations are 0.23, 0.23, and 0.15 for the total ICU, Uncaring, and Callousness subscale scores, respectively (all $p < 0.01$). For the parent version, the Uncaring scale correlated with the Callousness scale at 0.42 ($p < 0.001$) and Unemotional scale

Table 3 Correlations between the revised ICU scales and behavioral problems

	Mean	SD	Anxious/Depressed		Withdrawal/Depressed		Somatic Complaints		ODD Symptoms	CD Symptoms
			CBCL	YSR	CBCL	YSR	CBCL	YSR		
α			0.79	0.73	0.70	0.56	0.80	0.75		
Mean	3.03		5.59	5.59	1.57	4.08	1.83	5.95	3.84	0.77
SD	3.20		3.95	3.95	2.04	2.71	2.55	3.80	3.03	1.39
Parent-Report										
Total ICU	15.10	7.72	0.30** (0.30**)	0.07 (0.07)	0.42** (0.41**)	0.17** (0.18**)	0.21** (0.22**)	0.04 (0.04)	0.38** (0.37**)	0.36** (0.35**)
Callousness	1.93	2.44	0.36** (0.36**)	0.11 (0.10)	0.37** (0.37**)	0.13* (0.13*)	0.25** (0.26**)	0.08 (0.08)	0.29** (0.29**)	0.34** (0.34**)
Uncaring	9.34	4.92	0.22** (0.22**)	0.07 (0.07)	0.30** (0.30**)	0.17** (0.17**)	0.14** (0.15**)	0.05 (0.05)	0.38** (0.38**)	0.30** (0.29**)
Unemotional	3.83	2.31	0.15** (0.16**)	-0.02 (-0.01)	0.36** (0.36**)	0.07 (0.08)	0.14* (0.14*)	-0.06 (-0.05)	0.14** (0.13*)	0.21** (0.20**)
Self-Report										
Total ICU	5.42	4.78	0.16** (0.16**)	0.24** (0.24**)	0.17** (0.16**)	0.27** (0.28**)	0.10 (0.10)	0.16** (0.17**)	0.12* (0.11*)	0.08 (0.08)
Callousness	2.52	3.21	0.11 (0.11*)	0.22** (0.22**)	0.02 (0.02)	0.28** (0.27**)	0.03 (0.03)	0.23** (0.23**)	0.05 (0.05)	-0.04 (-0.04)
Uncaring	2.89	3.16	0.13* (0.13*)	0.14* (0.14*)	0.23** (0.23**)	0.13* (0.15**)	0.12* (0.13*)	0.01 (0.03)	0.14* (0.12*)	0.16** (0.14*)

CBCL = Child Behavior Checklist. YSR = Youth Self-Report. Partial correlations controlling for sex are given in parentheses

** $p < 0.01$, * $p < 0.05$

at 0.49 ($p < 0.001$) and the Callousness correlated with the Unemotional scale at 0.25 ($p < 0.001$). For the self-report version, the Callousness correlated with the Uncaring scale at 0.13 ($p < 0.05$).

Discussion

In this study, we examined the psychometric properties of the self-report and parent versions of the ICU in a community sample of boys and girls. The main findings are (1) for the self-report ICU, a two-factor structure comprising one Callousness factor (7 items) and one Uncaring factor (6 items) fit adequately to the data and was found to be superior to the bifactor or three-factor model; the factor structure was invariant across sex, although boys scored higher on the Uncaring factor than girls; (2) for the parent-report ICU, a three-factor structure (Callousness, 7 items; Uncaring, 8 items; and Unemotional, 4 items) was supported; boys and girls did not differ on factor structure or mean levels of the three factors; (3) the revised ICU total score and factor scores were positively associated with ODD/CD symptoms and internalizing behavior problems. Findings suggest that future refinement of the ICU should take various raters into account.

In general, our findings on the self-report version bear some similarities to Houghton et al. (2013a)'s results. First, all the Unemotional items were excluded from the final model. This factor had poor internal reliability, and none of the items making up the Unemotional factor loaded onto the other two factors. Houghton and colleagues argued that children younger than 12 years old might not be able to “feel” the (affective) emotions of others, and that “they may not have had the experience to be able to attribute these emotions to themselves or others” (Houghton et al. 2013a). Thus self-report Unemotional items may not be useful in differentiating children with high or low CU traits. In fact, it has been argued that the Unemotional factor may be tapping into a distinct construct from CU and further examination of this factor is needed in future studies to determine the need to include these items as part of future (self-report) version of the ICU (Kimonis et al. 2013).

Meanwhile, our findings are different from Houghton et al. (2013a)'s in a few important ways. First, we found boys to have higher scores than girls on Uncaring (i.e., a lack of caring about one's performance in tasks and for others' feelings), whereas Houghton et al. failed to find any sex difference in the mean levels of factor scores. Second, one pair, instead of 8 pairs of error items were correlated in our final model. These different findings may be caused by several factors, including the different characteristics of the samples involved in the two studies. Houghton et al.'s study included Australian children from grades 3 to 7, with a wide age range from 7 to 13 years, and it was purely school-based. In contrast, our sample was

community based; it composed of American children aged 8 to 10 years, and the majority of them were African-Americans and from low SES families. These inconsistencies may call for further study on the potential effects of ethnicity, social background, and developmental factors on the factorial structure of self-report ICU.

For the parent-report version, we found that the three-factor model fit the data adequately well, after four Callousness items⁴ and one Unemotional item were deleted due to low item-scale correlation, and 8 pairs of errors were correlated. The internal reliability was acceptable for two of the three subscales, with coefficient alphas being 0.83 for the Uncaring factor and 0.71 for the Callousness factor. The internal reliability was marginal (0.63) for the Unemotional factor. It was consistent with prior studies that also showed lower levels of reliability for the Unemotional factor, partly due to fewer items, than for the other two factors (e.g., Essau et al. 2006). Also in line with prior literature (Frick et al. 2003), the internal reliability estimates for the self-report scales were lower than for the parent version (Table 2). Note that the cross-informant correlations for the ICU total score and subscales ranged from 0.15 to 0.23, slightly lower than the correlations found in past studies of CU traits ($r = 0.29\text{--}0.57$; Frick et al. 2003; Roose et al. 2010). Taken together, ratings from two different informants seem to provide overlapping but different perspectives of CU traits. Although youth self-reports may provide an introspective and private view of their own behaviors, our findings seem to suggest that parent reports on these traits may be more accurate and reliable, at least in children at this age range.

As expected, total ICU score and factor scores were all significantly associated with the number of ODD/CD symptoms. Positive relationships were also found between CU scores and internalizing behaviors, indicating that CU traits may relate to parents or children themselves endorsing them as being socially withdrawn, isolated, or low in mood. These findings are in line with prior research that links the occurrence of CU traits with ODD and CD symptoms (Barry et al. 2000; Feilhauer et al. 2012) and internalizing problems (Hawes et al. 2014; Waller et al. 2015), and advance this past work in showing that these associations extend to both self- and parent-report ICU in children under 12 years. The high correlations between ICU scores and counts of ODD/CD symptoms provide some support to the ICU's convergent validity. However, its discriminant validity seems to be questionable because psychopathy and anxiety/depression are believed to be unique concepts and valid measures of each should not correlate too highly (Bagozzi et al. 1991). One possible

⁴ The four Callousness items were: item 2 (*Does not seem to know "right" from "wrong"*); item 8 (*Is concerned about the feelings of others*); item 10 (*Does not let feelings control him/her*); item 21 (*The feelings of others are unimportant to him/her*). The one Unemotional item was item 6 (*Does not show emotions*).

explanation for these positive associations is that we did not examine the entire concept of psychopathy. Although CU traits are considered the central features of psychopathy, it is important to examine a broader range of psychopathic traits, including the grandiose-manipulative and daring-impulsive dimensions, to get a more comprehensive understanding of psychopathic traits in children and adolescents (Salekin 2015). Finally, it is worth noting that significant correlations emerging with self-reported ICU scores are predominately YSR scores, and those emerging with parent-reported ICU are predominately CBCL and DISC-IV scores (Table 3), indicating the influence of shared method variance.

The current findings should be considered in light of several limitations. First, the present study was community-based, so the findings may not be generalized to children presenting with elevated CU traits, including those in clinical, institutional or referral-based settings. In addition, African-American participants and low-income families were slightly overrepresented in our sample. Second, the reliabilities for some subscales were relatively low, although even lower reliabilities (around 0.40) have been reported in some studies (see review by Waller et al. 2013). These low reliabilities highlight the need for continued investigation of the construct of CU traits among youth samples. Third, the current study was cross-sectional. Additional time points of measurement would have allowed the investigation of trajectories of change over time. Investigating how CU traits and its subfactors change across ages, and linking these changes to behavioral changes during adolescence and adulthood should be explored in future.

In summary, the current study compared two-factor, bifactor, and three-factor structure for the self-report and parent versions of ICU in a community sample of children. Findings indicate that the factorial structure is same for boys and girls, but different for various informants (i.e., children themselves vs. parents). In addition, the ICU total and factor scores were positively associated with ODD/CD symptoms and internalizing behavioral problems, demonstrating ICU's satisfactory convergent but questionable discriminant validity. Together with previous research, current findings demonstrate better reliability for the parent than for the self-report version of the ICU, and underscore the need to further refine this instrument for the two informants separately.

Acknowledgments Research reported in this publication was supported by the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health to the first author under Award Number SC2HD076044. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. We would like to thank the Psychophysiology Lab staff for their assistance in collecting data and the families for participation.

Compliance with Ethical Standards

Conflicts of Interest Yu Gao and Wei Zhang declare that they have no conflict of interest.

Experiment Participants This study was approved by the university Internal Review Board (IRB) and all study procedures, including informed consent, followed the protocol approved by the IRB.

References

- Achenbach, T. M. (1991). *Integrative guide for the 1991 CBCL/4–18, YSR, and TRF profiles*. Burlington: University of Vermont, Department of Psychiatry.
- Bagozzi, R. P., Yi, Y., & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, 36(3), 421–458.
- Barker, C., Pistrang, N., & Elliott, R. (1994). *Research methods in clinical and counseling psychology*. Chichester: John Wiley.
- Barry, C. T., Frick, P. J., DeShazo, T. M., McCoy, M., Ellis, M., & Loney, B. R. (2000). The importance of callous–unemotional traits for extending the concept of psychopathy to children. *Journal of Abnormal Psychology*, 109(2), 335–340.
- Benesch, C., Görtz-Dorten, A., Breuer, D., & Döpfner, 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.
- 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: a construct-validated analysis in an at-risk sample. *Assessment*, 20(5), 532–544.
- Byrd, A. L., Kahn, R. E., & Pardini, D. A. (2013). A validation of the inventory of callous-unemotional traits in a community sample of young adult males. *Journal of Psychopathology and Behavioral Assessment*, 35(1), 20–34.
- Carmines, E. G., & McIver, J. P. (1981). *Analyzing models with unobserved variables: analysis of covariance structures*. In G. Bohnstedt & E. Borgatta (Eds.), *Social measurement: Current issues*, (pp. 65–115). Beverly Hills, CA: Sage.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255.
- 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.
- Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Callous-unemotional traits in a community sample of adolescents. *Assessment*, 13(4), 454–469.
- Ezpeleta, L., Osa, N. D. L., Granero, R., Penelo, E., & Domènech, J. M. (2013). Inventory of callous-unemotional traits in a community sample of preschoolers. *Journal of Clinical Child & Adolescent Psychology*, 42(1), 91–105.
- 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.
- 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.
- Frick, P. J. (2004). *The inventory of callous-unemotional traits. Unpublished rating scale*. New Orleans: University of New Orleans.
- Frick, P. J., & Dickens, C. (2006). Current perspectives on conduct disorder. *Current Psychiatry Reports*, 8(1), 59–72.
- 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.
- Frick, P. J., Lilienfeld, S. O., Ellis, M., Loney, B., & Silverthorn, P. (1999). The association between anxiety and psychopathy dimensions in children. *Journal of Abnormal Child Psychology*, 27(5), 383–392.
- 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.
- Hawes, S. W., Byrd, A. L., Henderson, C. E., Gazda, R. L., Burke, J. D., Loeber, R., & Pardini, D. A. (2014). Refining the parent-reported inventory of callous–unemotional traits in boys with conduct problems. *Psychological Assessment*, 26(1), 256–266.
- Houghton, S., Hunter, S. C., & Crow, J. (2013a). 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.
- Houghton, S., Hunter, S. C., Khan, U., & Tan, C. (2013b). Interpersonal and affective dimensions of psychopathic traits in adolescents: development and validation of a self-report instrument. *Child Psychiatry & Human Development*, 44(1), 51–69.
- Hu, L. T., & Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling. Concepts, issues, and applications* (pp. 76–99). London: Sage.
- Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7 user's reference guide*. Chicago: Scientific Software International.
- Kimonis, E. R., Branch, J., Hagman, B., Graham, N., & Miller, C. (2013). The psychometric properties of the inventory of callous–unemotional traits in an undergraduate sample. *Psychological Assessment*, 25(1), 84–93.
- 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.
- Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1991). Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology*, 19(1), 75–95.
- Munoz, L. C., & Frick, P. J. (2007). The reliability, stability, and predictive utility of the self-report version of the antisocial process screening device. *Scandinavian Journal of Psychology*, 48(4), 299–312.
- Nunnally, J. C., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.,). New York, NY: McGraw-Hill.
- Pardini, D., Stepp, S., Hipwell, A., Stouthamer-Loeber, M., & Loeber, R. (2012). The clinical utility of the proposed DSM-5 callous-unemotional subtype of conduct disorder in young girls. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(1), 62–73.
- 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.
- Salekin, R. T. (2015). *Psychopathy in childhood: toward better informing the DSM-5 and ICD-11 conduct disorder specifiers*. Personality Disorders: Theory, Research, and Treatment. Advance online publication. doi:10.1037/per0000150.

- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Shaffer, D., Fisher, P., & Lucas, C. (2004). The diagnostic interview schedule for children (DISC). In M. Hersen (Ed.), *Comprehensive handbook of psychological assessment* (pp. 256–270). Hoboken, NJ: John Wiley & Sons.
- Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & Schwab-Stone, M. E. (2000). NIMH diagnostic interview schedule for children version IV (NIMH DISC-IV): description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child & Adolescent Psychiatry*, 39(1), 28–38.
- Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*, 25(2), 173–180.
- Tanaka, J. S., & Huba, G. J. (1985). A fit index for covariance structure models under arbitrary GLS estimation. *British Journal of Mathematical and Statistical Psychology*, 38(2), 197–201.
- Waller, R., Gardner, F., & Hyde, L. W. (2013). What are the associations between parenting, callous–unemotional traits, and antisocial behavior in youth? A systematic review of evidence. *Clinical Psychology Review*, 33(4), 593–608.
- Waller, R., Hyde, L. W., Grabbell, A. S., Alves, M. L., & Olson, S. L. (2015). Differential associations of early callous-unemotional, oppositional, and ADHD behaviors: multiple domains within early-starting conduct problems? *Journal of Child Psychology and Psychiatry*, 56(6), 657–666.
- Willoughby, M. T., Mills-Koonce, W. R., Waschbusch, D. A., & Gottfredson, N. C. (2015). An examination of the parent report version of the inventory of callous-unemotional traits in a community sample of first-grade children. *Assessment*, 22(1), 76–85.