The Reliability and Validity of the Inventory of Callous Unemotional Traits: A Meta-Analytic Review

Elise M. Cardinale and Abigail A. Marsh

Abstract

In the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders, a Limited Prosocial Emotions specifier was added to the conduct disorder diagnostic criteria to designate a subgroup of children who exhibit callous unemotional (CU) traits. The Inventory of Callous Unemotional Traits (ICU) is the only dedicated measure of CU traits and was influential in the development of the Limited Prosocial Emotions specifier. Despite its role in the research and diagnosis of CU traits, some questions have persisted regarding the internal consistency and validity of the ICU and its three subscales: callous, uncaring, and unemotional. Results of a meta-analysis revealed acceptable internal consistency and external validity for total ICU, callous, and uncaring scores, but not unemotional scores. These results support the utility of the total ICU, callous, and uncaring scales, but indicate weaknesses in the scale or construct of unemotionality as it relates to interpersonal callousness, uncaring, and antisociality.

Keywords

Inventory of Callous Unemotional Traits, Limited Prosocial Emotions, psychopathy, aggression, meta-analysis

Extensive empirical evidence implicates callous unemotional (CU) traits in the development of severe and persistent antisocial behaviors (Frick, Stickle, Dandreaux, Farrell, & Kimonis, 2005; Frick & White, 2008) and in poor responses to traditional treatment (D. J. Hawes & Dadds, 2005; Wilkinson, Waller, & Viding, 2016). The Inventory of Callous Unemotional Traits (ICU; Kimonis et al., 2008), the only dedicated measure of CU traits, was influential in the development of the recent Limited Prosocial Emotions specifier to identify CU traits in youths with conduct disorder (American Psychiatric Association, 2013; Frick & Moffitt, 2010), and has been used as a research tool across a wide variety of samples and contexts (Feilhauer, Cima, & Arntz, 2012; Horan, Brown, Jones, & Aber, 2015; Piher, Etter, Schmid, & Kimonis, 2015). The ICU was originally constructed as a single scale, but is now commonly viewed as having a three-factor bifactor structure, consisting of one overarching CU dimension and three subfactors: the callous, uncaring, and unemotional subscales. These subscales are widely used as research tools in studies investigating the development of psychopathic traits and antisocial behaviors. However, no quantitative meta-analysis has yet confirmed their reliability or their associations with relevant outcomes.

The ICU was developed in response to concerns about the poor internal reliability of CU subscale scores of youth-adapted measures of psychopathy, for example, the Hare Psychopathy Checklist: Youth Version (Forth, Kosson, & Hare, 2014) and the Antisocial Process Screening Device (Frick & Hare, 2001; Muñoz & Frick, 2007). The ICU was created by expanding the CU subscale of the Antisocial Process Screening Device into a dedicated measure of CU traits, consisting of 24 items, half of which are positively worded and half of which are negatively worded. Since their development, self-report and observer-report versions of the ICU have been used to assess CU traits in a wide range of samples, including both males and females (Colins, Andershed, Hawes, Bijnste, & Pardini, 2015; Piher et al., 2015), and detained children, clinically evaluated children, and children recruited from the community (Kimonis et al., 2008; Pechorro, Ray, Barroso, Maroco, & Gonçalves, 2016; Piher et al., 2015). Systematic reviews of the literature support total ICU scores as a valid continuous measure of CU traits, as total scores reliably demonstrate acceptable internal consistency and strong associations with variables associated with the development of psychopathy, such as antisocial...
behavior (Frick & Ray, 2015; Frick, Ray, Thornton, & Kahn, 2014).

Although the ICU was originally constructed as a unidimensional measure of CU traits, subsequent exploratory factor analyses revealed a three-factor bifactor structure, with one overarching CU factor within which are nested three subfactors: callousness, or reduced empathic responding (e.g., “The feelings of others are unimportant to me”), uncaring, or lack of concern about performance and relationships (e.g., “I work hard on everything I do,” reverse scored), and unemotionality, or impoverished emotional experience and expression (e.g., “I hide my feelings from others;” Byrd, Kahn, & Pardini, 2013; Ciucci, Baroncelli, Franchi, Golmaryami, & Frick, 2014; Essau, Sasagawa, & Frick, 2006; Houghton, Hunter, & Crow, 2013). As a result, the ICU now includes an 11-item callous subscale, 8-item uncaring subscale, and 5-item unemotional subscale.

Although the three-factor bifactor structure has emerged as the best-fit model across a variety of sample populations, concerns have been raised regarding the ICU’s subscales and the constructs they represent. Perhaps the most significant concern is that the apparent factor structure of the ICU is a spurious effect that can be better explained by method variance. Specifically, common wording of the items that comprise the subscales may be the cause of covariation among these items, rather than the items capturing unique underlying constructs (Frick & Ray, 2015; Ray, Frick, Thornton, Steinberg, & Cauffman, 2016). Ten of the 11 items that load on the callous dimension are positively worded, meaning they correspond to increased CU traits (e.g., “I do not care who I hurt to get what I want”), whereas all 8 items that load on the uncaring dimension are negatively worded (e.g., “I always try my best”). Therefore, it is possible that the callous and uncaring factors emerged as an artifact of the wording of the items rather than because they capture distinct underlying constructs (Frick & Ray, 2015). One consideration, however, is that item wording cannot easily explain the emergence of the unemotional subscale, which contains three negatively worded and two positively worded items.

Other questions have been raised regarding the unemotional subscale, including that this subscale may be internally inconsistent with or unrelated to the overarching construct of CU traits. Supported by evidence that the unemotional subscale shows smaller correlations with total ICU scores despite relatively high interitem correlations (Ciucci et al., 2014; Houghton et al., 2013; Kimonis et al., 2008), and evidence that 78% of genetic variance related to the unemotional scale is specific to the unemotional factor and not explained by variance in callous-uncaring scores (Henry, Pingault, Boivin, Rijswijk, & Viding, 2016), some suggest that the unemotional subscale may designate a dimension that is phenotypically distinct from the overarching CU dimension. Moreover, while the callous and uncaring subscales are consistently strong predictors of both concurrent (Ansel, Barry, Gillen, & Herrington, 2015; Cima, Raine, Meesters, & Popma, 2013; Fanti, Panayiotou, Lazarou, Michael, & Georgiou, 2016) and future (Kahn, Byrd, & Pardini, 2013) aggressive and antisocial behaviors, the correspondence between the unemotional subscale and antisocial behaviors is more variable with some studies even finding that unemotionality corresponds to decreases in antisocial behaviors, including aggression (Berg et al., 2013; Essau et al., 2006).

Various moderator variables have been proposed as explaining conflicting or counterintuitive findings across studies that use the ICU. Gender may be relevant, for example, as boys tend to receive significantly higher scores than girls across all three subscales (Ciucci & Baroncelli, 2014; Essau et al., 2006; Fanti, Frick, & Georgiou, 2009). The detained status of the sample may also affect the measured correspondence between ICU scores and outcomes, given that detained samples tend to contain larger proportions of children with very high ICU scores (Colins & Andershed, 2015; Kahn et al., 2013; Kimonis et al., 2008) and more severe antisocial behaviors (Colins et al., 2010; Teplin, Abram, McClelland, Dulcan, & Mericle, 2002). Last, the use of the self-report versus other-report version of the ICU may influence the reliability of ICU scores and their associations with various outcomes (Kimonis, Kennealy, & Goulter, 2016; Picentini, Cohen, & Cohen, 1992). Together, this evidence suggests the importance of considering moderator variables like these when evaluating the factor structure, reliability, and validity of the ICU and its subscales.

The current study used quantitative meta-analytic strategies to systematically examine the internal consistency of the ICU and its three subscales, their association with the ICU’s parent construct of psychopathy, and relevant psychological and behavioral outcomes, including internalizing and externalizing symptoms and aggression. We assessed the internal consistency of the three subscales and total score of the ICU by examining both Cronbach’s alpha and intercorrelations among scores on the ICU. We then assessed the convergent validity of the ICU and its subscales by examining correlations between these measures and measures of psychopathy. Finally, we assessed the external validity of the ICU and its subscales by examining relationships with externalizing symptoms (including aggression, nonviolent delinquency, and hyperactivity), internalizing symptoms, and empathic tendencies. We predicted that total ICU scores would demonstrate strong internal consistency and external validity, but that not all subscales would exhibit comparable internal consistency and validity. Specifically, we predicted that the callous and uncaring subscales would both show strong internal consistency and external validity, but that the unemotional subscale would correspond weakly with measures of
psychopathy and externalizing. We also considered variables that may moderate the internal consistency and validity of the ICU, including gender, detention status, and ICU respondent.

**Method**

**Data Sources and Study Selection**

A literature search was conducted in May 2016 using PubMed and Google Scholar to identify relevant studies. The search term “Inventory of Callous Unemotional Traits” was used for an initial search. Next, to limit our search to those papers that used the ICU and reported statistics on all three subscales, we added a search requirement for the term “Uncaring,” which is the only subscale name that is not used when discussing the construct of CU traits more generally.

The combined literature search returned 633 results. After removing duplicates, 569 articles were evaluated using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for systematic reviews and meta-analyses (Moher, Liberati, Tetzlaff, & Altman, 2010), 489 articles were excluded based on the following criteria: (a) article was not written in English, (b) article was not empirical work, (c) study did not use the ICU, and (d) the three original subscales of the ICU were not investigated as part of the study. Following a full-text assessment of the remaining 80 papers, 5 additional papers were excluded because neither Cronbach’s alpha nor Pearson’s correlation coefficients with relevant outcome measures were reported for at least one of the original ICU subscales (Supplemental Table S1). We also sent a request for unpublished data to members of the Society for the Scientific Study of Psychopathy in September 2016 and obtained unpublished data from two groups, resulting in two new samples and additional unpublished data for a previously published sample (Henry et al., 2016). This resulted in a total sample size of 27,947 subjects from 115 samples.

**Data Synthesis and Analysis**

**Internal Consistency.** Internal consistency was assessed through examination of Cronbach’s $\alpha$ for ICU total scores and subscales scores and correlations among the three subscales of the ICU as well as between each of the ICU subscales with the total score of the ICU.

**Outcome Measures**

**Measures of psychopathy.** The relationship between ICU scores and measures of three facets of psychopathy were examined. Affective, interpersonal, and behavioral facet scores were compiled based on the three-factor model of psychopathy (Cooke & Michie, 2001). The affective facet consisted of assessments of deficient emotional experience including shallow affect, lack of empathy, and lack of remorse or guilt. The interpersonal facet consisted of assessments of an arrogant and deceitful interpersonal style including glibness/superficial charm, conning/manipulation, pathological lying, and grandiose sense of self-worth. The behavioral facet consisted of assessments of an impulsive/responsible behavioral style and presence of antisocial behaviors (Hall, Benning, & Patrick, 2004; Johansson, Andershed, Kerr, & Levander, 2002). Last, a total psychopathy score was compiled from the relevant studies. For each sample, a single effect size was calculated for the relationship between scores on the ICU and each facet of psychopathy (affective, interpersonal, behavioral, and total psychopathy) by taking the average of all reported Pearson’s correlation coefficients within each facet (see Text S1 for measures included in each facet).

**External criteria.** The relationship between ICU scores and relevant behavioral outcomes was also assessed. Internalizing outcomes included anxiety, psychological withdrawal, and depressive symptoms. Externalizing outcomes included reactive aggression, proactive aggression, nonviolent delinquency, and hyperactivity. An overarching externalizing variable was evaluated in addition to each of these distinct and nonoverlapping forms of externalizing. The reactive and proactive aggression outcomes were comprised only of explicit measures of these forms of aggression. The nonviolent delinquency outcome was composed of measures assessing nonviolent criminal behaviors (e.g., theft, vandalism, and substance use). The hyperactivity outcome consisted of measures of hyperactivity, impulsivity, and attentional difficulties. Last, we calculated empathy outcomes assessing the relationship between scores on the ICU and empathic tendencies. Again, for each of these outcomes, a single effect size for each sample was calculated for the relationship between total and subscale scores on the ICU and each of the external criteria by taking the average of all reported Pearson’s correlation coefficients within each type of outcome (see Text S1 for measures included in each outcome).

**Study Characteristics.** We examined the potential moderating effects of various study characteristics on the relationship between ICU scores and outcomes. These study characteristics included the gender composition of the sample, use of community-recruited or detained samples, and, which respondent completed the ICU. The effect of gender was examined by comparing samples composed of only males, only females, and both genders. The effect of detention status was investigated by comparing samples composed of only detained or only nondetained participants; studies only reporting statistics for combined detained and nondetained participants were not included in these analyses. All studies...
were also identified as using either self-report or other-report, the latter of which collapsed across parent- and teacher-report versions of the ICU.

**Pooled Effect Sizes.** For total ICU and subscales scores, pooled effect sizes were calculated to assess internal consistency, correlations between each subscale and total ICU scores, correlations among subscales, correlations with measures of psychopathy, and correlations with psychological and behavioral outcomes. Separate pooled effect sizes were calculated across all studies and for subgroups who varied in terms of our moderator variables. All pooled effect sizes were calculated using Hedges and Olkin’s random-effects approach. We employed a random-effects approach in order to account for anticipated heterogeneity in effect sizes across the studies included in this meta-analysis due to differences in study populations and procedures. Hedge and Olkin’s random-effects approach takes into account between-study variance, and, therefore, results in more conservative estimates with larger standard errors relative to a fixed-effects approach (Hedges & Olkin, 1985). This approach has been used in topically similar past meta-analyses (Blais, Solodukhin, & Forth, 2014; Longman, Hawes, & Kohlhoff, 2016; Ray et al., 2013; Smith & Lilienfeld, 2015; Walters, 2003).

In order to investigate whether each subscale demonstrated acceptable internal consistency and strong associations with external criteria, we evaluated the resulting pooled effect sizes through comparison of the 95% confidence intervals to the following criteria: for pooled Cronbach’s alpha, values greater or equal to 0.70 indicates acceptable internal consistency of the subscale for use as a research tool (Peterson, 1994), and for pooled Pearson’s correlation coefficients, large, medium, and small effect sizes are defined as 0.50, 0.30, and 0.10, respectively (Cohen, 1988). For moderator analyses, we compared across each study category for each of the four scores on the ICU. For example, the internal consistency of the self-report and other-report versions of the unemotional subscale were compared. Lack of overlap between 95% confidence intervals of the pooled effect size for independent samples indicated significant differences between effect sizes (Cumming & Finch, 2005).

**Publication Bias.** The Egger Test was conducted to assess potential publication biases in favor of statistically significant findings and large effect sizes in the included samples. For each relationship examined, we used the Erasmus Research Institute’s Meta-Essentials Workbooks for Meta-Analysis (van Rhee, Suurmond, & Hak, 2015) to calculate an estimate for the intercept of the regression line of the standard normal deviate regressed against the inverse standard error for each study. In the absence of publication bias, one would expect the studies to be symmetrically dispersed in a funnel plot such that the larger studies would have effect sizes that cluster around the pooled effect size, while smaller studies would show a more disperse pattern of smaller and larger effects. The resulting Eggers regression line for this symmetrical funnel plot should run through an intercept of 0. When large effects result primarily from smaller studies, the regression line is forced below the origin, resulting in a negative intercept. Publication bias is inferred when the estimated intercept for each analysis is significantly different from zero using the recommended significance threshold of $p < .10$ (Egger, Smith, Schneider, & Minder, 1997). This approach provides a nonsubjective assessment of publication bias that is equivalent to the assessment of the degree of funnel plot asymmetry.

**Results**

**Internal Consistency**

Examination of the pooled Cronbach’s alpha across all studies revealed acceptable internal consistency for the total ICU, $\alpha = 0.83$, as well as the uncaring, $\alpha = 0.80$, callous, $\alpha = 0.75$, and unemotional subscales, $\alpha = 0.71$ (Table 1). Pooled effect sizes for the correlations between the callous, $r = 0.79$, uncaring, $r = 0.82$, and unemotional subscales, $r = 0.63$, with the total ICU were all large. We also examined the intercorrelations among the three subscales and found a medium to large effect size for the correlation between the callous and uncaring subscales $r = 0.45$ and a medium to small effect size for the correlation between the uncaring and unemotional subscales, $r = 0.30$, and the callous and unemotional subscales, $r = 0.24$ (Table 1).

**Convergent Validity**

To assess convergent validity of the ICU, we examined the relationship between scores on the ICU and measures of psychopathy. First, we investigated the pooled association with total psychopathy scores across samples and found medium to large effect sizes for the associations between psychopathy and the total ICU, $r = 0.47$, callous subscale, $r = 0.43$ and uncaring subscale, $r = 0.40$. The association with the unemotional subscale, $r = 0.16$, was small. This pattern of findings remained when examining the associations with the behavioral and interpersonal facets such that again, the pooled associations with the total ICU score, callous subscale, and uncaring subscale all had medium effect sizes but the pooled association with the unemotional subscale was small. For pooled associations with the affective facet, which contains items most directly related to CU traits, the total ICU score, $r = 0.46$, callous subscale, $r = 0.40$, and uncaring subscale, $r = 0.34$, had medium effect sizes. A small pooled effect size was again found for the unemotional subscale, $r = 0.22$ (Table 1; Figure 1).
Table 1. Pooled Effect Sizes Across All Studies.

<table>
<thead>
<tr>
<th>Measure of Psychopathy</th>
<th>Callous</th>
<th>Uncaring</th>
<th>Unemotional</th>
<th>Total ICU</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$k$</td>
<td>$n$</td>
<td>ES</td>
<td>95% CI</td>
</tr>
<tr>
<td>Internal consistency</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cronbach’s $\alpha$</td>
<td>72</td>
<td>27,106</td>
<td>0.75</td>
<td>[0.72, 0.77]</td>
</tr>
<tr>
<td>Correlations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uncaring</td>
<td>57</td>
<td>19,350</td>
<td>0.45</td>
<td>[0.40, 0.50]</td>
</tr>
<tr>
<td>Total ICU</td>
<td>35</td>
<td>13,358</td>
<td>0.79</td>
<td>[0.76, 0.82]</td>
</tr>
<tr>
<td>Measures of psychopathy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affective</td>
<td>26</td>
<td>7,993</td>
<td>0.40</td>
<td>[0.32, 0.47]</td>
</tr>
<tr>
<td>Behavioral</td>
<td>21</td>
<td>5,705</td>
<td>0.35</td>
<td>[0.32, 0.38]</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>22</td>
<td>6,272</td>
<td>0.38</td>
<td>[0.34, 0.42]</td>
</tr>
<tr>
<td>Total score</td>
<td>16</td>
<td>4,253</td>
<td>0.43</td>
<td>[0.37, 0.48]</td>
</tr>
<tr>
<td>Externalizing</td>
<td>22</td>
<td>13,230</td>
<td>0.18</td>
<td>[0.14, 0.22]</td>
</tr>
<tr>
<td>Reactive aggression</td>
<td>16</td>
<td>4,544</td>
<td>0.35</td>
<td>[0.28, 0.42]</td>
</tr>
<tr>
<td>Proactive aggression</td>
<td>15</td>
<td>4,004</td>
<td>0.44</td>
<td>[0.39, 0.49]</td>
</tr>
<tr>
<td>Nonviolent delinquency</td>
<td>18</td>
<td>3,952</td>
<td>0.30</td>
<td>[0.21, 0.39]</td>
</tr>
<tr>
<td>Hyperactivity</td>
<td>14</td>
<td>8,054</td>
<td>0.36</td>
<td>[0.30, 0.42]</td>
</tr>
<tr>
<td>Empathy</td>
<td>12</td>
<td>3,592</td>
<td>-0.25</td>
<td>[-0.34, 0.15]</td>
</tr>
</tbody>
</table>

Note. The following abbreviations were used for reporting key meta-analytic statistics: $k =$ number of samples included in the analyses; $n =$ total number of participants across all samples included; ES = pooled effect size (For measures of internal consistency, this refers to the pooled Cronbach’s alpha). For correlations, this refers to the pooled Pearson’s correlation coefficient; 95% CI = 95% confidence interval for the pooled effect size; $I^2 =$ percentage of variation across studies that is due to heterogeneity rather than chance; Egger Test = assessment of publication bias where *$p < .10$ indicates potential for publication bias.
We assessed the external validity of the ICU by examining its associations with psychological and behavioral outcomes. Pooled associations between externalizing outcomes and the total ICU, $r = 0.34$, callous subscale, $r = 0.35$, and uncaring subscale, $r = 0.29$, all had a medium effect size. However, the association between the unemotional subscale and externalizing outcomes was close to zero, $r = 0.05$ (Table 1; Figure 1). This pattern of findings (medium effect sizes for associations with the total ICU, callous subscale, and uncaring subscale but small effect sizes for associations with the unemotional subscale) persisted for specific externalizing outcomes: reactive aggression, proactive aggression, nonviolent delinquency, and hyperactivity (Table 1). Particularly, strong associations were found between proactive aggression and scores on the total ICU, $r = 0.41$, callous subscale, $r = 0.44$, and uncaring subscale, $r = 0.33$, in keeping with consistently observed links between this form of aggression and interpersonal callousness (Muñoz & Frick, 2012; Thomson & Centifanti, 2017). Again, however, associations with the unemotional subscale were near zero, $r = 0.06$.

Pooled associations with internalizing outcomes revealed small positive effect sizes with the total ICU score, $r = 0.17$, callous subscale, $r = 0.18$, uncaring subscale, $r = 0.09$, and unemotional subscale, $r = 0.10$ (Table 1). Last, all scores on the ICU were negatively associated with empathy and had medium effect sizes. The largest negative association was observed for the total ICU, $r = -0.42$, followed by the uncaring subscale, $r = -0.35$, callous subscale, $r = -0.25$, and unemotional subscale, $r = -0.22$ (Table 1; Figure 1).

**Publication Bias**

Examination of the Egger Test revealed no evidence of publication bias for the majority of outcomes. Exceptions include potential publication bias for the internal consistency of the uncaring subscale and total ICU score, correlations among some subscales (uncaring and callous scores, unemotional and callous scores) and for correlations between total ICU and uncaring scores. We also found evidence of publication bias for the correlation between the uncaring subscale and externalizing outcomes. However,
after examining the funnel plots, we determined that these results were being driven by a single very large study (n = 5,092, ~5 times larger than the second largest sample included) skewing the regression line. After removing this study, no evidence of publication bias remained.

**Study Characteristics**

**ICU Respondent.** To investigate the effect of the respondent on ICU scores and associations, we calculated pooled alphas separately for samples that used the self-report versus other-report ICU. As seen in Supplemental Table S2, internal consistency was uniformly higher for studies using the other-report relative to the self-report ICU. The other-report version of the ICU had significantly higher internal consistency for the total ICU, $\alpha = 0.87$, uncaring subscale, $\alpha = 0.85$, and unemotional subscale, $\alpha = 0.80$, than studies using the self-report ICU. Furthermore, while the internal consistency of the total ICU, $\alpha = 0.81$, callous subscale, $\alpha = 0.73$, and uncaring subscale, $\alpha = 0.78$, all reached acceptable levels among studies using the self-report ICU, the same was not true for the unemotional subscale, $\alpha = 0.66$.

By contrast, convergent validity with psychopathy tended to be lower for studies using the other-report ICU. For associations with measures of psychopathy, the association between the interpersonal facet and both total ICU scores and callous subscale scores were significantly smaller when using the other-report ICU (total ICU $\bar{r} = 0.24$, callous subscale $\bar{r} = 0.24$) than the self-report ICU (total ICU $\bar{r} = 0.43$, callous subscale $\bar{r} = 0.41$).

However, external validity was generally higher for the other-report versions of the ICU. Studies using this measure found significantly larger associations between externalizing outcomes and scores on the total ICU $\bar{r} = 0.44$, callous subscale, $\bar{r} = 0.46$, and uncaring subscale, $\bar{r} = 0.41$, relative to studies using the self-report ICU (total ICU $\bar{r} = 0.31$, callous subscale $\bar{r} = 0.32$, uncaring subscale $\bar{r} = 0.25$). We observed similar patterns for associations with hyperactivity and nonviolent delinquency. However, only uncaring subscale scores were more strongly associated with reactive aggression among studies using the other-report ICU, $\bar{r} = 0.36$, than the self-report ICU, $\bar{r} = 0.22$. No differences emerged for associations with proactive aggression.

**Detained Status.** As seen in Supplemental Table S3, the internal consistency of scores on the ICU was largely unaffected by the detained status of the sample. The one exception was that the callous subscale was more internally consistent in detained, $\alpha = 0.81$, than nondetained samples, $\alpha = 0.73$. Additionally, correlations between the uncaring and unemotional subscales were significantly larger in detained samples, $\bar{r} = 0.38$, than nondetained samples, $\bar{r} = 0.27$.

When considering associations with psychopathy, the association between total psychopathy scores and the callous subscale was larger in detained samples, $\bar{r} = 0.55$, than nondetained samples, $\bar{r} = 0.37$. The associations between the affective facet and total ICU score, $\bar{r} = 0.63$, callous subscale, $\bar{r} = 0.52$, and unemotional subscale, $\bar{r} = 0.35$, were significantly larger in comparison with the nondetained sample (total ICU $\bar{r} = 0.34$, callous subscale $\bar{r} = 0.32$, unemotional subscale $\bar{r} = 0.15$).

The ICU and its subscales were equally strongly associated with all externalizing outcomes in both detained and community samples. However, the relationship between internalizing outcomes and scores on the unemotional subscale was positive in nondetained samples, $\bar{r} = 0.11$, but negative in detained samples, $\bar{r} = −0.11$. The association between empathy and the uncaring subscale was also larger in detained samples, $\bar{r} = −0.46$, than nondetained samples, $\bar{r} = −0.29$.

**Gender.** We calculated pooled effect sizes separately for studies that included male-only, female-only, and mixed gender samples. Only two analyses revealed effects of gender. First, as seen in Supplemental Table S4, associations between the affective facet of psychopathy and callous subscale scores were significantly smaller for mixed gender samples, $\bar{r} = 0.32$, in comparison with female-only samples, $\bar{r} = 0.51$. There were no observed differences with male-only samples. There were no effects of gender on relationships with psychological or behavioral outcomes, with the exception of larger associations between internalizing symptoms with both total ICU scores and callous subscale scores in female-only samples (total ICU scores $\bar{r} = 0.29$, callous subscale $\bar{r} = 0.38$) than in mixed gender samples (total ICU score $\bar{r} = 0.15$, callous subscale $\bar{r} = 0.14$). Again, no differences emerged with male-only samples.

**Discussion**

The results of this study provide the first meta-analytic evidence for the internal consistency and validity of the ICU as a research assessment of CU traits. The total ICU score, callous subscale, and uncaring subscale appear to be highly internally consistent and externally valid assessments. Across a variety of samples, these measures are consistent predictors of a variety of clinically serious outcomes in youths, including externalizing behaviors as a class, and reactive and proactive aggression, nonviolent delinquency, and hyperactivity specifically. The scales also predict higher levels of psychopathic traits and reduced empathy. But, the findings also support concerns regarding the validity and utility of the unemotional subscale (Byrd et al., 2013; S. W. Hawes et al., 2014; Henry et al., 2016; Kimonis et al., 2008; Kimonis, Branch, et al., 2013; Roose, Bijttebier, Decoene, Claes, & Frick, 2010; Waller et al., 2015). Although this
subscales maintain consistently acceptable internal consistency across samples and is strongly correlated with total ICU scores, it is not closely associated with most of the outcomes that the ICU typically predicts, including any form of externalizing. In all cases, the relationship between the unemotional subscale and these outcomes is less than or equal to $r = 0.10$. Also unlike callous and uncaring scores, unemotional scores were only weakly associated with both psychopathy total and facet scores.

Moderator variables minimally affected the observed patterns, although a few meaningful patterns emerged. First, the other-report version of the ICU appears to be a more internally consistent assessment than the self-report ICU, especially for the total ICU, uncaring, and unemotional scores. Second, partly consistent with the hypothesis that the callous subscale discriminates best at high levels of CU traits, we found that the callous subscale was more internally consistent and showed stronger associations with measures of psychopathy in detained samples, although it was no better at predicting behavioral outcomes in detained samples. Third, both the detained status and gender of the sample moderated the relationship between ICU scores and internalizing outcomes. Although both effects were small, the association between the unemotional subscale and internalizing outcomes was positive in nondetained samples and negative in detained samples. Additionally, total ICU and callous subscale scores were more strongly associated with internalizing outcomes in female-only samples relative to mixed gender samples. Importantly, these observed differences are only differences of relative magnitude, with the direction and approximate size of effects tending to remain consistent across samples, and versions of the measure. None of these moderator variables accounted for the consistently low external validity of the unemotional subscale score.

What, then, explains the poor performance of the unemotional subscale? The unemotional subscale is smaller than the other two subscales (containing only five items). Nevertheless, its internal consistency is acceptable and only slightly lower than the other subscales. Our results cannot rule out the possibility that the emergence of the three subscales is driven at least in part by method variance (Frick & Ray, 2015). The predominantly positive versus negative wording of the items in the callous and uncaring subscales may drive their seemingly divergent external associations. For example, callous subscale scores tend to be more strongly related to psychopathic traits and externalizing outcomes than uncaring scores, consistent with the argument that endorsements of positively worded items (e.g., “I do not care who I hurt to get what I want”) can better discriminate among children with higher levels of CU traits, resulting in stronger associations with more severe symptomologies (Ray, Pechorro, & Goncalves, 2016). Method variance alone cannot, however, explain the poor performance of the unemotional subscale, which contains similar numbers of positively and negatively worded items.

Another possibility is that the items that comprise the unemotional subscale capture heterogeneous emotional processes with inconsistent relationships to key external constructs. For example, four of the five items in this subscale refer not to the experience of emotion but to the outward expressions of it, for example, “I hide my feelings from others,” and “I do not show my emotions to others.” Respondents may construe such items as referring to being withdrawn, anhedonic, or shy, which are traits not closely linked to psychopathy or antisociality. This could explain the seemingly paradoxical finding that unemotional scores are positively associated with internalizing symptomology, despite internalizing symptoms including increases in emotions like anxiety, sadness, and guilt. However, the positive relationship between the unemotional subscale and internalizing symptoms is nonetheless small ($r = 0.10$), such that the current results cannot be interpreted in support of using the unemotional subscale as an assessment of internalizing symptomologies in children with disruptive behavior.

Another consideration is that overall emotional reactivity and responsiveness may be constructs that are too diffuse to capture easily in a single small scale, particularly because not all forms of emotionality may be equally strongly associated with callousness. Although some early characterizations of psychopathy, the parent construct of the ICU, included global poverty of emotion as a defining trait (Hare, 1998, 1999), empirical research has not generally confirmed the global nature of affective deficits. Instead, psychopathy and callousness appear to be associated with decreases in some forms of emotional experience (de Wied, van Boxtel, Matthys, & Meeus, 2012; Kimonis et al., 2006; Kimonis, Frick, Cauffman, Goldweber, & Skeem, 2012; Loney, Frick, Clements, Ellis, & Kerlin, 2003), but increases in other forms of emotional experiences, for example, anger (Blair, 2012; S. W. Hawes et al., 2014; Hicks & Patrick, 2006; Kimonis et al., 2012; Urben et al., 2016), and unrelated or minimally related to emotions such as happiness and disgust (Dawel, O’Kearney, McKone, & Palermo, 2012; Jones, Happé, Gilbert, Burnett, & Viding, 2010; Marsh & Blair, 2008; Marsh et al., 2011).

In contrast, impoverished experiences of fear are consistently linked to psychopathy and callousness. CU traits have been linked to reduced skin conductance responses to threat (Kimonis, Frick, Muñoz, & Aucoin, 2008), reduced subjectively experienced fear (Jones et al., 2010; Marsh et al., 2011), and trait fearlessness (Pardini, Lochman, & Frick, 2003), even after controlling for severity of conduct problems (Pardini, Lochman, & Powell, 2007). Reduced personal experience of fear in CU youths is also mirrored by their selectively reduced responsiveness to nonverbal fear cues in others (Blair, Buchani, Colledge, & Scott, 2005;
Marsh et al., 2008; Muñoz, 2009; S. F. White et al., 2016). Furthermore, even in the context of CU traits, fear insensitivity remains a significant predictor of antisocial behaviors (Kimonis et al., 2006; Kimonis, Frick, Fazeekas, & Loney, 2006), and incorporation of items indexing deficits in fear responding may improve the prediction of aggressive and delinquent behavior (Cardinale et al., 2017; Kimonis, Frick, Muñoz, & Aucoin, 2007; Lozier, Cardinale, VanMeter, & Marsh, 2014). Together these findings suggest that measurements of reduced fearfulness, rather than global unemotionality, may more precisely capture affective deficits linked to psychopathy, callousness, and associated antisocial behaviors. It is also possible that targeted measures of physiological responses associated with CU traits, such as reduced amygdala activation, might correspond to unemotionality in ways that would illuminate the nature of this subscale (Blair, 2013).

Should the unemotional subscale be excluded from the ICU altogether? Some recent studies have taken this approach, excluding all items from this subscale prior to conducting analyses (S. W. Hawes et al., 2014; Houghton et al., 2013; Waller et al., 2015). The creation of an abridged version of the ICU that includes only the uncaring and callous subscales (which may measure two distinct constructs or which may simply capture CU traits with varying levels of severity) may be a justifiable approach. Alternately, items on the unemotional subscale could be replaced with items more precisely tailored to capture affective deficits associated with callous and uncaring traits as well as external outcomes like aggression and delinquency. Items related to risk insensitivity and fearlessness, for example, could be candidates for inclusion (Kimonis et al., 2007).

These considerations are relevant not only to the ICU scale but to the Limited Prosocial Emotions specifier, which was created from four items on the ICU, two of which index callousness, one which indexes uncaring, and one which indexes unemotionality. Currently, this specifier requires that two of the four items be endorsed in order for Limited Prosocial Emotions to be diagnosed in a child with conduct disorder. As a result, a child who exhibits no callousness, but only unemotionality and uncaring (a lack of concern about performance in important activities) could receive a diagnosis identical to a child exhibiting a lack of empathy and remorse. The results of this meta-analysis suggest that a child who only exhibits shallow affect and uncaring behavior may not represent the same level of risk for externalizing outcomes and aggression as a child in whom overtly callous traits are observed. Even more important from a treatment standpoint, symptoms of conduct disorder in a child with only uncaring and unemotional symptoms may reflect a distinct etiology—such as an internalizing disorder or anhedonia—than symptoms in children in whom the remaining items in the specifier are observed (Henry et al., 2016).

Some limitations should be considered in interpreting the present findings. Data from only three unpublished samples (N = 5,534) were available for inclusion in our analyses. As such we cannot rule out the possibility of publication bias in our results, especially for our moderator analyses in which the number of samples was often small (as few as two studies in some instances). Mitigating this concern, however, were results of the Eggers Test, which revealed little evidence that publication bias was influencing the majority of our analyses across all samples. Furthermore, the finding that the unemotional subscale has small relationships with external criteria is unlikely to be affected by the inclusion of unpublished data, given that unpublished findings tend to be biased toward small or nonsignificant effects.

Additionally, the current study only examines one form of reliability. As such our conclusions are limited to the internal consistency of the ICU and cannot be used to infer interrater reliability or temporal stability. There may also be influential moderator variables that were not examined in the current study. For example, potential moderating effects of interview versus self-report-based ratings could be investigated in future work through comparison of the ICU and interview measures of CU traits, including the Hare Psychopathy Checklist: Youth Version (PCL:YV, Forth et al., 2014) and the recently developed Clinical Assessment of Prosocial Emotions (CAPE; Frick, 2013). Because the available data did not allow us to directly test for the effects of this and other possible moderating variables, we attempted to account for unmeasured variables through the use of random-effects models, which account for expected heterogeneity in effect sizes due to differences in the study populations and procedures (e.g., socioeconomic status, age of the participants, experimental setting). As a result, pooled effect sizes represent an estimate of the average effect (not a common effect) that is modeled to allow for variations due to unmeasured external criteria (Riley, Higgins, & Deeks, 2011). As such, the present results are not as directly generalizable across contexts, populations, and testing procedures as would be the case for a fixed-effects meta-analysis in which no heterogeneity was present.

Caution is therefore warranted in extending the conclusions of this meta-analysis to the use of the ICU outside of the research setting. Various factors may contribute to psychological assessments performing more weakly in the field relative to laboratory settings (Edens & Boccaccini, 2017). Such assessments include Hare’s Psychopathy Checklist–Revised (Hare, 2003), which in some legal and forensic settings has been found to demonstrate lower interrater reliability (Edens, Boccaccini, & Johnson, 2010) and smaller effect sizes for the prediction of recidivism (S. W. Hawes, Boccaccini, & Murrie, 2013). Our findings also cannot be used to draw conclusions about the origins of CU traits, which, like most other clinical constructs, exhibit a
developmental trajectory that reflects the influence of both heritable risk factors and various contextual factors and life experiences, including parenting and socioeconomic status, all of which can influence the emergence of CU traits and subsequent behavioral problems (Waller & Hyde, 2017). Therefore, while our results generally support the utility of the ICU, and specifically the total ICU, callous subscale, and uncaring subscale scores, further research is needed to determine how the ICU should best inform clinical, forensic, or correctional treatment and decision making.

Despite some limitations, the results of this meta-analysis support strong internal consistency, convergent validity, and external validity of the total ICU, and its callous and uncaring subscales, as assessments of CU traits. However, our results also support concerns regarding the validity and utility of the unemotional subscale. Despite the unemotional subscale’s acceptable internal consistency, the correlations between this subscale and measures of psychopathy as well as aggression and nonviolent delinquency, were clearly divergent from the patterns otherwise observed. These findings have important implications for understanding the structure of the ICU and for the conceptualization of unemotionality as central to the construct of CU traits.

**Authors’ Note**

There has been no prior dissemination of the data presented in this article. For access to all underlying research materials please contact the first author.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by NIH/NICHD R03 HD064906-01.

**References**

References marked with “*” were used in the meta-analysis.


