An Examination of the Parent Report Version of the Inventory of Callous-Unemotional Traits in a Community Sample of First-Grade Children

By: Michael T. Willoughby, W. Roger Mills-Koonce, Daniel A. Waschbusch, Nisha C. Gottfredson, the Family Life Project Investigators


Made available courtesy of Sage Publications: http://dx.doi.org/10.1177/1073191114534886

***© The Authors. Reprinted with permission. No further reproduction is authorized without written permission from the authors & Sage Publication. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. ***

Abstract:

Background. The Inventory of Callous-Unemotional Traits is a self- and other report questionnaire of callous-unemotional behaviors that is increasingly widely used in research and clinical settings. Nonetheless, questions about the factor structure and validity of scales remain. Method. This study provided the first large-scale (N = 1,078) investigation of the parent report version of the Inventory of Callous-Unemotional Traits in a community sample of school-age (first-grade) children. Results. Confirmatory factor analysis indicated that a two-factor model that distinguished empathic-prosocial (EP) from callous-unemotional (CU) behaviors provided the best fit to the data. EP and CU were moderately to strongly correlated with each other (ϕ = −.67, p < .001) and with oppositional defiant disorder and conduct disorder (ODD/CD) behaviors (ϕ_{ODD/CD, EP} = −.55; ϕ_{ODD/CD, CU} = .71, ps < .001). Individual differences in EP and CU behaviors explained unique variation, beyond that attributable to ODD/CD behaviors, in peer-, teacher-, and parent relationship quality. Moreover, whereas EP moderated the effects of ODD/CD in the prediction of student–teacher relationship quality, CU moderated the effects of ODD/CD in the prediction of peer and parent relationship quality. Conclusions. Results are discussed with respect to the use of the ICU with school-age children.

Keywords: callous-unemotional | conduct problems | Inventory of Callous-Unemotional Traits | factor structure

Article:

Callous-unemotional (CU) behaviors refer to low empathy and fear, poor recognition of fear or distress in others, reduced reactivity to challenging events, and possibly an overfocus on reward and insensitivity to punishment (Frick & White, 2008; Kotler & McMahon, 2005). CU behaviors
are one dimension of the broader construct of psychopathy, which also includes narcissism and impulsive-antisocial behavior. Over the past decade, there has been interest in downward extensions of CU to child and adolescent samples (Frick & Viding, 2009). There is growing evidence in support of using CU to understand better the heterogeneity among youth with conduct problems (Frick, Ray, Thornton, & Kahn, 2013), which was the primary rationale for the inclusion of the “limited prosocial emotions” specifier into the revised diagnostic criteria for conduct disorder (CD; APA, 2013b).

A decade ago, the Antisocial Process Screening Device (APSD; Frick & Hare, 2002) was among the most frequently used questionnaires to measure CU in childhood and adolescent samples. Frick (2004) subsequently developed the Inventory of Callous-Unemotional traits (ICU) to overcome the limitations of the CU subscale of APSD. Limitations of the CU subscale of the APSD included (a) the small number (six) of items on the CU subscale that contributed to poor to modest internal consistency, (b) the use of a 3-point Likert-type rating scale that limited scale variability, and (c) the predominant use of positively worded items that had the potential to contribute to response set bias (summarized in Kimonis et al., 2008). In contrast to the CU subscale of the APSD, the ICU included a mix of 24 positively and negative worded items, all of which were rated on a 4-point Likert-type rating scale. The 4 items from the CU subscale of the APSD that consistently demonstrated the best psychometric properties were used to generate items for each of the proposed four factors of the ICU. The availability of parent report, teacher report, and youth self-report versions of the ICU in the public domain has resulted in this scale being increasingly widely used in clinical and research settings.

Nearly all of the initial psychometric work on the ICU focused on the youth (adolescent) self-report version. Whereas the original development of the ICU posited four factors (which corresponded to the four items that were drawn from the APSD), the initial empirical investigation indicated that a bifactor model, which included a general factor on which all items loaded and three specific (callousness, uncaring, unemotional) factors on which subsets of items loaded, provided the best fit to the data (Essau, Sasagawa, & Frick, 2006). Three subsequent studies came to the same conclusions (Fanti, Frick, & Georgiou, 2009; Kimonis et al., 2008; Roose, Bijttebier, Decoene, Claes, & Frick, 2010). These results provided empirical support for the creation of both overall and subscale scores for the ICU and demonstrated cross-cultural (participants were American, Dutch, German, and Greek adolescents) consistency of results. However, a closer inspection of these results revealed that although the bifactor model provided the best relative fit of the models considered in each study (i.e., it fit better than simple one-[undifferentiated] and three-factor models), the bifactor model exhibited mediocre global model fit in all three studies (e.g., significant likelihood ratio tests of global fit, root mean square error of approximation [RMSEA] ≥ .07, comparative fit index [CFI] < .95) and required either the omission of items (e.g., Kimonis et al., 2008) or the inclusion of a substantial number of residual correlations (e.g., Essau et al., 2006) to achieve this mediocre fit. Four more recent studies, which have relied on a combination of exploratory and confirmatory factor methods, have
reported mixed results regarding the factor structure of youth (and in one case adult) self-reports of the ICU, with researchers advocating for between two and five factors (Byrd, Kahn, & Pardini, 2013; Feilhauer, Cima, & Arntz, 2012; Houghton, Hunter, & Crow, 2013; Kimonis, Branch, Hagman, Graham, & Miller, 2013).

Far less research has been conducted on the parent report version of the ICU. A small subset of the adolescent participants in the Roose et al. (2010) study had parent reports on the ICU. Despite very small sample sizes (Ns < 60), the authors reported that the bifactor parameterization was preferable to the simple one- and three-factor models—though global model fit was again mediocre. Hawes et al. (2014) recently reported the results of a larger scale (N = 250) evaluation of the parent report version of the ICU using a clinical sample of boys with significant conduct problems. Consistent with other recent studies, they noted that the previously proposed factor structures for the ICU fit their data poorly. As such, they used item response theory methods to identify and discard 12 ICU items that exhibited poor psychometric properties. A two-factor model, which distinguished callous from uncaring items, provided excellent fit for the remaining 12 items. Hawes et al. (2014) indicated that whereas the callous items discriminated among higher levels of CU, uncaring items discriminated at lower levels of CU. For the purposes of the present study, it is noteworthy that whereas all of the items on their revised callous scale represented the presence of callous behaviors, all of the items on their revised uncaring scale represented the presence of empathic and prosocial (EP) behaviors (these items were reverse-scored to indicate uncaring behaviors).

The distinction between items that characterize EP versus CU behaviors is potentially of both theoretical and practical importance. A “limited prosocial emotions” specifier was introduced into the recently revised diagnostic criteria for CD (APA, 2013b). Per the Diagnostic and Statistical Manual of Mental Disorders (5th ed.; DSM-5) documentation, this specifier “applies to those individuals with a more serious pattern of behavior characterized by a callous and unemotional interpersonal style across multiple settings and relationships” and is intended to facilitate individualized treatment planning and to spur research (APA, 2013a). The specifier was operationalized as a lack of remorse or guilt, evidence of callousness/lack of empathy, unconcern about performance, or shallow and/or deficient affect. These indicators implicitly treat low EP and high CU behaviors as conceptually equivalent. A primary objective of this study is to empirically evaluate this assumption. Two previous studies indirectly acknowledged the distinction between the absence of EP and the presence of CU behaviors. Although both Houghton et al. (2013) and Hawes et al. (2014) indicated that a two-factor solution provided the best relative fit to their ICU item data, they labeled their latent variables callous and uncaring. Although this terminology was consistent with that used by Frick (2004) in the development of the ICU, it obscured the fact that the items on the ICU that were used to measure uncaring behaviors represent the presence of EP behaviors (i.e., uncaring behaviors were inferred from low ratings on items that described the presence of EP behaviors). In addition to testing the one-,
three-, and bifactor models that have been tested in previous studies, the current study also considered a two-factor model that distinguished EP and CU behaviors.

To the extent that EP and CU factors are dissociable, this would raise questions about whether the “limited prosocial emotions” specifier for CD should be operationalized as strictly the absence of EP behaviors, the presence of CU behaviors, or their combination. If EP and CU behaviors were dissociable, this would also raise the question of whether individual differences in (the absence of) EP and/or (the presence of) CU behaviors were uniquely and perhaps differentially associated with relevant outcomes (e.g., relationship quality across persons and settings), as well as whether EP and/or CU behaviors moderated the effect of conduct problems on these outcomes. A secondary objective of this study was to test these questions.

In sum, this study tested competing factor structures for the parent report version of the ICU. In addition to considering the one-, three-, and bifactor model specifications that have been the focus of most previous studies, which primarily considered the adolescent self-report version of the ICU, this study also tested a two-factor model that distinguished EP from CU behaviors. Building on the results of Hawes et al. (2014), we hypothesized that this two-factor model would provide the best relative fit to the data. A secondary objective of this study was to test whether individual differences in behaviors from the ICU (especially EP and CU) moderated the effects of oppositional defiant disorder and conduct disorder (ODD/CD) problems in the prediction of children’s relationship difficulties with their parents, teacher, and peers. Consistent with the introduction of the limited prosocial emotion qualifier in the DSM-5 criteria for CD, we hypothesized that among children with elevated levels of ODD/CD, those who also exhibited low levels of EP and/or high levels of CU would experience the greatest relationship difficulties.

Method

Participants

The Family Life Project is a prospective longitudinal study of families residing in six low-wealth counties in Eastern North Carolina and Central Pennsylvania (three counties per state) that were selected to be indicative of the Black South and Appalachia, respectively. Complex sampling procedures were employed to recruit a representative sample of 1,292 children whose families resided in one of the six counties at the time of the child’s birth. Low-income families in both states and African American families in North Carolina were oversampled; however, through the use of weighted analyses, all of our inferences generalize back to the six-county study area as if participants were selected using simple random sampling. Full details of the sampling plan and study design appear elsewhere (Vernon-Feagans, Cox, & the Family Life Project Investigators, 2013).

The current study included children (N = 1,078; 50% male; M = 7.3 SD = 0.3 years old) whose families participated in the first-grade home visit and who had nonmissing
parent-rated disruptive and CU behaviors, which represented 83% of the total sample. Families and children who were enrolled in the study but who did not participate in the first-grade assessments (N = 214) did not differ from study participants (N = 1,078) with respect to state of residence (42% vs. 40% residing in Pennsylvania, p = .64), living in a household that was recruited into the low-income stratum (76% vs. 78% poor, p = .47), primary caregiver educational status at study enrollment (80% vs. 80% with a high school degree/GED or beyond, p = .93), child gender (55% vs. 50% male, p = .22), or race (37% vs. 44%, p = .10).

**Measures**

**Inventory of Callous-Unemotional Traits Rating Scale.** The ICU rating scale (Frick, 2004) consists of 24 items, each rated on a 4-point Likert-type scale (0 = not at all, 1 = just a little, 2 = pretty much, 3 = very much), that are intended to represent individual differences in CU behaviors. This study used parent reports of the 24 ICU items at the first-grade home visit. The factor structure and psychometric properties of this instrument were the focus of this study.

**Disruptive Behavior Disorders (DBD) Rating Scale.** The DBD (Pelham, Evans, Gnagy, & Greenslade, 1992; Wright, Waschbusch, & Frankland, 2007) consists of DSM-5-referenced symptoms for ADHD (attention deficit hyperactivity disorder), ODD, and CD that are rated on 4-point Likert-type scale (0 = not at all, 1 = just a little, 2 = pretty much, 3 = very much). This study used parent reports for eight ODD and nine CD symptoms (select CD symptoms that were inappropriate for first-graders were omitted) from the first-grade home visit as indicators of a latent variable of ODD/CD problems.

**Strengths and Difficulties Questionnaire (SDQ).** Parents and teachers completed the SDQ when the children were in first grade. The SDQ consists of 25 items, each of which is rated on a 3-point Likert-type rating scale (0 = not at all, 1 = just a little, 2 = very much), that are combined to form five 5-item subscales, including emotional, conduct, hyperactivity, and prosocial behaviors, as well as peer relationship difficulties. The reliability and validity of the SDQ are well established (Goodman, 2001; Goodman & Goodman, 2009). In this study, both the parent-reported (N = 1,078; M = 0.3, SD = 0.3; α = .54) and the teacher-reported (N = 892; M = 0.3, SD = 0.3; α = .53) peer problems (e.g., child prefers to play along; child gets picked on or bullied by others) scales were used as indicators of the latent construct of peer relationship difficulties. Although the modest levels of internal consistency for peer scales would typically be a source of concern (i.e., if these scales were used as measured outcomes), this was not the case here because these scales were used as indicators of a broader latent construct of peer relationship difficulties.

**Peer Relationship Ratings.** Teachers rated four items on a 6-point Likert-type scale (0 = almost never, 1 = rarely, 2 = sometimes, 3 = often, 4 = very often, 5 = almost always). These items were derived from the Excluded by Peers subscale of the Child Behavior Scale developed by Ladd and Profilet (1996). The items included “is liked by classmates,” “is disliked by classmates,” “is left out or ignored by classmates,” and “is teased or picked on by classmates.” Consistent with our
recent work (Willoughby, Blanton, & Family Life Project Investigators, 2013), we operationalized peer impairment by a combination of not being liked (i.e., ratings of 0 = almost never or 1 = rarely for “is liked by classmates”) and being activity disliked (i.e., ratings of 2 = sometimes or greater for “is disliked,” “is left out or ignored,” or “is teased or picked on”); 3% (N = 23/892) of the children met this criterion. We also considered a dichotomous rating of whether the child had at least one friend (“Regardless of whether this child is popular or unpopular, does she or he have a special, close, ‘best friend’?”); 69% (N = 616/892) of children were reported to have at least one close friend.

Student–Teacher Relationship Scale (STRS). The STRS (Pianta, 2001) is a 28-item measure of teacher-rated perceptions of their relationships with individual students. Each item was rated on a 5-point Likert-type scale (1 = definitely does not apply to 5 = definitely applies). Given previous concerns about the convergent validity and discriminant validity of the dependency scale (Doumen et al., 2009), teachers only completed the 8-item closeness (N = 892; M = 4.2, SD = 0.7; α = .84) and the 7-item conflict (N = 892; M = 1.7, SD = 0.9; α = .92) scales of the STRS in this study.

Parenting Daily Hassles (PDH) Rating Scale. The PDH (Crnic & Greenberg, 1990) is a 20-item self-report measure of perceived parenting stressors and hassles. Parents rated the frequency (5-point Likert-type rating from 1 = never to 5 = constantly) and intensity (5-point Likert-type rating from 1 = no hassle to 5 = big hassle) of each item (e.g., “The kids demand that you entertain or play with them”; “difficulties getting privacy”). The total intensity score (mean of 20 intensity ratings) was used as an indicator of parent relationship quality (N = 479; M = 2.3, SD = 0.8; α = .92). Note that this scale was available only for a subset of the sample; it was eliminated in an effort to reduce the overall length of the first-grade home visit to 2.5 hours.

Caregiver Helplessness Rating (CHR) Scale. The CHR (George & Solomon, 2011) is a 25-item self-report measure of parent’s impression of their relationship with their child. Each item is rated on a 5-point Likert-type scale (1 = not at all characteristic to 5 = very characteristic). We used the 12-item caregiver fear and helplessness subscale (“When I am with my child, I often feel out of control”; “I often feel like there is nothing I can do to discipline my child”; “I feel that I am a failure as a caregiver”) as an indicator of parent relationship quality (N = 479; M = 1.3, SD = 0.4; α = .76). Like the PDH, this scale was available only for a subset of the sample (i.e., it was eliminated from the protocol to reduce the overall length of the first-grade home visit).

Analytic Strategy

The first research question concerned the factor structure of the ICU and was tested using confirmatory factor analytic (CFA) models. Five CFA models were considered- one-factor (undifferentiated structure), two-factor (EP, CU), three-factor (callous, unemotional, uncaring), and two bifactor models (general, callous, unemotional, uncaring; general, EP, CU). The weighted least squares means and variance adjusted (WLSMV) estimator was used for all CFA
models to accommodate the ordinal nature of item-level data (i.e., ICU and DBD items were rated on 4-point Likert-type scales). The second research question concerned the contributions of EP, CU, and ODD/CD behaviors to the prediction of parent, peer, and teacher relationship quality and was tested using structural equation models (SEMs). Latent interactions were estimated using the latent moderated structural equation approach, which uses a quasi-maximum likelihood estimator (Klein & Muthen, 2007).

For models involving the WLSMV estimator, model fit was evaluated using the likelihood ratio chi square test, as well as CFI and RMSEA fit indices, where values of CFI ≥ .95 and RMSEA < .05 were indicative of good fit (Hu & Bentler, 1999; Yu, 2003). Indices of global fit were not available for models involving the quasi-maximum likelihood estimator. All models were estimated using Version 7.1 of Mplus software (Muthen & Muthen, 1998-2013).

Results

Factor Structure of the ICU

Five CFA models were fit to the ICU items (see Table 1 for a synopsis of model fit). Whereas the simple one- (undifferentiated), two- (EP, CU), and three- (uncaring, unemotional, callous) factor models converged, estimation problems were encountered for both bifactor models. Whereas one of the bifactor models (i.e., that with a general factor and three specific factors: callous, uncaring, and unemotional) did not converge, the other bifactor model (i.e., that with a general factor and two specific factors: EP and CU) converged but had a not positive definite psi matrix, which implied model overfitting (i.e., more latent factors were estimated than was necessary). Although the likelihood ratio test statistics were statistically significant for the one-, two-, and three-factor models, the fit statistics favored the two-factor over the one- (undifferentiated) and three-factor (callous, uncaring, unemotional) models. Considering the results of the two-factor model, significant latent variances indicated that there were individual differences in both EP and CU behaviors (ϕ2s = .63 and .68, respectively, ps < .0001). The EP and CU factors were negatively correlated (ϕ = −.66, p < .001). Item statistics for all ICU items are summarized in Table 2. With the exception of Item 10 (standardized λ = .01, p = .86), all items had statistically significant and moderate- to large-sized factor loadings on their respective factors. Given our preference to evaluate the ICU as designed, we retained Item 10 in subsequent models, despite the nonsignificant factor loading.

Table 1. Synopsis of CFA Model Fit for ICU Items.

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>χ²(df)</th>
<th>p</th>
<th>CFI</th>
<th>RMSEA [90% CI]</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>One factor (undifferentiated)</td>
<td>2740.7 (252)</td>
<td>&lt;.0001</td>
<td>.87</td>
<td>.10 [.09, .10]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2</td>
<td>Two factor (EP, CU)</td>
<td>1447.7 (251)</td>
<td>&lt;.0001</td>
<td>.94</td>
<td>.07 [.06, .07]</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3</td>
<td>Three factor (callous, uncaring, unemotional)</td>
<td>2115.7 (249)</td>
<td>&lt;.0001</td>
<td>.90</td>
<td>.08 [.08, .09]</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
Table 2. Observed Frequencies and Factor Loadings for ICU Items.

<table>
<thead>
<tr>
<th>Items</th>
<th>Rating (%)</th>
<th>Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not at all</td>
<td>Some</td>
</tr>
<tr>
<td>Tries not to hurt others’ feelings</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>Apologizes to persons he/she has hurt</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Works hard on everything</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Does things to make others feel good</td>
<td>2</td>
<td>22</td>
</tr>
<tr>
<td>Is concerned about the feelings of others</td>
<td>4</td>
<td>17</td>
</tr>
<tr>
<td>Always tries his/her best</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>Feels bad/ guilty . . . done something wrong</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>It is easy to tell how he/she is feeling</td>
<td>3</td>
<td>23</td>
</tr>
<tr>
<td>Expresses his/her feelings openly</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>Easily admits to being wrong</td>
<td>13</td>
<td>56</td>
</tr>
<tr>
<td>Is very expressive and emotional</td>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td>Is concerned about schoolwork</td>
<td>13</td>
<td>30</td>
</tr>
<tr>
<td>Does not let feelings control him/her</td>
<td>35</td>
<td>47</td>
</tr>
<tr>
<td>Seems very cold and uncaring</td>
<td>92</td>
<td>6</td>
</tr>
<tr>
<td>. . . not care who he/she hurts . . . to get .</td>
<td>83</td>
<td>13</td>
</tr>
<tr>
<td>Does not care if he/she is in trouble</td>
<td>74</td>
<td>20</td>
</tr>
<tr>
<td>Does not care about doing things well</td>
<td>76</td>
<td>19</td>
</tr>
<tr>
<td>Shows no remorse when . . . wrong</td>
<td>79</td>
<td>15</td>
</tr>
<tr>
<td>Does not show emotions</td>
<td>85</td>
<td>11</td>
</tr>
<tr>
<td>Does not like to put the time into . . .</td>
<td>61</td>
<td>31</td>
</tr>
<tr>
<td>. . . not . . . know “right” from “wrong”</td>
<td>82</td>
<td>12</td>
</tr>
<tr>
<td>The feelings of others are unimportant . .</td>
<td>82</td>
<td>11</td>
</tr>
<tr>
<td>Hides his/her feelings from others</td>
<td>68</td>
<td>25</td>
</tr>
<tr>
<td>Does not care about being on time</td>
<td>63</td>
<td>26</td>
</tr>
</tbody>
</table>

Note. ICU = Inventory of Callous-Unemotional Traits; EP = empathic-prosocial; CU = callous-unemotional. N = 1,078. All factor loadings significant at p < .001, except Item 10.

Criterion Validity

A six-factor CFA model was estimated that included latent variables for EP, CU, ODD/CD behaviors, as well as three latent variables that represented peer, teacher, and parent relationship
difficulties. The six-factor model fit the data well, $\chi^2(1, 100) = 2580.1, p < .0001$, CFI = .95, RMSEA [90% confidence interval] = .035 [.033, .037]. All of the latent variances and factor loadings (including Item 10 on the ICU, standardized $\lambda = .18$, $p = .003$) were statistically significant ($ps \leq .001$). As is summarized in Table 3, individual differences in EP and CU behaviors were moderately to strongly correlated with each other ($\phi = -.67$) and with ODD/CD behaviors ($\phi = -.55$ and $.71$, respectively). Moreover, EP, CU, and ODD/CD behaviors were all moderately to strongly correlated with peer, teacher, and parent relationship problems ($|\phi s| = .30-.63, ps < .001$).

### Figure 1. Main effects structural equation model.

A SEM was then estimated in which the latent variables of peer, teacher, and parent relationships were regressed on the latent variables of EP, CU, and ODD/CD. This model was identical in fit to the six-factor CFA that was reported above; however, it provided a test of the unique contributions of each dimension of behavior in the prediction of relational outcomes. As summarized in Figure 1, individual differences in EP ($\beta = -.14, p = .007$), CU ($\beta = .39, p < .001$), and ODD/CD ($\beta = .16, p = .009$) behaviors all made unique contributions to the prediction of peer relationship difficulties (latent $R^2 = .39$). Individual differences in EP ($\beta = -.33, p < .001$) and ODD/CD ($\beta = .17, p = .02$), but not CU ($\beta = .01, p = .92$), behaviors made unique contributions to teacher-student relationship quality (latent $R^2 = .19$). Finally, individual differences in EP ($\beta = .22, p = .01$), CU ($\beta = .39, p = .001$), and disruptive ($\beta = .47, p < .001$) behaviors all made unique contributions to the prediction of parent relationship quality (latent $R^2 = .45$); however, the point estimate for EP behaviors was in the opposite direction (sign) of the observed bivariate association (compare path in Figure 1 to corresponding value in Table 3). This was indicative of
a suppression effect, multicollinearity among predictors, and/or model misspecification (e.g., omission of an interaction term).

To test the relative magnitude of EP and CU contributions to each of the three outcomes, the SEM was reestimated three times (once per outcome). In each model, an equality constraint was imposed on the regression coefficients that linked EP and CU to a given outcome. A nested likelihood ratio test indicated whether the imposition of an equality constraint resulted in worse model fit relative to the baseline model that had not imposed any equality constraints (significant tests indicated that EP and CU exerted differential effects for a given outcome). CU was more strongly associated with peer problems than EP, $\chi^2(1) = 38.3, p < .001$. CU and EP were not differentially predictive of parent relationship quality, $\chi^2(1) = 3.4, p = .06$. EP was more strongly associated with student–teacher relationship quality than was CU, $\chi^2(1) = 21.5, p < .001$.

The final set of SEMs tested whether EP and/or CU moderated the effect of ODD/CD behaviors for each outcome. This was intended to approximate the use of EP and/or CU as “specifiers” of the effects of ODD/CD on peer, parent, and teacher relationship outcomes. For each outcome, an initial model included main effects of ODD/CD, EP, and CU, along with both ODD/CD × EP and ODD/CD × CU interaction terms. A subsequent model omitted those interaction terms that were clearly not statistically significant. There was evidence of statistically significant ODD/CD × CU interactions for peer and parent relationship outcomes ($ps = .001$ and $.013$, respectively). Conversely, there was evidence of a statistically significant ODD/CD × EP interaction ($p = .004$) for student–teacher relationship quality. The conditional associations between ODD/CD behaviors and each outcome are depicted in Figures 2 to 4. For each outcome, increasing levels of ODD/CD behaviors were more strongly associated with relationship problems at either (a) higher versus lower levels of CU or (b) lower versus higher levels of EP. These results supersede those reported in Figure 1 (i.e., the main effects of ODD/CD, EP, and CU on outcomes are replaced here by conditional effects).
**Figure 2.** Conditional association of ODD/CD and peer relationship difficulties by CU. Note. ODD/CD = oppositional defiant disorder and conduct disorder; CU = callous-unemotional.

**Figure 3.** Conditional association of ODD/CD and parent helplessness and hassles by CU. Note. ODD/CD = oppositional defiant disorder and conduct disorder; CU = callous-unemotional.

**Discussion**
The ICU is increasingly used by researchers and clinicians who wish to measure individual differences in CU behaviors, often as a means for subdividing youth with elevated conduct problems. Although previous psychometric evaluations of the ICU had primarily involved adolescents’ self-reports of their CU behaviors, there is also growing interest in the measuring of CU in childhood (Viding & McCrory, 2012). The primary purpose of this study was to present the first large-scale evaluation of parent ratings of the ICU involving a representative community sample of first-grade children.

A two-factor model that distinguished EP from CU behaviors provided the best relative fit to the data. The results of the two-factor model indicated that (a) there were individual differences in parent ratings of both EP and CU behaviors, (b) EP and CU behaviors were moderately negatively correlated, and (c) all but one item exhibited statistically significant factor loadings on the intended factor (this item subsequently had a significant loading in models that included outcomes). Our results were consistent with those of Hawes et al. (2014) who evaluated parent ratings of the ICU using a clinic sample of boys with conduct problems, as well as with those of Houghton et al. (2013) who evaluated child self-reports of ICU. An important distinction is that whereas we characterized low ratings on some ICU items as indicative of the absence of empathetic and prosocial behaviors, previous studies have characterized these same items as indicative of the presence of uncaring behaviors (by reverse-scoring them). Although this distinction is entirely semantic, we believe that it is important, particularly in light of the recent modification to the diagnostic criteria for CD. The newly introduced limited prosocial emotions specifier for CD treats low EP and high CU behaviors as conceptually equivalent. To the extent that the ICU items are characterized as “callous” or “uncaring,” this may perpetuate the notion of the equivalence of these behaviors. This contradicts our CFA results (and those of both Hawes et al. and Houghton et al.), which indicated that these were correlated but distinct dimensions of behavior. More generally, the absence of EP behaviors should not be considered as evidence for the presence of CU behaviors.
Figure 4. Conditional association of ODD/CD and student–teacher relationship quality by EP. 

Note. ODD/CD = oppositional defiant disorder and conduct disorder; EP = empathic-prosocial.

To date, most studies have indicated that a bifactor solution provided the best fit to the ICU data. This provided support for the creation of an overall ICU score, along with specific scale scores (e.g., callous, uncaring, unemotional). Our results did not provide support for the creation of an overall scale score. Rather, they provide support for the creation of EP and CU scores. All of the studies that provided support for the creation of an overall ICU score were based on adolescent self-report data. In contrast, our study and that of Hawes et al. (2014), neither of which provide support for the creation of an overall ICU score, were based on parent reports of school-age children. It is conceivable that the factor structure of the ICU may change across development. However, it is equally possible that the factor structure differs by informant. Our study design and sample did not permit us to test these questions, though this is clearly an important direction for future research.

SEMs provided unambiguous support for the incremental value of EP and CU behaviors, above and beyond that attributable to individual differences in ODD/CD behaviors, in the prediction of peer, teacher, and parent relationship problems. Whereas the presence of CU behaviors appeared to be a stronger determinant of parent perceptions of their relationship with children, the absence of EP behaviors appeared to be a stronger determinant of teachers’ perceptions of their relationship with children. Although teachers appeared to prioritize the absence of EP behaviors in their own relationships with children, they indicated that the presence of CU behaviors was a stronger determinant of peer relationship difficulties. It will be important for future research to
replicate these findings, particularly using peerbased metrics of relationship quality (e.g., sociometric ratings). The differential contributions of EP and CU behaviors to parent, peer, and teacher relationship difficulties help validate the distinction between these correlated but dissociable dimensions of behavior.

Although EP and CU explained unique variation in all three outcomes, evidence of ODD/CD × EP (for teacher problems) and ODD/CD × CU (for parent and peer problems) interactions were most directly relevant to questions about how to best operationally define a specifier for CD. For all three of the outcomes considered, higher levels of ODD/CD behaviors were more strongly associated with relationship dysfunction conditional on either lower levels of EP or higher levels of CU behaviors. Building on the CFA results, these SEM results demonstrated the value of distinguishing EP from CU behaviors. The combination of high levels of ODD/CD behaviors with either low levels of EP or high levels of CU behaviors appears to differentially affect relationship quality across contexts. Distinguishing the specific dimensions of EP and/or CU behaviors that interact with ODD/CD may clarify intervention development activities (e.g., whereas increasing EP behaviors may enhance teacher-student relationships, reducing CU behaviors may enhance peer and/or parent-child relationships).

This study had at least three strengths. First, the representative sampling frame and community sample improved generalizability relative to convenience and/or clinic samples. Second, our use of multi-informant assessments (whereas EP, CU, and ODD/CD behaviors were exclusively measured by parents, relationship quality outcomes were primarily—though not exclusively—measured by teachers) helped eliminate shared method variance as an explanation for the reported associations. Third, by addressing all study questions within a latent variable framework, we provided more powerful tests that were not compromised by measurement error. The ability to estimate and test interactions between latent variables was particularly novel.

This study also suffered from at least five limitations. First, although the study was based on a large representative sample, the absolute number of children exhibiting clinically elevated levels of ODD/CD or CU behaviors was small. It will be important for future studies to test the clinical utility of EP versus CU behaviors in children who meet diagnostic criteria for CD. Second, we relied exclusively on parents as informants of ODD/CD, EP, and CU behaviors. This departs from clinical best practice, in which parent and teacher reports of behaviors (and self-reports among adolescents) across settings would be preferred. An important direction for future research involves clarifying how to best use multiple informants of EP and CU behaviors, particularly given that they are often weakly correlated. Third, this sample consisted entirely of first-grade children. Although the measurement of CU behaviors in children this young (and younger) appears valid (Hyde et al., 2013; Willoughby, Waschbusch, Moore, & Propper, 2011), it is unclear whether the ICU necessarily includes the best characterization of EP and CU behaviors for children in this age range; moreover, it is unclear whether the structure of EP and CU items undergoes changes across development. Fourth, parent ratings of relationship quality
were only available for a subset of the sample. Although the reduced sample size was accommodated through full information maximum likelihood estimation, the standard errors of coefficients predicting this outcome may have been increased. Fifth, and most important, whereas all of the items that we characterized as EP were positively worded, all of the items that characterized as CU were negatively worded. Frick et al. (2013) included positively and negatively worded items in the ICU to reduce the likelihood of response set biases—not due to a substantive interest in EP versus CU behaviors. It will be important for future studies to use supplementary sources of information to help disambiguate whether the distinction between EP and CU items that we have advocated here is substantively meaningful or simply a methodological artifact of the way in which items have been written.

In conclusion, this study provided support for a two-factor model of the ICU that distinguished EP from CU behaviors. Consistent with a growing body of research, children characterized by high levels of ODD/CD behaviors and either low EP or high CU behaviors experienced the most peer problems, had the most conflictual relationships with their teachers, and had caregivers who reported the greatest levels of helplessness and parent-related stresses. These results provide general support for the inclusion of the new specifier in the revised diagnostic criteria to CD. However, these results also demonstrated the utility of distinguishing low levels of EP (e.g., lacks empathy) versus high levels of CU behaviors (e.g., lacks remorse). The recently introduced limited prosocial emotions specifier of CD does not explicitly attend to this distinction but perhaps should. Because of its inclusion of both EP and CU behaviors, the ICU appears to be an important tool for clinicians and researchers wishing to distinguish among youth with serious conduct problems.

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: Support for this research was provided by the Eunice Kennedy Shriver National Institute of Child Health and Human Development Grant P01 HD39667, with cofunding from the National Institute on Drug Abuse. The Family Life Project Key Investigators in the second phase of the study included Lynne Vernon-Feagans, Martha Cox, Clancy Blair, Peg Burchinal, Linda Burton, Keith Crnic, Patricia Garrett-Peters, Mark Greenberg, Roger Mills-Koonce, and Michael Willoughby.

References


