Article

A Comparison of Parent, Teacher, and Youth Ratings on the Inventory of Callous–Unemotional Traits

Assessment
1–15

DOI: 10.1177/10731911211047893
journals.sagepub.com/home/asm
SAGE

Tatiana M. Matlasz[1], Paul J. Frick[1,2], and Julia E. Clark[1]

Abstract

The current study compared the validity of self-, parent-, and teacher-report versions of the Inventory of Callous–Unemotional Traits (ICU), a widely used measure of callous–unemotional (CU) traits, at several different ages. Participants (N = 236, 60.6% girls) were children in Grades 3, 6, and 8 (M_{age} = 11.55, SD = 2.23) from a public school system in the southern United States. We tested the association of all three ICU versions with several validators: parent- and teacher-reported conduct problems, peer nominations of characteristics associated with CU traits, and sociometric peer nominations of social preference. Results revealed an interaction between the ICU version and grade in the overall level of CU traits reported, with teacher-report leading to the highest ratings in sixth grade and being higher than parent-report in third grade. Furthermore, the validity of the different versions of the ICU varied somewhat across grades. Specifically, findings support the validity of both teacher- and self-report in third grade, but self-report was the only version to show strong validity in the eighth grade.

Keywords

Inventory of Callous–Unemotional Traits, ICU, assessment, informants, self-report, validity

Callous–unemotional (CU) traits are defined by a lack of remorse or guilt, a lack of empathy, shallow or deficient emotions, and a failure to care about performance in important activities (Frick & Ray, 2015). These traits have been used to define the affective components of psychopathy in adult samples (Hare & Neumann, 2008) and the affective components of conscience in child samples (Frick et al., 2014a). It is important to note that in both adult and child samples there are other dimensions of psychopathy that are highly related to conduct problems (CP) in children (see Salekin, 2017; Sica et al., 2020). For example, Salekin (2017) summarized research on one widely used model of psychopathy that includes CU traits, but also includes an impulsive lifestyle facet, an interpersonal facet, as well as CP themselves. This research suggests that the impulsive lifestyle and the interpersonal facet are highly correlated with CP (see, e.g., Frick et al., 2000; Colins et al., 2018, using two different ways of measuring these constructs in different age groups), making them less useful for specifying subtypes of children with CP, and more suitable as general risk factors for CP or as part of the definition of CP. In fact, they have been used that way for many decades in the classification in childhood psychopathology (see Burns, 2000, for a discussion of this). Specifically, the impulsive lifestyle dimension forms part of the criteria for Attention-deficit/Hyperactivity Disorder (ADHD), which has a long history of research as a risk factor for serious CP (see Waschbusch, 2002). Similarly, the interpersonal facet includes deceitfulness, lying, and conning others as a defining feature, and deceitfulness is one of four symptom clusters included in the diagnostic criteria for conduct disorder (CD), based on a long history of research showing their associations with other covert forms of CP (Frick et al., 1993).[1]

In contrast, a significant amount of research has indicated that the CU dimension of psychopathy is less strongly correlated with CP than the other dimensions of psychopathy (see Salekin, 2017) and it is the one dimension that consistently changes the relation of CP with other important variables, such as their association with emotional reactivity to emotional distress (Kimonis et al., 2006)[1] and various dimensions of parenting (Edens et al., 2008). Thus, children with CP high on these traits often show a number of different correlates and characteristics compared with children with CP who are not high on CU traits (e.g., reduced

[1]Louisiana State University, Baton Rouge, LA, USA
[2]Australian Catholic University, Brisbane, Queensland, Australia

Corresponding Author:
Tatiana M. Matlasz, Department of Psychology, Louisiana State University, 236 Audubon Hall, Baton Rouge, LA, 70803, USA.
Email: tmatla1@lsu.edu

emotional reactivity to distress vs. enhanced emotional reactivity, respectively; Viding et al., 2012), and these findings have led to theories specifying how these different characteristics could reflect different underlying causes to CP in the two groups (Blair et al., 2014; Frick et al., 2014b). These different characteristics have also been used to explain why those with elevated CU traits often end treatment with more severe CP than other youth with CP (Wilkinson et al. 2016)[2]. As a result, unlike the other dimensions of psychopathy, CU traits are better considered as a specifier for CP (i.e., designating important subtypes of children with CP), rather than as a risk factor for or as part of the defining features of CP. Finally, unlike the other dimensions of psychopathy, CU traits have traditionally not been reflected in diagnostic classification systems for childhood psychopathology. To change this, the *Diagnostic and statistical manual of mental disorders–Fifth edition* (*DSM-5*) for the first time added the specifier of "with Limited Prosocial Emotions" to the diagnosis of CD, which is defined by the presence of significant levels of CU traits (American Psychiatric Association [APA], 2013). A similar specifier was added to the diagnoses of Oppositional Defiant and Conduct-Dissocial Disorders in the 11th Edition of the International Classification of Disease (World Health Organization, 2018) to define distinct subgroups of children with these diagnoses based on their level of CU traits.[3]

Thus, many dimensions that have often been associated with psychopathy are considered important for the research and clinical classification of children with CP. However, research on these dimensions clearly shows that they have different associations with CP and as result, should not be combined into a total score when studying them in relation to CP. However, early measures of CU traits were largely confined to subscales on measures of psychopathy and this led to relatively short measures that did not have strong psychometric properties (see Kotler & McMahon, 2005; Sharp & Kine, 2008, for reviews). As it became clear that CU traits should be measured separately from other dimensions of psychopathy, more comprehensive measures of this construct have been developed.

The most widely used measure of CU traits is the Inventory of Callous–Unemotional Traits (ICU; Kimonis et al., 2008). The ICU is a 24-item behavior rating scale, which was designed to overcome limitations of other measures by (a) providing a comprehensive coverage of the key features of CU traits, specifically those included in the DSM-5 specifier; (b) including both positively (i.e., higher ratings indicating higher levels of CU traits) and negatively (i.e., higher ratings indicating lower levels of CU traits) worded items; and (c) using a 4-point rating format that leads to a significant range of responses, while also avoiding a central tendency (i.e., middle rating; Frick & Ray, 2015). Furthermore, the ICU has three versions, including self-, teacher-, and parent-report versions, has been translated into over 25 different languages, and has been used widely in research, with over 250 published studies using the ICU in samples ranging in age from 3 years to young adulthood (https://faculty.lsu.edu/pfricklab/). Cardinale and Marsh (2020) conducted a meta-analysis of 115 samples (N = 27,947) using the self-report ICU and reported that the pooled Cronbach's alpha across all studies was .83 and that the pooled association with general externalizing behavior was r = .34, with proactive aggression was r = .41, and with empathy was r = -.42.

Importantly, while a number of factor analyses of the ICU have identified distinct item clusters (i.e., callousness, uncaring, and unemotional), these factor analyses consistently show that an overarching factor is needed to achieve adequate model fit and that total scores formed by unit weighing of the individual items on the ICU are largely determined by this broad factor for both the self-report and parent-report versions (Ray & Frick, 2018). Furthermore, the subfactors appear to be largely determined by method variance, with a callousness dimension largely composed of items scored in the positive direction (e.g., "I do not care if I get into trouble") and an uncaring dimension largely composed of items that are inversely scored to create total scores from the ICU (e.g., "I try not to hurt others" feelings; Ray et al., 2016). Finally, when method variance is considered in factor analyses, a model that includes the four item clusters (a) from which the ICU was developed, (b) which correspond to the DSM-5 specifier, and (c) which each contribute to an overarching construct, is supported (Kliem et al., 2020; Koutsogiorgi et al., 2020). This factor structure has been shown to be invariant across boys and girls (Kliem et al., 2020).[4]

Thus, while there is support for the validity of the total score of the ICU, most of this work has used the self-report version in adolescent samples (Cardinale & Marsh, 2020; Deng et al., 2019). A few studies have tested the validity of informant versions, with positive findings for the validity of the parent-report (Benesch et al., 2014; Gao & Zhang, 2016; Hawes et al., 2014; Kimonis Fanti, & Singh, 2014; McDonald et al., 2018; Waller et al., 2015; Willoughby et al., 2014) and teacher-report (Ezpeleta et al., 2013; Henry et al., 2018; Kimonis et al., 2016; Roose et al., 2010; Ueno et al., 2019) versions. Deng et al. (2019) provided a meta-analysis on the reliability of the different versions of the ICU (i.e., self-, teacher-, and parent-report) and reported that the internal consistency was higher for parent (α = .83) and teacher (α = .88) versions, compared with the self-report version (α = .79). While these findings suggest that all three versions possess adequate reliability, the differences in reliability could raise concerns about measurement invariance across these versions. However, these findings need to be interpreted cautiously given that the pooled estimate for the self-report version was based on substantially more effect sizes (k = 113) than either parent (k = 23) or

teacher ($k = 7$) reports. As a result, it is possible that the differences across versions may be due to confounds with other variables, such as teacher-report being used more often in younger samples (Deng et al., 2019). Furthermore, there are other pieces of evidence to support the measurement invariance across different versions, such as the similarity in factor structure across the different versions (Roose et al., 2010; Ueno et al., 2019), including the consistent finding supporting a general CU factor (Ray & Frick, 2018).

However, a critical issue for establishing whether or not the construct is measured similarly across the different versions is whether the scores are similarly correlated with theoretically important criteria across self-, parent-, and teacher-report. Unfortunately, the available research utilizing more than one reporter and comparing the correlates to the different versions of the ICU is quite limited. Such tests of measurement validity are important for any measure of personality and psychopathology in children and adolescents, given the reports from different sources are not highly correlated with each other and can each capture important variations in children's emotions and behavior across settings (De Los Reyes & Kazdin, 2005). However, the importance of multiple reporters may be especially critical for assessing CU traits for the diagnosis of the Limited Prosocial Emotions specifier based on the *DSM-5* criteria, which states that characteristics of the specifier should be displayed "in multiple relationships and settings" and should not be "just occasional occurrences in some situations" and, as a result, asserts that "multiple information sources are necessary" for the diagnosis (APA, 2012). Thus, it is important to test the differential validity of various versions of the ICU in the assessment of CU traits to guide this diagnostic process.

In one study that did compare the different versions of the ICU, White, Cruise, and Frick (2009) compared parent- and self-report ratings on the ICU in a sample of 94 adolescent boys ($M_{age} = 15.22$ years, $SD = 1.48$) detained for a sex offense. Both parent- and youth self-report on the ICU were associated with risk factors for later offending, although the parent ratings showed more consistent correlations across the different types of risk factors and accounted for unique variance in risk factors when controlling for self-report, whereas scores from the self-report ICU did not account for unique variance when controlling for parent-report. Gao and Zhang (2016) gave the parent- and self-report versions of the ICU to a sample of 340 boys and girls aged 8 to 10 years. They reported that ratings from both reporters were correlated with measures of behavior problems. However, there were strong effects of shared method variance, with parent-report on the ICU showing stronger correlations with parent-reported behavior problems and youth self-report on the ICU showing stronger correlations with youth-reported behavior problems. Similarly, Berg et al. (2013) compared the correlations between self-report and caretaker-report in a sample of 70 adolescents aged 13 to 17 years in the foster care system. For the most part, caregiver-report on the ICU was most highly correlated with other caregiver-report measures, whereas self-report on the ICU was most highly correlated with other self-report measures. Finally, Lin et al. (2019) compared caregiver and self-report versions of the ICU in a sample of adolescents (aged 12 to 19 years) who were detained for delinquent behavior. As with previous studies showing strong effects of method variance, youth self-report on the ICU was largely correlated with youth self-report of delinquency but not parent-report of externalizing behavior, whereas parent-report on the ICU was associated with parent-reported externalizing behavior but not youth-report of delinquency. However, this study did report that both youth and parent versions were similarly correlated with official records of delinquent behavior, illustrating the importance of having external criteria that do not share methods when evaluating the validity of the different versions of the ICU.

As would be expected from the more limited use of the teacher-report version of the ICU, only a very few studies to date have directly compared the validity of the teacher-report of the ICU with other versions of the scale. Roose et al. (2010) studied 450 adolescents ($M_{age} = 16.67$, $SD = 1.34$) and reported that both the self-report and a composite of the parent and teacher versions were associated with self-reported measures of antisocial behavior, empathy, and prosocial beliefs. Again there was evidence for the effects of shared method variance, with the self-report showing somewhat higher correlations than the informant-report version with the self-reported variables used in this study. Importantly, this study did not report correlations for the parent- and teacher-report versions separately. Docherty et al. (2017) compared self, parent, teacher, and staff reports on the ICU in a sample of community and detained adolescents ($N = 634$; $M_{age} = 16.18$, $SD = 1.31$). Correlations between the versions were fairly low, ranging from $r = .13$ to .24, though ICU scores from all sources were significantly associated with the outcomes of aggression, violence, and detained status with parent-report being a stronger indicator of outcomes compared to self and teacher reports. Similarly, Barhight et al. (2017) compared parent-, teacher-, and self-report of CU traits measured by the ICU in a sample of 95 children in the fourth and fifth grades and reported correlations across versions ranging from $r = .25$ to .38 (Barhight et al., 2017). All three versions were significantly correlated with measures of aggression but again, there was evidence for strong method effects, with versions of the ICU being more highly associated with the same reporter for aggression.

In perhaps the largest test of all three version of the ICU to date, Ueno et al. (2019) collected all three versions in a large sample ($n = 1,339$) of nonreferred German children aged 6 to 18 years. Both the self-report and teacher versions

of the ICU were correlated with another self-report measure of CU traits and all three versions of the ICU were associated with measures of externalizing behavior. Importantly, this study included a large age range but did not test whether the validity of the different raters differed by age. This limitation is important, given suggestions that the importance of teacher-report may be greater in young children but declines as the child approaches adolescence, as any one teacher is less likely to have the child in class throughout the day (De Los Reyes & Kazdin, 2005; Frick et al., 2020). Furthermore, as a child approaches adolescence, the validity of the child's self-report may increase, as the child is better able to report on internal thoughts, feelings, and preferences and the child's behavior may not be as closely monitored by parents and teachers (De Los Reyes & Kazdin, 2005).

Current Study

In summary, despite the need for a multirater assessment of CU traits and the availability of different versions of the ICU for self, parent, and teacher report, much of the research using this scale to date has focused on the self-report version in adolescents, with very few studies testing the validity of the teacher-report version. Furthermore, even fewer studies have compared the validity of the different versions of the ICU to each other and rarely have they included a validator that does not share method (i.e., same reporter) with one of the ICU versions. Finally, these tests of the differential validity of the different versions need to consider potential developmental changes in the relative utility of the different reporters.

To begin to address these important limitations in past work, we compared the validity of the self-, parent-, and teacher-report versions of the ICU in an ethnically diverse sample of children in Grades 3, 6, and 8. First, we investigated whether or not there were differences in the overall level of CU traits across raters and whether any such differences varied by grade. Second, we investigated the correlations between the different versions of the ICU and several important validators. The first validator was parent and teacher reports of CP, which has been one of the most commonly used variables to test the validity of the ICU (Cardinale & Marsh, 2020). While CU traits are less strongly correlated with CP than other dimensions of psychopathy, past research has clearly shown that they are modestly correlated (e.g., around $r = .33$; Frick et al., 2014b). Thus, we predicted based on past work that all three versions of the ICU would be significantly correlated with these ratings of behavior problems but that this would be stronger for the parent and teacher ratings due to shared method variance with our measure of CP. Importantly, we also used several peer nominations as validators in order to have an external criterion that did not share methods with any of the ICU versions to provide a strong test of the differential validity of the various raters. Specifically, we included a well-validated measure of social preference using sociometric peer nominations, as previous research has demonstrated an association between CU traits and peer rejection (Barry et al., 2008; Graziano et al., 2016; Piatgorsky & Hinshaw, 2004; Waller et al., 2017). Furthermore, we included several other peer nominations that were collected specifically to capture the mean and socially aloof characteristics associated with CU traits (Haas et al., 2018; Muñoz et al., 2008). Third, we tested whether raters differed in their validity across grades. Based on past work with measures of other types of personality and behavior (Frick et al., 2020), we predicted that teacher reports would have stronger associations with validators at younger ages. Finally, we tested the incremental validity of the various raters in several ways. Specifically, we tested whether the different ratings would account for unique variance in the different validators, with our hypothesis again being that the incremental contribution of teacher-report would be greatest at the younger ages.

Method

Participants

Participants were 236 children and adolescents recruited from two rural public school systems in the southern United States, from the third ($n = 71$, 30.1%), 6th ($n = 56$, 23.7%), and eighth ($n = 109$, 46.2%) grades. Participants ranged in age from 8 to 15 years ($M = 11.55$; $SD = 2.23$) and 60.6% were girls. By parental report, the sample primarily identified as Black, Afro-Caribbean, or African American (39.4%) and non-Hispanic Caucasian (36.4%), with smaller portions identifying as Biracial (12.3%), Latino or Hispanic American (5.1%), and other ethnic minorities (East Asian or Asian American, 1.7%; Middle Eastern or Arab American, 0.8%, Native American or Alaskan Native, 0.8%; Other, 0.8%). The remaining 2.7% of the sample did not report their race or ethnicity. The number of parents who were married (45.3%) and unmarried (46.6%) was evenly distributed; 8.1% did not report their marital status. In addition, 84.9% of parents reported having less than a college degree and the majority of the sample reported their total household income as less than $40,000 (71.6%).

Measures

Callous–Unemotional Traits. The ICU (Kimonis et al., 2008) is a 24-item measure of CU traits. Items were answered on a 4-point Likert-type scale (0 = *not at all true*, 1 = *somewhat true*, 2 = *very true*, 3 = *definitely true*). Participants missing more than two thirds of the items for any version were removed from the sample ($n = 11$) and this exclusion resulted in the final sample size of 236 children and

Table 1. Correlations Among the Main Study Variables.

			ICU version				Validators			
Variables	*M* (*SD*)	Range	Self	Parent	Teacher	HRC	CP	Social Pref	Mean/Cold	Not Nice
ICU version										
Self	18.99 (8.10)	3.00, 50.00	—	.22**	.29***	.61***	$.30_a$***	−.28***	$.34_a$***	.24***
Parent	16.56 (9.46)	1.00, 45.00	—	—	.18**	.50***	.41***	−.14*	.20**	$.13_a$*
Teacher	20.50 (13.04)	0, 53.00	—	—	—	.78***	.44***	−.20**	.15*	.32***
HRC	34.69 (10.72)	12.00, 66.00	—	—	—	—	.56***	−.32***	.28***	.34***
Validators										
CP	25.02 (7.81)	18.00, 58.00	—	—	—	—	—	−.37***	.37***	.28***
Social Pref	0.10 (1.46)	−6.02, 3.55	—	—	—	—	—	—	−.45***	−.59***
Mean/Cold	0.003 (0.70)	−0.56, 5.17	—	—	—	—	—	—	—	.12
Not Nice	−0.09 (0.90)	−3.75, 1.05	—	—	—	—	—	—	—	—

Note. Significant pairwise comparisons ($p < .05$) are denoted subscript "a," which differs significantly from teacher. ICU = Inventory of Callous–Unemotional Traits; HRC = highest resolved composite; CP = conduct problems; Social pref = social preference.
*$p < .05$. **$p < .01$. ***$p < .001$.

adolescents. For those missing less than a third of the items, their ICU score was prorated using the mean score from the available items. The ICU total score showed high internal consistency for each of the versions: youth self-report ($\alpha = .77$), parent-report ($\alpha = .85$), and teacher-report ($\alpha = .93$), which is very similar to the internal consistency found in past research for these different versions of the scale (e.g., Deng et al., 2019). As shown in Table 1, the correlation between versions ranged from $r = .18$ between parent- and teacher-report to $r = .29$ between self- and teacher-report (all ps $< .01$).

The primary focus of analyses was on the individual associations between the different ICU versions and the validation measures. However, to determine how combining across raters would influence the associations with other variables, a cross-rater composite was formed by taking the highest rating on each item by any version to create the "highest rating composite" (HRC) based on the recommendations by of Piacentini et al. (1992) for weighing information across raters when assessing childhood psychopathology, in general, and based on the recommendations of Frick and Hare (2001) for assessing CU traits specifically. The justification for this type of composite is that underreporting is more likely than overreporting on measures of psychopathology and CU traits. As a result, having one report that is lower than another may not necessarily reflect less of the trait and the higher score may actually be a more accurate indicator of the trait level in the person being assessed.

Conduct Problems. The Disruptive Behavior Disorders Scale (DBD; Pelham et al., 1992) is a 39-item measure of the symptoms included in the *DSM-5* diagnostic criteria for ADHD, oppositional defiant disorder (ODD) and CD (APA, 2012). Items were answered on a 4-point Likert-type scale from 1 (*not at all*) to 4 (*very much*). For the current study, only the items from the ODD and CD subscales were used and combined into a composite of CP. Ratings from the DBD have been significantly correlated with diagnoses of ODD and CD based on structured interviews in samples of children aged 7 to 12 years of age (Waschbusch et al., 2007) and has been associated with clinical diagnoses across various cultural groups (Gerdes et al., 2013). Teacher ratings on both the ODD and CD subscales have also shown strong reliability in a school aged sample ($\alpha = .95$ and .75, respectively; Pelham et al., 1992). In the current study, parent and teacher ratings were correlated $r = .26$ ($p < .001$). The ratings from these two informants were combined by taking the highest score from either informant for each item (Piacentini et al., 1992). This multirater composite showed a very high level of internal consistency in this sample ($\alpha = .92$).

Peer-Nominations. Given their ability to capture a unique perspective on youth's functioning, peer nominations are the most commonly used method of peer-referenced assessment (McMullen et al., 2014), and further, have been used without any evidence of negative effects on children. For example, Mayeux et al. (2007) completed a sociometric exercise with 91 third graders and then interviewed them and their teachers. Their results indicated that children were not hurt or upset by the procedures nor did the participants feel that their peers treated them differently following the testing. Importantly, peer rejection has been associated with aggression and CP in multiple studies (e.g., Hooijsma et al., 2020; Waller et al., 2017) and peer rejection assessed by peer nominations at age 5 has predicted being arrested, dropping out of school, and using illicit substances through age 27 (Lansford et al., 2016). In addition, peer nominations of social preference have been inversely correlated with self-report and informant measures of CU traits (Barry et al., 2008; Graziano et al., 2016) and peer nominations of

empathy have been negatively correlated with self-report of CU traits ($r = -.26$, $p < .01$; van Baardewijk et al., 2008).

In the current study, peer nominations were collected following the guidelines provided by Cillessen (2009). Specifically, for each possible nomination, participants could make an unlimited number of nominations and were allowed to nominate anyone within their grade at their school. In all grades, only the nominations of participating children were coded and used for analyses and self-nominations were omitted. For each item, the number of nominations received were summed, then standardized using the proportion score method in which the number of nominations received was divided by the number of nominators in that grade and school, with positively worded items being reverse-scored following standardization. A social preference score was determined by taking standardized nominations for the question "who do you like the least?" and subtracting them from the standardized nominations of "who do you like the most," consistent with past research (see McMullen et al., 2014). Thus, higher scores indicate greater "liked most" relative to "liked least" nominations (i.e., less peer rejection).

Other peer nominations were collected to more specifically assess characteristics that have been associated with CU traits. Again, a combination of both positive and negative peer nominations was used based on factor analysis (Matlasz et al., 2021). The *mean/cold* composite consisted of five nominations: "who is mean?" "who doesn't care who they hurt?" "who always has to get his or her own way?" "who doesn't care about having friends?" and "who is hard to get to know well?" ($\alpha = .77$). The second composite consisted of three reverse-scored items, "who is nice?" "who can you trust?" and "who is easy to make friends with?" which together, represent a pattern of someone who is potentially *not nice* ($\alpha = .86$).

Procedure

After receiving approval from the Louisiana State University Institutional Review Board, we obtained permission from the superintendent of the school system and the principals at the elementary and middle schools within this system. After receiving approval from the schools, we approached teachers in the third, sixth, and eighth grades, who sent a description of the study home with the children, along with parental consent forms and parent-report measures (i.e., ICU, DBD, and demographics). Upon receiving parental consent, children were asked for their assent to participate. Parents and children were fully informed of all study procedures, including the fact that the personality of the child would be rated by their classmates. All child-report measures were administered during the school day, in a group setting, on school computers. For all participating children (i.e., those who returned parental consent and provided assent), teachers were instructed to complete the ICU and the DBD. To compensate teachers for their time and effort, and to encourage participation, we offered the school $10.00 per participating child to go toward purchasing classroom supplies. Class (i.e., grade) sizes ranged from 73 to 134, with an average class size of 100. Across all four schools, classroom participation rates ranged from 42% to 54% of eligible students participating, with 54.0% participating across the third grades, 42.67% in the sixth grades, and 48.0% participating in the eighth grade, with a weighted average (i.e., weighted by classroom size) participation rate of 49%.

Data Analysis

All main analyses were conducted using SPSS Statistics v24. First, a 3 × 3 mixed analysis of variance with ICU version type (i.e., self, parent, and teacher) as the within-subjects factor and grade (i.e., third, sixth, and eighth) as the between-subjects factor was conducted to determine if there were main effects of rater or grade on ICU ratings, as well as interactions between grade and rater. Second, we tested the zero-order correlations between the various scores from the ICU (i.e., total scores from each rater and two cross-rater composites) and the various validators (i.e., parent- and teacher-rated CP and peer nominations). Third, we tested whether any of these correlations with the ICU were moderated by grade. To do this, a series of hierarchical regression analyses were run with the main effects for the ICU score and grade entered as predictors in the first step and their interaction added in the second step. Last, to examine if any of the versions were uniquely associated with the validators after controlling for the other versions, a series of simultaneous regression analyses were conducted with self, parent, and teacher reports of the ICU entered simultaneously as predictors of the four outcome variables.

Participants for the current study were drawn from a sample ($N = 289$) used in a larger study. Only those who had all three versions of the ICU were included in the current study, which resulted in the final sample size of 236. Given that the sample was collected for a different purpose, power analyses for our statistical tests are post-hoc. Thus, for each analysis we determined the minimum effect size that could be detected as significant with the sample size of 236 at a power of .95 for each of our analyses using G-Power. For the 3 × 3 mixed analysis of variance with version type as the within-subjects effect and child's grade as the between-subjects effect, our power analysis indicated that our sample size allowed us to detect a within-factor effect size (i.e., version) of $f^2 = .01$ and an interaction effect size (i.e., version by grade) of $f^2 = .02$ at the $p < .05$ level with a power of .95. Since no study has previously tested an interaction between rating format and grade (or age), it was hard to determine if this was sufficient power to detect

potential interactions. However, interaction effects smaller than this would likely not lead to meaningful differences in raters across grades. The next set of analyses were the correlations between the multiple versions of the ICU and the various validators. Our sample size provided a power of .95 (one-tailed) to detect a correlation of at least .21 as being significant at the $p < .05$ levels. A review of past research suggested the ICU's average correlation with CP across 105 studies was .33 (Frick et al., 2014b). Finally, we conducted multiple regression analyses with five predictors and our goal was to detect unique variance for each predictor. Our sample size provided a power of .95 to detect an effect size of $f^2 = .06$ ($\beta = .25$) at the $p < .05$ level. White et al. (2009) reported that self and parent reports on the ICU independently predicted a history of antisocial behavior in a sample adolescents who were arrested for sexual offenses, with β's of .28 and .33 for self- and parent-report, respectively.

Results

Preliminary Analyses

Bivariate correlations revealed that of the demographic variables measured, only marital status and gender were significantly correlated with the main study variables, with age, race/ethnicity, and parental education demonstrating no significant correlations with any of the ICU variables or validators. Specifically, marital status was correlated with CP ($r = -.19$, $p < .01$), indicating that children whose parents were married were rated as showing lower levels of CP. Gender was associated with several of the variables, suggesting that girls were more likely to both rate themselves as lower on CU traits ($r = -.15$, $p < .05$) and receive lower ratings of CU traits by teachers ($r = -.31$, $p < .001$); received less nominations for the Not Nice peer dimension ($r = -.22$, $p < .01$); and received higher ratings of social preference ($r = .16$, $p < .05$). Given that gender was the only variable associated with both the ICU and the validators, it was the only demographic variable used as a covariate in the main analyses.

Group Differences Between ICU Versions and Grade

To test the first study hypotheses, we conducted a 3 × 3 analysis of covariance with ICU version (self, teacher, and parent) as the within-subjects factor and grade (third, sixth, and eighth) as the between-subjects factor, controlling for gender. Mauchly's test indicated that the assumption of sphericity was violated ($\chi^2 = 18.93$, $p < .001$); therefore, the following statistics reported were corrected using the Huynh–Feldt estimates. The results of these analyses revealed no significant main effect of grade on level of CU traits. However there was a significant main effect of version on CU traits, $F(1.89, 439.08) = 22.64$, $p < .001$, partial $\eta^2 = .09$. Pairwise comparisons indicated that there were significant differences between self and parent reports of CU traits ($p < .01$), with self-report being higher than parent-report. Also, there were significant differences between self and teacher reports of CU traits ($p < .01$) and parent and teacher reports ($p < .001$), with teachers reporting the highest levels of CU traits.

Furthermore, there was a significant interaction between ICU version and grade, $F(3.79, 439.08) = 5.49$, $p < .001$, partial $\eta^2 = .05$, suggesting that the differences in youths' levels of CU traits reported by the different ICU versions varied between grades. Post hoc analyses were conducted examining all possible 2 × 3 interactions (i.e., child–parent, child–teacher, teacher–parent) with grade to determine which differences were leading to the significant interaction. The interactions emerged between self and teacher reports, $F(2, 232) = 7.52$, $p < .01$, partial $\eta^2 = .06$, and between teacher and parent reports, $F(2, 232) = 6.20$, $p < .01$, partial $\eta^2 = .05$. These interactions are graphed in Figure 1. As illustrated in this figure, teacher-report was higher than parent-report in third grade and was higher than both self- and parent-report in sixth grade. However, there were no differences between the level of CU traits reported across versions in Grade 8.

Differences in the ICU Versions' Associations With Conduct Problems and Peer Nominations

Zero-Order Correlations. The correlations between the different ICU versions and the multirater composites from the ICU and the variables used to test the validity of these indices of CU traits are provided in Table 1. As can be seen in Table 1, scores on the different versions of the ICU were significantly correlated with each other and with the HRC ICU score.[5] As can also be seen in Table 1, all of the scores from the ICU were significantly correlated with the various validators. We also tested for significant differences in the strength of the correlations between each of the three versions of the ICU with each validator using Fisher's *r*-to-*z* transformation based on the recommendations of Steiger (1980). These are also provided in Table 1. To summarize the results of these correlations, the correlation between HRC ICU and parent–teacher rated CP was quite substantial ($r = .56$, $p < .001$), with this largely accounted for by parent and teacher ratings. For the peer nominations, the HRC again tended to show the highest correlations, although this was generally due to the self-report and teacher-report versions.

Moderation by Grade. The results of the hierarchical multiple regression analyses to determine if there were interactions between each ICU version and grade level in the ICU's association with the four validators (controlling for gender)

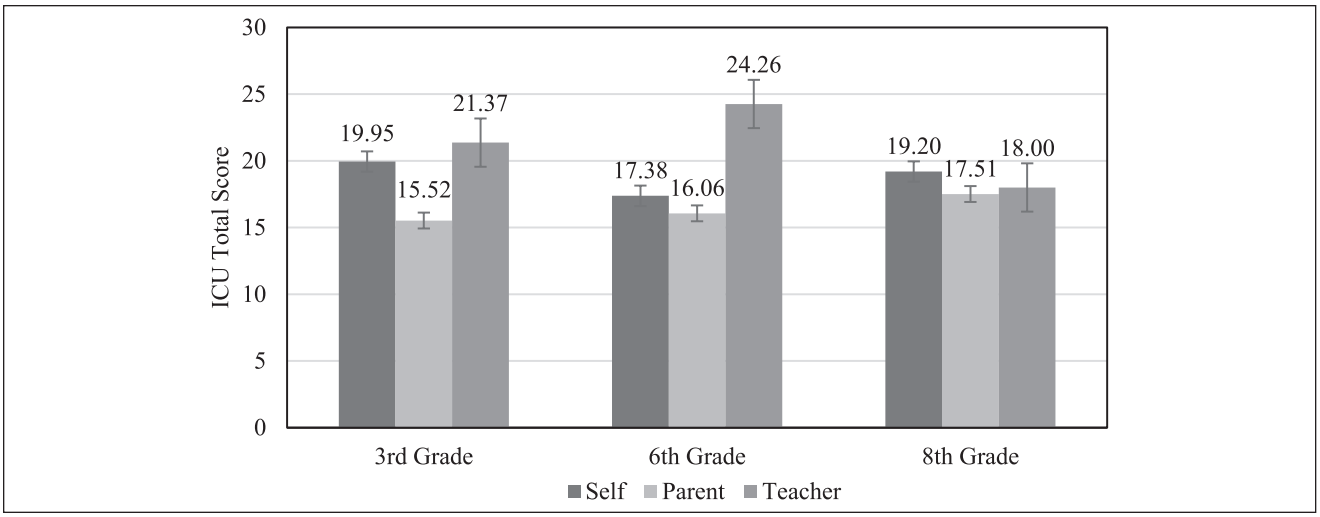

Figure 1. Estimated marginal means of levels of CU traits.
Note. ICU = Inventory of Callous–Unemotional Traits.

Table 2. Results of Hierarchical Regression Analyses With Teacher-Report ICU.

	CP		Social Preference		Mean/Cold		Not Nice	
Versions	*b* (*SE*)	β	*b* (*SE*)	β	*b* (*SE*)	β	*b* (*SE*)	β
Self-report								
Gender	0.15 (0.99)	0.01	0.34 (0.20)	0.11	−0.01 (0.10)	−0.01	−0.24 (0.12)	−0.13*
Grade	−0.07 (0.22)	−0.02	0.01 (0.04)	0.02	−0.001 (0.02)	−0.004	0.01 (0.03)	0.03
T-ICU	0.27 (0.04)	0.44***	−0.02 (0.01)	−0.16*	0.01 (0.004)	0.15*	0.02 (0.004)	0.28***
R^2	.19***		.04**		.01		11***	
Parent-report								
Gender	0.31 (0.95)	0.02	0.30 (0.19)	0.10	−0.002 (0.10)	−0.001	−0.23 (0.12)	−0.13
Grade	−0.09 (0.21)	−0.02	0.01 (0.04)	0.02	−0.002 (0.02)	−0.01	0.01 (0.03)	0.03
T-ICU	0.74 (0.11)	1.24***	−0.13 (0.02)	−1.15***	0.04 (0.01)	0.80**	0.05 (0.01)	0.69**
T-ICU × Grade	−0.07 (0.02)	−0.83***	0.02 (0.003)	1.03***	−0.01 (0.002)	−0.69**	−0.004 (0.002)	−0.42*
R^2	.25***		.13***		.05**		.12***	
ΔR^2	.06***		.10***		.04**		.02*	

Note. Gender was controlled for in all regression analyses reported above. We also repeated the tests of these interactions while also controlling for the interaction of gender by ICU and this did not influence the significant interactions with grade for any validator. β = standardized beta coefficient; *b* = unstandardized beta coefficient; *SE* = standard error; ICU = Inventory of Callous–Unemotional Traits; T-ICU = teacher-report ICU; ICU Grade = teacher-report ICU by grade interaction; CP = conduct problems.
*$p < .05$. **$p < .01$. ***$p < .001$.

revealed that there were no significant interactions of grade with self- and parent-report versions for any of the validators (ΔR^2 ranging from .000 to .01, all $ps > .05$). However, there were significant interactions with teacher-report for all four validators (ΔR^2 ranging from .02 to .10) and these are reported in Table 2.

To explore these interactions, the correlations between teacher-report and each validator were calculated separately by grade and are provided in Table 3. For ease of interpretation, gender was not controlled for in these analyses. Also, for comparison purposes, the correlations between the validators and the other ICU formats and the pairwise comparison showing the differences among versions are provided in Table 3. The reason for the interactions between teacher-report and grade is clear from these correlations. Teacher-reported CU traits had strong correlations with the four outcomes in Grades 3 and 6 but only one remained significant (i.e., with parent- and teacher-rated CP) at Grade 8. Also evident from this table is that at Grade 8, the self-report version of the ICU was the only report format to be correlated with peer nomination measures, the validators in which there was no shared method variance with the ICU ratings.

Table 3. Correlations Among Main Study Variables by Grade.

Grades	CP	Social Preference	Mean/Cold	Not Nice
Third grade (n = 71)				
Self	.31$_a$**	−.31$_a$**	.44***	.15$_a$
Parent	.44***	−.21$_a$	.30*	.09$_a$
Teacher	.65***	−.62***	.36**	.40**
HRC	.63***	−.52***	.35**	.35**
Sixth grade (n = 56)				
Self	.36**	−.30*	.08	.38**
Parent	.41**	−.31*	.15	.37**
Teacher	.61***	−.40**	.29*	.59***
HRC	.69***	−.48***	.28*	.64***
Eighth grade (n = 109)				
Self	.30**	−.27**	.35$_a$***	.25**
Parent	.44$_a$***	−.04	.15	.04
Teacher	.21*	.08	−.003	.15
HRC	.42***	−.12	.23*	.17

Note. Significant pairwise comparisons ($p < .05$) are denoted by the subscript "a," which differs significantly from teacher. ICU = Inventory of Callous–Unemotional Traits; Self = self-report ICU; Parent = parent-report ICU; Teacher = teacher-report ICU; HRC = highest rating composite ICU; CP = conduct problems.
*$p < .05$. **$p < .01$. ***$p < .001$.

Incremental Contributions of the Different Versions

In Table 4, all versions of the ICU were entered together in a simultaneous multiple regression model to test their incremental association with each of the validators, controlling for the other versions (and gender). Given the findings of differences in the correlation between teacher-report and the validators across grades, these analyses were conducted both in the full sample and for each grade separately. In the full sample, all three versions of the ICU contributed unique variance to the prediction of CP, whereas only the self-report version contributed unique variance to the prediction of all three of the peer nomination variables. However, these results in the full sample obscure differences across grade. In the third grade, teacher-report tended to show the most consistent association with the validators, explaining unique variance in all but one (i.e., mean/cold) of the peer nomination variables. By the eighth grade, however, teacher-report did not explain significant incremental variance in any of the outcome variables, even the one (i.e., CP) that also included teacher-report. In contrast, self-report was the only variable that explained unique variance in all four validators.

Discussion

The primary focus of the current study was to test and compare the utility of the different versions of the ICU across children in three grades prior to high school, using at least some validators that did not share the same rater used to complete the ICU. Such an investigation is quite timely and important, as the *DSM-5* criteria for the "with Limited Prosocial Emotions" specifier to a diagnosis of CD requires information from different sources across several settings. Unfortunately, very little work to date has systematically studied potential differences between the three reporter versions of the ICU across age or considered the role of shared method variance in the validity of the different versions of the ICU.

We first tested whether there were version or grade effects on the absolute levels of CU traits rated on the ICU. Our results revealed a grade by ICU version interaction that was largely due to differences in teacher ratings across grade. Specifically, analyses revealed that teacher ratings of CU traits were significantly higher than parent ratings on the ICU in third grade and significantly higher than both self and parent ratings in the sixth grade. By the eighth grade, there were no differences in the mean level of CU traits rated across the three versions. Given that one of the only other studies to compare versions across age did not find differences in teacher ratings cross age in a large sample (N = 955*)* of German children aged 6 to 18 years (Ueno et al., 2019), our results need to be replicated in other samples. However, the wider age range of this German sample, with a large number of children at or above our top age (where we did not find rater differences), and the absence of self-report prior to the age of 13 years in the German sample, all may account for the inconsistencies in findings. However, our findings do suggest that the same norms for CU traits cannot be used across versions, at least at younger ages, and they highlight potential changes in teacher ratings of CU traits across development.

Table 4. Incremental Validity of the Different Versions of the ICU.

	CP		Social Preference		Mean/Cold		Not Nice	
Versions	*b* (*SE*)	β	*b* (*SE*)	β	*b* (*SE*)	β	*b* (*SE*)	β
Overall								
Gender	0.46 (0.91)	0.03	0.28 (0.20)	0.10	0.02 (0.09)	0.02	−0.22 (0.12)	−0.12
Grade	−0.21 (0.20)	−0.06	0.02 (0.04)	0.02	−0.01 (0.02)	−0.02	0.01 (0.03)	0.03
Self	0.12 (0.06)	0.12*	−0.04 (0.01)	−0.23**	0.03 (0.01)	0.30***	0.02 (0.01)	0.14*
Parent	0.27 (.05)	0.33***	−0.01 (0.01)	−0.07	0.01 (0.01)	0.13*	0.004 (0.01)	0.05
Teacher	0.21 (0.04)	0.35***	−0.01 (0.01)	−0.09	0.002 (0.004)	0.04	0.02(0.01)	0.24**
R^2	.31***		.09***		.12***		.12***	
Third grade								
Gender	1.52 (1.38)	0.09	0.59 (0.27)	0.21*	−0.15 (0.18)	−0.09	0.27 (0.20)	0.16
Self	−0.05 (0.09)	−0.06	−0.01 (0.02)	−0.04	0.03 (0.01)	0.32**	−0.01 (0.01)	−0.08
Parent	0.34 (0.07)	0.38***	−0.02 (0.02)	−0.12	0.02 (0.01)	0.22*	0.01 (0.01)	0.07
Teacher	0.46 (0.07)	0.66***	−0.07 (0.01)	−0.53***	0.01 (0.01)	0.16	0.04 (0.01)	0.47**
R^2	.54***		.40***		.23***		.14**	
Sixth grade								
Gender	0.07 (2.32)	0.004	0.24 (0.41)	0.09	0.13 (0.20)	0.11	−0.20 (0.25)	−0.11
Self	0.20 (0.13)	0.18	−0.03 (0.02)	−0.16	0.002 (0.01)	0.02	0.02 (0.01)	0.20
Parent	0.12 (0.10)	0.15	−0.02 (.02)	−0.12	0.003 (0.01)	0.05	0.01 (0.01)	0.08
Teacher	0.36 (0.09)	0.51***	−0.03 (.02)	−0.26	0.02 (0.01)	0.33	0.03 (0.01)	0.45**
R^2	.39***		.16*		.02		.37***	
Eighth grade								
Gender	0.67 (1.28)	0.05	−0.09 (0.32)	−0.03	0.09 (0.13)	0.07	−0.40 (0.18)	−0.21*
Self	0.18 (0.08)	0.20*	−0.06 (0.02)	−0.31**	0.03 (0.01)	0.37***	0.03 (0.01)	0.21*
Parent	0.29 (0.07)	0.38***	0.001 (0.02)	0.01	0.01 (0.01)	0.09	0.001 (0.01)	0.01
Teacher	0.06 (0.05)	0.11	0.02 (0.01)	0.15	−0.004 (0.01)	−0.09	0.003 (0.01)	0.04
R^2	.23***		.06*		.11**		.08*	

Note. Gender was controlled for in all regression analyses reported above. ICU = Inventory of Callous–Unemotional Traits; β = standardized beta coefficient; *b* = unstandardized beta coefficient; *SE* = standard error; CP = conduct problems.
*$p < .05$. **$p < .01$. ***$p < .001$.

In terms of the overall sample, the self-report version of the ICU seemed to show the most consistent correlations with the measures chosen as validators for this study across the three grade levels. As seen in Table 1, this version showed the most consistent correlations across both the measure of CP and the various peer nominations, being related to peer rejection and to being considered "mean" and "not nice" by peers. In fact, the self-report version of the ICU often did not differ greatly in the strength of associations with the validators from the multirater composite (i.e., HRC) that combined self-, parent-, and teacher-report on the ICU. Such findings support the practice in much of the past research that used the ICU by relying on the self-report version (Cardinale & Marsh, 2020; Deng et al., 2019; Frick et al., 2014b).

However, these overall findings mask some important differences in validity across the ICU versions depending on the child's grade. Specifically, we found an interaction between ICU version type and grade in the associations between the ICU and the various validators, which was largely due to changes in the validity of teacher ratings across the grades studied (see Table 2). That is, teacher ratings showed some of the best validity, both overall (Table 3) and incremental to the other raters (Table 4) in Grade 3 but showed some of the poorest validity by Grade 8. This change is consistent with research on the assessment of other forms of child psychopathology, suggesting that the validity of teacher-report declines as a child leaves elementary school (van der Ende & Verhulst, 2005; Verhulst & van der Ende, 1992). This decline in the validity of teacher-report likely is due to teachers spending less time with children in later grades, when children are more likely to change classes (and thus, change teachers) during the day (Frick et al., 2020). Interestingly, with some notable exceptions, the self-report of the ICU showed fairly consistent levels of validity across the grades studied and clearly emerged as the most valid rater by Grade 8. It is possible that this finding is due to the fact that as children's cognitive abilities develop, they are better able to identify and report on their feelings and internal states, leading to greater consistency

between their self-reported CU traits and characteristics or behaviors observed by others as they get older (De Los Reyes & Kazdin, 2005).

All of these findings should be interpreted in the context of several study limitations. First, although we were able to compare ICU ratings at three different age groups, we did not obtain data for youth below third grade or above eighth grade. Thus, our ability to generalize beyond this age range is questionable. This limitation is particularly important for our findings supporting the validity of the self-report version across the various ages, given that self-report on many measures of child psychopathology decreases below the age of 9 years (Frick et al., 2020). Second, our findings are limited by the cross-sectional design of the study in which all measures were collected at a single time point; thus, we were not able to make any inferences regarding the predictive validity of the various versions for the ICU. As such, the results of the current study need to be replicated using a longitudinal design to examine potential differences in the prediction of outcomes across the various versions. Third, although the sample was fairly diverse in terms of race and ethnicity, family structure, and socio economic status, participants were recruited from public schools in a rural area of the southern United States, which may limit the generalizability of the findings to other samples, especially those in more urban regions. Fourth, the sample size was much smaller for the sixth grade compared with the third and eighth grades, and across all grades, the classroom participation rates for the peer nominations were fairly low (e.g., 42% to 54%). However, past work has suggested that participation rates as low as 40% demonstrated good reliability when assessing overt aggression and popularity, specifically (Marks et al., 2013). Furthermore, Prinstein (2007) demonstrated that nominations collected from a small subsample of youth (i.e., 10% of the grade) correlated moderately to highly with nominations collected from the full sample of youth. Another issue related to our sample size is the ability to detect interactions, which typically are much smaller than main effects in social science research (Blake & Gangestad, 2020; McClelland & Judd, 1993). However, for the most important interactions for the study aims, the nonsignificant interactions between self-report and grade and parent-report and grade all contributed less than 0.01% of additional variance over the main effects in their association with validators. Thus, we are confident that such small effects would not have resulted in important differences in the validity of these raters across grade, even if they became significant with a larger sample. Finally, we were limited in the number and types of validators that were used to test the differential validity of the various version of the ICU. As noted previously, one reason that CU traits is an important construct relative to other related constructs is not its correlation with CP but the fact that it moderates the relation of CP with theoretically important variables, such as emotional reactivity to distress cues or responsiveness to cues of punishment (Frick et al., 2014b). Thus, it would be important to further test the validity of the different versions of the ICU using other variables that have proven to be important to the construct of CU traits.

Despite these limitations, the findings have important implications for the use of the ICU. First, research has often recommended combining different reporters for measuring child psychopathology, given that each method is capturing something unique that might not be captured by any single reporter alone (Achenbach et al., 1987; De Los Reyes & Kazdin, 2005). Our results support this practice to some extent, given that combining across ICU versions to form a composite score from the ICU generally led to the strongest and most consistent correlations across the different measures used as validators and across the grade levels studied. However, it is also important to note that these correlations were not always substantially better than what was found for certain single raters. As a result, for certain purposes (e.g., some research studies), ratings from single raters may provide fairly similar information to what would be obtained from combining across different reporters.

Second, what appeared to be the best version or best combination of versions seemed to vary across grade. In the third grade, the findings for the self-report and teacher-report versions were very similar. In contrast, by eighth grade the self-report format was clearly superior to the other versions and the validity of the teacher-report was questionable. As a result, if a single reporter is desired across this developmental age range, it would seem best to use the self-report format. If a second reporter is desired, the validity of teacher-report is clearly best in third grade and still seems to add important information to self-report in the sixth grade. By the eighth grade, however, other versions did not add to the prediction of peer nominations of social preference or nominations of being mean and cold (Table 4). Furthermore, adding additional reporters to the self-report on the ICU decreased the correlations with these validators that did not share methods (Table 3). Thus, these findings support the widely used practice of relying on self-report on the ICU in adolescent samples (Cardinale & Marsh, 2020; Deng et al., 2019). However, they also point to the need to further explore other versions, especially teacher-report prior to adolescence, a method that has heretofore been largely underutilized in past research (Deng et al., 2019).

Declaration of Conflicting Interest

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) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Tatiana M. Matlasz https://orcid.org/0000-0003-2111-0340

Notes

1. To illustrate the overlap between dimensions of psychopathy and criteria for childhood disruptive behavior disorders, a recent study by Colins et al. (2020) administered a commonly used measure of psychopathic traits in children and adolescents and reported that the correlation between impulsive lifestyle dimension of psychopathy with a measure of ADHD symptoms was $r = .84$ ($p < .001$). Furthermore, five of the nine items used to assess the interpersonal dimension of psychopathy were related to lying and conning other people (i.e., "Lies often to avoid problems," "seems to lie more than other children of the same age," often lies to get what he/she wants," "to frequently lie seems to be completely normal for him/her," "to get people to do what he/she wants, he/she often finds it efficient to con them"). Lying and conning is part of definition of CD.
2. It is important to note that the review of Wilkinson et al. (2016) was not consistent with the contention that those children and adolescents high on CP and CU traits were unresponsive to treatment. Instead, the results suggest that those with elevated CU traits tended to enter treatment with more severe CP and, despite improving with treatment, still had more severe CP at posttreatment.
3. The *DSM-5* and ICD-11 definitions for the Limited Prosocial Emotions all share the same four symptoms: lack of remorse or guilt, a lack of empathy, shallow or deficient emotions, and a failure to care about performance in important activities. However, the ICD-11 definition also includes a failure to respond to punishment, which is not assessed by the ICU.
4. Some have questioned whether the unemotional items on the ICU should be included in the total score, given that they are less correlated with the other items and are less correlated (and sometimes uncorrelated) with measures of CP and aggression (Hawes et al., 2014). However, as noted in the discussion of the unique nature of CU traits, relative to other dimensions of psychopathy, the importance of the construct is that it predicts important correlates independent of CP not in its ability to predict CP. Thus, the fact that these items load on the overarching factor in factor analyses (Ray & Frick, 2018) and are negatively correlated with empathy but less correlated with CP (Cardinale & Marsh, 2020), make them important for measuring the construct.
5. We also explored two other ways to combine information across different versions to see if multirater composites led to scores from the ICU that were more highly associated with key validators than scores from any version of the ICU. In addition to the HRC, we also created a composite by summing scores across the three raters and by taking a resolved score of the two best versions, which in this case, were teacher and self. The correlations between these three composites were quite high: HRC and summed composite, $r = .92$; HRC and self-teacher resolved, $r = .91$, and summed composite and self-teacher resolved, $r = .87$; all ps $< .001$. As would be expected from these very high correlations, the results were unchanged irrespective of which composite was used.

References

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101*(2), 213-232. https://doi.org/10.1037/0033-2909.101.2.213

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Author.

Barhight, L. R., Hubbard, J. A., Swift, L. E., & Konold, T. R. (2017). A multitrait-multimethod approach to assessing childhood aggression and related constructs. *Merrill-Palmer Quarterly*, *63*(3), 367-395. https://doi.org/10.13110/merrpalmquar1982.63.3.0367

Barry, T. D., Barry, C. T., Deming, A. M., & Lochman, J. E. (2008). Stability of psychopathic characteristics in childhood: The influence of social relationships. *Criminal Justice and Behavior*, *35*(2), 244-262. https://doi.org/10.1177/0093854807310508

Benesch, C., Görtz-Dorten, A., Breuer, D., & Döpfner, M. (2014). Assessment of callous-unemotional traits in 6 to 12 year-old children with callous-unemotional traits by parent ratings. *Journal of Psychopathology and Behavioral Assessment*, *36*(4), 519-529. https://doi.org/10.1007/s10862-014-9420-7

Berg, J. M., Lilienfeld, S. O., Reddy, S. D., Latzman, R. D., Roose, A., Craighead, L. W., Pace, T. W. W., & Raison, C. L. (2013). The inventory of callous and unemotional traits: A construct-validational analysis in an at-risk sample. *Assessment*, *20*(5), 532-544. https://doi.org/10.1177/1073191112474338

Blair, R. J. R., Leibenluft, E., & Pine, D. S. (2014). Conduct disorder and callous–unemotional traits in youth. *New England Journal of Medicine*, *371*(23), 2207-2216. https://doi.org/10.1056/nejmra1315612

Blake, K. R., & Gangestad, S. (2020). On attenuated interactions, measurement error, and statistical power: Guidelines for social and personality psychologists. *Personality and Social Psychology Bulletin*, *46*(12), 1702-1702. https://doi.org/10.1177/0146167220913363

Burns, G. L. (2000). Problem of item overlap between the Psychopathy Screening Device and attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder rating scales. *Psychological Assessment*, *12*(4), 447-450. https://doi.org/10.1037/1040-3590.12.4.447

Cardinale, E. M., & Marsh, A. A. (2020). The reliability and validity of the Inventory of Callous Unemotional Traits: A meta-analytic review. *Assessment*, *27*(1), 57-71. https://doi.org/10.1177/1073191117747392

Cillessen, A. H. N. (2009). Sociometric methods. In K. H. Rubin, W. M. Bukowski, & B. Laursen (Eds.), *Social, emotional, and personality development in context: Handbook of peer interactions, relationships, and groups* (pp. 82-99). Guilford Press.

Colins, O., Andershed, H., Salekin, R. T., & Fanti, K. (2018). Comparing different approaches for subtyping children with conduct problems: Callous-unemotional traits only versus the multidimensional psychopathy construct. *Journal of Psychopathology and Behavioral Assessment*, *40*(1), 6-15. https://doi.org/10.1007/s10862-018-9653-y

Colins, O. F., Fanti, K. A., & Andershed, H. (2020). The *DSM-5* limited prosocial emotions specifier for conduct disorder:

comorbid problems, prognosis, and antecedents. *Journal of the American Academy of Child and Adolescent Psychiatry*, *60*(8), 1020-1029. https://doi.org/10.1016/j.jaac.2020.09.022

De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, *131*(4), 483-509. https://doi.org/10.1037/0033-2909.131.4.483

Deng, J., Wang, M. C., Zhang, X., Shou, Y., Gao, Y., & Luo, J. (2019). The Inventory of Callous Unemotional Traits: A reliability generalization meta-analysis. *Psychological Assessment*, *31*(6), 765-780. https://doi-org.libezp.lib.lsu.edu/10.1037/pas0000698.supp

Docherty, M., Boxer, P., Huesmann, L. R., O'Brien, M., & Bushman, B. (2017). Assessing callous-unemotional traits in adolescents: Determining cutoff scores for the inventory of callous and unemotional traits. *Journal of Clinical Psychology*, *73*(3), 257-278. https://doi.org/10.1002/jclp.22313

Edens, J. F., Skopp, N. A., & Cahill, M. A. (2008). Psychopathic features moderate the relationship between harsh and inconsistent parental discipline and adolescent antisocial behavior. *Journal of Clinical Child and Adolescent Psychology*, *37*(2), 472-476. https://doi.org/10.1080/15374410801955938

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. https://doi.org/10.1080/15374416.2012.734221

Frick, P.J., Barry, C.T., & Kamphaus, R.W. (2020). *Clinical assessment of children's personality and behavior* (4th ed.). Springer.

Frick, P. J., Bodin, S. D., & Barry, C. T. (2000). Psychopathic traits and conduct problems in community and clinic-referred samples of children: Further development of the Psychopathy Screening Device. *Psychological Assessment*, *12*(4), 382-393. https://doi.org/10.1037/1040-3590.12.4.382

Frick, P.J., Lahey, B.B., Loeber, R., Tannenbaum, L.E., Van Horn, Y., Christ, M.A.G., Hart, E. A., & Hansen, K. (1993). Oppositional defiant disorder and conduct disorder: A meta-analytic review of factor analyses and cross-validation in a clinic sample. *Clinical Psychology Review 13*(4), 319-340. https://doi.org/10.1016/0272-7358(93)90016-F

Frick, P. J., & Ray, J. V. (2015). Evaluating callous-unemotional traits as a personality construct. *Journal of Personality*, *83*(6), 710-722. https://doi.org/10.1111/jopy.12114

Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014a). A developmental psychopathology approach to understanding callous-unemotional traits in children and adolescents with serious conduct problems. *Journal of Child Psychology and Psychiatry*, *55*(6), 532-548. https://doi.org/10.1111/jcpp.12152

Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014b). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychological Bulletin*, *140*(1), 1-57. https://doi.org/10.1111/jcpp.12152

Gao, Y., & Zhang, W. (2016). Confirmatory factor analyses of self- and parent-report inventory of callous-unemotional traits in 8- to 10-year-olds. *Journal of Psychopathology and Behavioral Assessment*, *38*(3), 331-340. https://doi.org/10.1007/s10862-015-9527-5

Gerdes, A. C., Lawton, K. E., Haack, L. M., & Hurtado, G. D. (2013). Assessing ADHD in Latino families: Evidence for moving beyond symptomatology. *Journal of Attention Disorders*, *17*(2), 128-140. https://doi.org/10.1177/1087054711427396

Graziano, P. A., Ros, R., Haas, S., Hart, K., Slavec, J., Waschbusch, D., & Garcia, A. (2016). Assessing callous-unemotional traits in preschool children with disruptive behavior problems using peer reports. *Journal of Clinical Child & Adolescent Psychology*, *45*(2), 201-214. https://doi.org/10.1080/15374416.2014.971460

Haas, S. M., Becker, S. P., Epstein, J. N., & Frick, P. J. (2018). Callous-unemotional traits are uniquely associated with poorer peer functioning in school-aged children. *Journal of Abnormal Child Psychology*, *46*(4), 781-793. https://doi.org/10.1007/s10802-017-0330-5

Hare, R. D., & Neumann, C. S. (2008). Psychopathy as a clinical and empirical construct. *Annual Review of Clinical Psychology*, *4*(1), 217-246. https://doi.org/10.1146/annurev.clinpsy.3.022806.091452

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. https://doi.org/10.1037/a0034718

Henry, J., Dionne, G., Viding, E., Vitaro, F., Brendgen, M., Tremblay, R. E., & Boivin, M. (2018). Early warm-rewarding parenting moderates the genetic contributions to callous–unemotional traits in childhood. *Journal of Child Psychology and Psychiatry*, *59*(12), 1282-1288. https://doi.org/10.1111/jcpp.12918

Hooijsma, M., Huitsig, G., Dijkstra, J. K., Flache, A., & Veenstra, R. (2020). Being friends with or rejected by classmates: Aggression toward same- and cross-ethnic peers. *Journal of Youth and Adolescence*, *49*(3), 678-692. https://doi.org/10.1007/s10964-019-01173-1

Kimonis, E. R., Fanti, K. A., Anastassiou-Hadjicharalambous, X., Mertan, B., Goulter, N., & Katsimicha, E. (2016). Can callous-unemotional traits be reliably measured in preschoolers? *Journal of Abnormal Child Psychology*, *44*(4), 625-638. https://doi.org/10.1007/s10802-015-0075-y

Kimonis, E. R., Fanti, K. A., & Singh, J. P. (2014). Using longitudinal methods to establishing cut-off scores for the parent-reported inventory of callous-unemotional traits. *Archives of Forensic Psychology*, *1*(1), 27-48. https://doi.org/10.1037/e500122015-043

Kimonis, E. R., Frick, P. J., Fazekas, H., & Loney, B. R. (2006). Psychopathy, aggression, and the emotional processing of emotional stimuli in non-referred girls and boys. *Behavioral Sciences and the Law*, *24*(1), 21-37. https://doi.org/10.1002/bsl.668

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., Aucoin, K. J., & Morris, A. S. (2008). Assessing callous–unemotional traits in adolescent offenders: Validation of the Inventory of Callous–Unemotional Traits. *International Journal of Law and Psychiatry*, *31*(3), 241-252. https://doi.org/10.1016/j.ijlp.2008.04.002

Kliem, S., Lohmann, A., Neumann, M., Glaubitz, C., Haselbach, S., Bergmann, M. C., & Baier, D. (2020). Factor structure of the inventory of callous-unemotional traits in a representative sample of German 9th grade students. *Journal of Abnormal Child Psychology*, *48*(1), 43-55. https://doi.org/10.1007/s10802-019-00590-x

Kotler, J. S., & McMahon, R. J. (2005). Child psychopathy: Theories, measurement, and relations with the development and persistence of conduct problems. *Clinical Child and Family Psychology Review*, *8*(4), 291-325. https://doi.org/10.1007/s10567-005-8810-5

Koutsogiorgi, C. C., Lordos, A., Fanti, K. A., & Michaelides, M. P. (2020). Factorial structure and nomological network of the Inventory of Callous-Unemotional Traits accounting for item keying variance. *Journal of Personality Assessment*, *103*(3), 312-323. https://doi.org/10.1080/00223891.2020.1769112

Lansford, J. E., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2016). A public health perspective on school dropout and adult outcomes: A prospective study of risk and protective factors from age 5 to 27 years. *Journal of Adolescent Health*, *58*(6), 652-658. https://doi.org/10.1016/j.jadohealth.2016.01.014

Lin, B., Kerig, P. K., & Adkins, D. E. (2019). Cross-validation of the Inventory of Callous-Unemotional Traits across reporters and genders in a sample of detained youth. *Psychological Assessment*, *31*(1), 15. https://doi.org/10.1037/pas0000636

Marks, P. E., Babcock, B., Cillessen, A. H., & Crick, N. R. (2013). The effects of participation rate on the internal reliability of peer nomination measures. *Social Development*, *22*(3), 609-622. https://doi.org/10.1111/j.1467-9507.2012.00661.x

Matlasz, T. M., Frick, P. J., & Clark, J. E. (2021). Understanding the social relationships of youth with callous-unemotional traits using peer nominations. *Journal of Clinical Child and Adolescent Psychology*. Advanced online publication. https://doi.org/10.1080/15374416.2020.1823847

Mayeux, L., Underwood, M. K., & Risser, S. D. (2007). Perspectives on the ethics of sociometric research with children. *Merrill-Palmer Quarterly*, *53*(1), 53-78. https://doi.org/10.1353/mpq.2007.0002

McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*(2), 376-390. https://doi.org/10.1037/0033-2909.114.2.376

McDonald, S. E., Ma, L., Green, K. E., Hitti, S. A., Cody, A. M., Donovan, C., Williams, J. H., & Ascione, F. R. (2018). Evaluation of the parent-report inventory of callous–unemotional traits in a sample of children recruited from intimate partner violence services: A multidimensional Rasch analysis. *Journal of Clinical Psychology*, *74*(3), 418-441. https://doi.org/10.1002/jclp.22497

McMullen, J. A., Veermans, K., & Laine, K. (2014). Tools for the classroom? An examination of existing sociometric methods for teacher use. *Scandinavian Journal of Educational Research*, *58*(5), 624-638. https://doi.org/10.1080/00313831.2013.838694

Muñoz, L. C., Kerr, M., & Besic, N. (2008). The peer relationships of youths with psychopathic personality traits: A matter of perspective. *Criminal Justice and Behavior*, *35*(2), 212-227. https://doi.org/10.1177/0093854807310159

Pelham, W. E., Jr., Gnagy, E. M., Greenslade, K. E., & Milich, R. (1992). Teacher ratings of *DSM-III-R* symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *31*(2), 210-218. https://doi.org/10.1097/00004583-199203000-00006

Piacentini, J. C., Cohen, P., & Cohen, J. (1992). Combining discrepant diagnostic information from multiple sources: Are complex algorithms better than simple ones? *Journal of Abnormal Child Psychology*, *20*(1), 51-63. https://doi.org/10.1007/bf00927116

Piatgorsky, A., & Hinshaw, S. P. (2004). Psychopathic traits in boys with and without attention-deficit/hyperactivity disorder: Concurrent and longitudinal correlates. *Journal of Abnormal Child Psychology*, *32*(5), 535-550. https://doi.org/10.1023/b:jacp.0000037782.28482.6b

Prinstein, M. J. (2007). Assessment of adolescents' preference- and reputation-based peer status using sociometric experts. *Merrill-Palmer Quarterly*, *53*(2), 243-261. https://doi.org/10.1353/mpq.2007.0013

Ray, J. V., & Frick, P. J. (2018). Assessing callous-unemotional traits using the total score from the Inventory of Callous-Unemotional Traits: A meta-analysis. *Journal of Clinical Child and Adolescent Psychology*, *49*(2), 190-199. https://doi.org/10.1080/15374416.2018.1504297

Ray, J. V., Frick, P. J., Thornton, L. C., Steinberg, L., & Cauffman, E. (2016). Positive and negative item wording and its influence on the assessment of callous-unemotional traits. *Psychological Assessment*, *28*(4), 394-404. https://doi.org/10.1037/pas0000183

Roose, A., Bijttebier, P., Decoene, S., Claes, L., & Frick, P. J. (2010). Assessing the affective features of psychopathy in adolescence: A further validation of the inventory of callous and unemotional traits. *Assessment*, *17*(1), 44-57. https://doi.org/10.1177/1073191109344153

Salekin, R.T. (2017). Research review: What do we know about psychopathic traits in children? *Journal of Child Psychology and Psychiatry*, *58*(11), 1180-1200. https://doi.org/10.1111/jcpp.12738

Sharp, C., & Kine, S. (2008). The assessment of juvenile psychopathy: Strengths and weaknesses of currently used questionnaire measures. *Child and Adolescent Mental Health*, *13*(2), 85-95. https://doi.org/10.1111/j.1475-3588.2008.00483.x

Sica, C., Ciucci, E., Baroncelli, A., Frick, P. J., & Patrick, C. J. (2020). Not just for adults: Using the Triarchic Model of Psychopathy to inform developmental models of conduct problems in adolescence. *Journal of Clinical Child and Adolescent Psychology*, *49*(6), 897-911. https://doi.org/10.1080/15374416.2019.1574228

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87*(2), 245-251. https://doi.org/10.1037//0033-2909.87.2.245

Ueno, K., Ackermann, K., Freitag, C. M., & Schwenck, C. (2019). Assessing callous–unemotional traits in 6-to 18-year-olds: Reliability, validity, factor structure, and norms of the German version of the Inventory of Callous–Unemotional Traits. *Assessment*, *28*(2), 567-584. https://doi.org/10.1177/1073191119847766

van Baardewijk, Y., Stegge, H., Andershed, H., Thomaes, S., Scholte, E., & Vermeiren, R. (2008). Measuring psychopathic traits in children through self-report: The development of the Youth Psychopathic Traits Inventory–Child Version. *International Journal of Law and Psychiatry*, *31*(3), 199-209. https://doi.org/10.1016/j.ijlp.2008.04.004

van der Ende, J., & Verhulst, F. C. (2005). Informant, gender and age differences in ratings of adolescent problem behaviour. *European Child & Adolescent Psychiatry*, *14*(3), 117-126. https://doi.org/10.1007/s00787-005-0438-y

Verhulst, F. C., & van der Ende, J. (1992). Agreement between parents' reports and adolescents' self-reports of problem behavior. *Journal of Child Psychology and Psychiatry*, *33*(6), 1011-1023. https://doi.org/10.1111/j.1469-7610.1992.tb00922.x

Viding, E., Sebastian, C. L., Dadds, M. R., Lockwood, P. L., Cecil, C. A., De Brito, S. A., & McCrory, E. J. (2012). Amygdala response to preattentive masked fear in children with conduct problems: The role of callous-unemotional traits. *American Journal of Psychiatry*, *169*(10), 1109-1116. https://doi.org/10.1176/appi.ajp.2012.12020191

Waller, R., Hyde, L.W., Baskin-Sommers, A.R., & Olson, S. L. (2017). Interactions between callous unemotional behaviors and executive function in early childhood predict later aggression and lower peer-liking in late-childhood. *Journal of Abnormal Child Psychology*, *45*(3), 597-609. https://doi.org/10.1007/s10802-016-0184-2

Waller, R., Wright, A. G., Shaw, D. S., Gardner, F., Dishion, T. J., Wilson, M. N., & Hyde, L. W. (2015). Factor structure and construct validity of the parent-reported Inventory of Callous-Unemotional Traits among high-risk 9-year-olds. *Assessment*, *22*(5), 561-580. https://doi.org/10.1177/1073191114556101

Waschbusch, D. A. (2002). A meta-analytic examination of comorbid hyperactive-impulsive-attention problems and conduct problems. *Psychological Bulletin*, *128*(1), 118-150. https://doi.org/10.1037/0033-2909.128.1.118

Waschbusch, D. A., Walsh, T. M., Andrade, B. F., King, S., & Carrey, N. J. (2007). Social problem solving, conduct problems, and callous-unemotional traits in children. *Child Psychiatry and Human Development*, *37*(4), 293-305. https://doi.org/10.1007/s10578-006-003-6

Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLOS ONE*, *11*(3), Article e0152719. https://doi.org/10.1371/journal.pone.0152719

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

Wilkinson, S., Waller, R., & Viding, E. (2016). Practitioner review: Involving young people with callous unemotional traits in treatment—does it work? A systematic review. *Journal of Child Psychology and Psychiatry*, *57*(5), 552-565. https://doi.org/10.1111/jcpp.12494

Willoughby, M. T., Mills-Koonce, W. R., Waschbusch, D. A., Gottfredson, N. C., & the Family Life Project Investigators. (2014). An examination of the parent report version of the Inventory of Callous Unemotional Traits in a community sample of 1st grade children. *Assessment*, *22*(1), 76-85. https://doi.org/10.1177/1073191114534886

World Health Organization. (2018). *The ICD-11 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Author.