Published in final edited form as:

J Clin Psychol. 2018 March ; 74(3): 418–441. doi:10.1002/jclp.22497.

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

Shelby Elaine McDonald, Ph.D.,

School of Social Work, Virginia Commonwealth University, Richmond, VA

Lin Ma, Ph.D.,

Department of Research Methods and Statistics, Morgridge College of Education, University of Denver, Denver, CO

Kathy E. Green, Ph.D.,

Department of Research Methods and Statistics, Morgridge College of Education, University of Denver, Denver, CO

Stephanie A. Hitti, B.S.,

Department of Psychology, Virginia Commonwealth University, Richmond, VA

Anna M. Cody, M.S.,

School of Social Work, Virginia Commonwealth University, Richmond, VA

Courtney Donovan, Ph.D.,

School of Education & Human Development, University of Colorado, Denver

James Herbert Williams, Ph.D., and

Graduate School of Social Work, University of Denver, Denver, CO

Frank R. Ascione, Ph.D.

Graduate School of Social Work, University of Denver, Denver, CO

Abstract

Objective—Our study applied multidimensional item response theory (MIRT) to compare structural models of the Inventory of Callous and Unemotional Traits Parent-Report Version (ICU; English and North American Spanish translations).

Method—Two hundred ninety-one maternal caregivers were recruited from community-based domestic violence services and reported on their children (77.9% ethnic minority; 47% female), who ranged in age from 7 to 12 years ($M=9.07$, $SD=1.64$). We compared nine models that were based on prior psychometric evaluations of the ICU.

Results—MIRT analyses indicated that a revised 18-item version comprised of two factors (Callous-Unemotional and Empathic-Prosocial) was more suitable for our sample. Differential item functioning was found for several items across ethnic and language groups, but not for child gender or age. Evidence of construct validity was found.

Conclusions—We recommend continued research and revisions to the ICU to better assess the presence of CU traits in community samples of school-age children.

Keywords

CU traits; domestic violence; children; Latinos; MIRT

Callous-unemotional (CU) traits in children and adolescents refer to a general lack of remorse, shallow affect, callousness, insensitivity to others' feelings, and/or hyper-regulated emotions. These traits, which are used as a clinical specifier for the course and treatment of conduct disorder (CD) in the *Diagnostic and Statistical Manual of Mental Disorders* (5th edition; *DSM-5*; American Psychiatric Association, 2013), are considered the early affective features of psychopathy in youth (Frick & White, 2008; Shenk, Dorn, Kolko, Rausch, & Insana, 2014). Research documents that these types of traits are stable over 1-, 2-, and 5-year periods among children between the ages of 7 to 17 (i.e., Frick, Kimonis, Dandreaux, & Farrell, 2003; Lynam et al., 2009; Lynam, Derefinko, Caspi, Loeber & Stouthamer-Loeber, 2007) and predictive of psychopathy in adulthood, even when controlling for childhood risk factors and CD (Burke, Loeber, & Lahey, 2007; Lynam, Caspi, Moffitt, Loeber, & Stouthamer-Loeber, 2007).

There is some evidence to suggest that exposure to adverse rearing environments, such as homes where child maltreatment occurs, increases children's risk of developing CU traits. It is hypothesized that early socialization plays an important role in the development of CU traits; in particular, early abnormal affective experiences, such as child maltreatment, may interfere with the development of morality and appropriate emotional responses to distress cues (Blair, 1995; Cleckley, 1982; Kochanska, 1993). As such, child abuse has been linked to lack of concern over the distress of others and empathy (Clark & Frick, 2016; Klimes-Dougan & Kistner, 1990). Moreover, recent research suggests that CU traits operate as an important mechanism underlying the link between child maltreatment and later risk behaviors (Carlson, Oshri, & Kwon, 2015).

Direct and indirect exposure of children to intimate partner violence (IPV) is increasingly being recognized as a form of childhood maltreatment in the U.S (Gilbert et al., 2009; MacMillan, Wathen, & Varcoe, 2013). Further, studies consistently show that approximately 45% of children exposed to IPV are also subject to direct physical aggression by a parent (Slep & O'Leary, 2005). Accordingly, there is some evidence to suggest that exposure to IPV increases children's risk for developing CU traits and is predictive of greater severity of these traits throughout young adulthood (Shenk et al., 2014; Weiler & Widom, 1996). A recent study found that among children exposed to IPV, CU traits, when examined alongside children's social problems, attention problems, internalizing and externalizing problems, and empathy, were important in distinguishing between children with asymptomatic, moderately maladjusted, and severely maladjusted profiles of adjustment (McDonald, Graham-

Bermann, Maternick, Ascione, & Williams, 2016). Moreover, prior studies indicate that psychopathic traits are a risk factor for chronic perpetration of IPV (Babcock, Green, Webb, & Yerington, 2005; Coyne, Nelson, Graham-Kevan, Keister, & Grant, 2010; Swogger, Walsh, & Kosson, 2007). For example, a study by Swogger and colleagues (2007) found that low empathy, callousness, and low impulsivity characterized men charged with domestic battery compared to perpetrators of different violent crimes.

Considering the substantial genetic etiology of CU traits (Henry, Pingault, Boivin, Rijdsdijk, & Viding, 2016), along with their association with childhood maladjustment, adult psychopathy, and the intergenerational transmission of IPV, it is important that future research and practice attend to these traits when examining children's exposure and response to family violence. Identification of CU traits in childhood is critical given that interventions targeted at reducing antisocial behavior are most effective and least costly when implemented early in childhood (Frick, 2016). Children may be less capable than caregivers to serve as informants of their own disruptive behavior problems (Loeber, Green, Lahey, & Stouthamer-Loeber, 1991); therefore, it is important that a reliable, succinct, and easily administered caregiver-report assessment be available to use in clinical practice and research with children exposed to IPV. The current study evaluated the utility of the parent-report version of the Inventory of Callous-Unemotional Traits (ICU; Frick, 2004) as a measure of CU traits among school-age children exposed to IPV.

Inventory of Callous-Unemotional Traits

A variety of instruments have been used to measure the presence of CU traits, such as the Antisocial Process Screening Device (APSD; Frick & Hare, 2001), the Psychopathy Checklist–Youth Version (Forth, Kosson, & Hare, 2003), and the Child Psychopathy Scale (Lynam, 1997). The ICU (Frick, 2004) is a measure that has garnered increased attention among studies of early psychopathy in youth during the past decade. The ICU was developed in response to the limitations of the Antisocial Process Screening Device (APSD; Frick & Hare, 2001), which only included four items that consistently loaded on a CU factor (i.e., “Does not show emotions to others,” “Concerned about the feelings of others,” “Cares about how well he/she does at schoolwork,” “Feels bad or guilty when he/she does something wrong”). Frick (2004) developed three negatively-worded and three positively-worded items to reflect behaviors consistent with each of the four APSD items, resulting in the 24-item ICU measure, which was conceptualized as a unidimensional model of CU traits. Notably, this measure can be downloaded free of charge and is available in the following five versions: Youth Self-Report, Parent Report, Teacher Report, Parent Report (Preschool Version), and Teacher Report (Preschool Version). In addition, the original English language version of the measure has been translated into nearly 20 other languages.

ICU factor structure

Several psychometric evaluations of the ICU have posited a bifactor model with three subfactors among various age ranges, translations, and clinical and community settings (Kimonis et al., 2008; Ciucci, Baroncelli, Franchi, Golmaryami, & Frick, 2014). The three intercorrelated factors, which make up the unique contributions to the overarching CU

construct have been labeled as follows: Callous, Uncaring, and Unemotional. The Callous factor captures behaviors demonstrating a lack of guilt, remorse, and empathy. Items loading on the Uncaring factor reflect a lack of caring about others' feelings as well as one's performance on tasks. The Unemotional factor reflects items pertaining to a lack of emotional expression (Essau, Sasagawa, & Frick, 2006; Gao & Zhang, 2016). The psychometric properties of the *youth* self-report version have been rigorously evaluated, and the majority of studies have found support for a factorial model, some with an overarching CU dimension and three subfactor structure (Henry et al., 2016) and others for a two-factor model with no overarching structure (Benesch, Görtz-Dorten, Breuer, & Döpfner, 2014; Houghton, Hunter, & Crow, 2013; Gao & Zhang, 2016). Moreover, many studies have reported evidence supporting construct validity of the measure. Specifically, the callousness and uncaring factors are consistently related to conduct problems, aggression, and antisocial behavior across studies (Essau et al., 2006; Fanti, Frick, & Georgiou, 2009). The uncaring factor is also consistently associated with empathy deficits (Hawes et al., 2014). Evidence of construct validity for the unemotional factor has been established with empathy and sensation seeking measures, yet this factor has not been consistently related to other behavioral correlates (Kimonis et al., 2016).

A few recent studies have conducted confirmatory factor analyses and reported on the factor structure of the English language version of the *parent*-report ICU (i.e., Gao & Zhang, 2016; Hawes et al., 2014; Henry et al., 2016; Horan, Brown, Jones & Aber, 2015; Roose, Bijttebier, Decoene, Claes, & Frick, 2010; Willoughby, Mills-Koonce, Waschbusch, Gottfredson, & Family Life Project Investigators, 2015). Among studies examining the ICU structure in community samples of adolescents, Roose et al. (2010), Henry et al. (2016), and Horan et al. (2015) found that a bifactor model with Callous, Unemotional, and Uncaring factors had the best relative fit of the models considered in their studies; similarly, Gao and Zhang (2016) found support for this model in a community sample of 8 to 10 year-old children. In contrast, Willoughby et al. (2015) and Hawes et al.'s (2014) findings did not lend support for the utility of the overall score of the parent report version. Hawes et al. (2014) found support for a two-factor (Callousness and Uncaring) structure in a sample of 250 boys between the ages of 6 and 12 years with only 12 of the original 24 items retained. Willoughby et al. (2015) found a two-factor model consistent with Hawes et al.'s (2014) results among parents of 1,078 first grade children; however, given that all of the items on Hawes et al.'s (2014) revised "uncaring" scale were comprised of items reflecting the *presence* of empathic and prosocial behavior (prior to reverse-scoring), these authors termed the factors "callous-unemotional" and "empathic-prosocial." Interestingly, Waller et al. (2015) recently reported support for the three-subfactor bifactor framework (general-specific model) and a two-factor model using a shortened ICU in a sample of parents of high-risk 9 year olds. However, the authors reported that the two-factor solution (comprised of callous and uncaring behaviors) could offer a more stable and parsimonious assessment of CU traits. Thus, there is need for further examination of the ICU.

Horan et al. (2015) argued that IRT techniques should be considered in future studies that seek to further examine the validity of the ICU. To our knowledge, Hawes et al. (2014) is the only study to date to have applied IRT analysis to parent-reported ICU data. However, this study examined boys exhibiting conduct problems and the sample was predominantly

African American (72%). Little is known about the dimensionality, item difficulty, response category use, and measurement invariance of the parent-report version of the ICU among community samples of school-age children. Moreover, no prior studies have examined differential item functioning across North American Spanish and English translations of the measure. Item functioning is a critical component for validity and it is essential that the ICU instrument include items showing a range of difficulty levels when administered across clinical and community settings. Thus, additional research attending to measure structure, item difficulty, and differential item functioning is needed to further investigate the psychometric properties of the parent report version of the ICU when administered in ethnically diverse, gender-balanced samples of parents of school-age children.

Current Study

Given prior research demonstrating great heterogeneity of psychosocial functioning among children exposed to IPV and the utility of CU traits in distinguishing children at high risk for severe maladjustment and persistent antisocial behavior, evaluation of the ICU's psychometric properties in a US sample of parents recruited from community-based IPV services is an important way to advance knowledge on the measure as well as assess the appropriateness of the ICU as a clinical assessment tool in this population. Moreover, we are unaware of any research that has tested multiple competing latent factor structures of the parent-report version of the ICU using multidimensional item response theory (MIRT); several studies have used classical test theory with a bifactor model, but use of item response theory to assess measure structure and item functioning is rare. Therefore, to address multiple gaps in the literature, our study had three primary aims. The first aim of our study was to replicate and extend knowledge of the psychometric properties of the ICU using MIRT in order to develop a more detailed understanding of the measure. We tested unidimensional, 2-factor (Empathic/Prosocial and Callous/Unemotional) and 3-factor (Callous, Unemotional, Uncaring) models with the 24-item ICU. We then deleted items that had been found to function poorly in prior research and tested 1-, 2-, and 3-factor structural models; next, we identified items that functioned poorly for our data and compared fit of the 1-, 2-, and 3-factor models. Expanding on the findings of Hawes et al. (2014) and Willoughby et al. (2015), we anticipated that a shortened, 2-factor model characterized by empathic-prosocial (EP) and CU factors would provide the best fit to our data. Our second aim was to examine invariance of the measure across demographic variables (Latina/non-Latina, racial majority/minority, gender, age) and whether the scale functioned similarly for caregivers completing the measure in English and Spanish. Finally, as noted by Ray, Frick, Thornton, Steinberg, and Cauffman (2016) and Waller et al. (2015), the ICU subfactors have not demonstrated consistent associations with external criteria across studies. Therefore, our final aim was to replicate and extend previous studies' efforts to explore convergent validity of the parent report version of the ICU. Consistent with previous studies such as Gao and Zhang (2016), Hawes et al. (2014) and Willoughby et al. (2015), this study examined the convergent validity of the ICU using correlations between the ICU subscales and standardized symptom scores pertaining to the following behavioral domains: aggression, internalizing behavior problems, externalizing behavior problems, total behavior problems, oppositional defiant (OD) problems, and conduct problems. In addition, we expanded on

prior work in this area by examining the validity of the ICU with maternal reports of children's empathy. We then tested the statistical significance of differences among correlations. We anticipated positive correlations between children's ICU scores and scores on the behavioral measures, and negative correlations between the ICU and empathy scores. Due to significant gender and ethnicity differences in ICU subscale scores, we then controlled demographic variables to determine the incremental prediction of ICU scores.

Method

This study involved a secondary analysis of existing data collected as part of a larger, mixed methods study examining the treatment of pets in families impacted by IPV. Two hundred ninety-one maternal caregiver-child dyads were recruited from 22 IPV agencies in a western U.S state between 2010 and 2016. To participate in the study, women were required to be over the age of 21 and have: (1) experienced IPV within the past year, (2) at least one child between the ages of 7–12 in the home, (3) lived with at least one companion animal in the home within the past year. If the participant had more than one eligible child between the ages of 7 and 12, they were able to select which child participated in the study. Designated domestic violence advocates at each agency were trained to recruit participants who were not in immediate crisis, obtain voluntary consent and assent, and administer standardized surveys to eligible women and children. Survey administrators scheduled interview appointments with participants and caregivers completed the surveys in a confidential location at the agency where they were recruited. Then, the staff member met with the child independently to administer child survey materials. Data analyzed in the current study reflect maternal caregiver data from the first data collection point. The Institutional Review Board at an institution of higher education approved these procedures. Further details on sample recruitment and participant characteristics can be found in the work of (AUTHOR CITATION).

Participants

Maternal caregivers ranged in age from 21 to 65 years with a mean age of 36.34 ($SD=7.43$). The majority of women identified as Latina/Hispanic (60.5%) with the remaining portion of the sample identifying as non-Latina White (26.8%), multi-racial/ethnic (6.60%), African American or Black (3.4%), American Indian or Alaska Native (1.7%), Pacific Islander (0.3%), and Asian (0.3%). The majority of children were Latino/Hispanic (55.3%). The ICU Parent Report Form—English Version was used with 66.3% of participants and the Spanish Version with 33.7% of participants. A summary of participant characteristics can be found in Table 1.

Measures

Spanish language versions of the measures were used when possible. For surveys without established Spanish translations, the English language version was translated to Spanish using the translation–back-translation procedure by a professional translator (Van de Vijver & Hambleton, 1996). In addition, trained survey administrators who were bilingual in Spanish and English provided feedback on the surveys following translation to confirm the appropriateness of the measure for Spanish-speaking study participants.

The Inventory of Callous and Unemotional Traits—Parent Report Form (ICU; Frick, 2004)—Child CU traits were assessed via maternal caregiver report on the 24-item parent-report version of the ICU. Items are rated on a four-point scale (ranging from 0 = not at all true to 3 = definitely true). The ICU is comprised of 12 positively worded and 12 negatively worded items. Reliability coefficients for the three factors proposed by Essau and colleagues (2006) were .85 for each factor. Willoughby et al. (2015) found that ICU items best fit a 2-factor model comprised of Callous-Unemotional (CU) and Empathic Prosocial (EP) subfactors. Although the authors did not report the reliability coefficients for the subscales, their results indicated significant latent variance was explained by the two subscales ($CU = \varphi^2 = .63$; $EP = .68$, $p < .001$). The North American Spanish version of the measure, which was translated and approved by the scale developer and a bilingual research assistant, was completed by 33.7% of our sample; the rest of participants completed the English language version. With the exception of one participant who identified as multi-racial, all participants who completed the Spanish translation of the measure identified as Latina. We are unaware of any published psychometric evaluation of the North American Spanish Language version of the ICU. In the current study, Cronbach's alpha for the English and Spanish language versions of the total scale were .70 and .81, respectively.

Child Behavior Checklist (CBCL 6/18; Achenbach & Rescorla, 2001)—Maternal caregivers completed the English or Spanish language version of the Child Behavior Checklist (CBCL 6/18; Achenbach & Rescorla, 2001). The CBCL is a widely used standardized measure of caregivers' reports of children's behavior problems and competencies. The measure allows for the calculation of raw and *T* scores that reflect multiple domains of socioemotional functioning and competence. Total scores and their respective *T* scores were strongly correlated in the current sample. Specifically, the following *T* scores, which are normed for age and gender were used in the current study: Aggressive Behavior (18 items; $\alpha = .91$), Internalizing Behavior Problems (32 items; $\alpha = .90$), Externalizing Behavior Problems (35 items; $\alpha = .92$), Total Behavior Problems (84 items; $\alpha = .94$), Oppositional Defiant Problems (5 items; $\alpha = .83$), and Conduct Problems (17 items; $\alpha = .86$). Standardized scoring procedures for the Oppositional Defiant and Conduct Problems scales were derived based on expert ratings of their similarity with various diagnoses on the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed., text rev.; *DSM-IV-TR*; American Psychological Association, 2000).

Griffith Empathy Measure (GEM; Dadds et al., 2008a)—The GEM is a parent-report survey that measures children's cognitive and affective empathy. The instrument is comprised of 23 items scored from strongly disagree (−4) to strongly agree (+4). Previous research has demonstrated the utility of the total score as a unidimensional construct capturing an overall dimension of empathy among children ages 4 to 18 years (Dadds et al., 2008a). Dadds et al. (2008b) reported adequate internal consistencies, test-retest reliability of scores over 1 week ($r > 0.89$) and 6 month periods ($r > 0.69$), as well as a stable factor structure across gender and age. The GEM total score and affective and cognitive empathy scales have been used in prior research to support the construct validity of the ICU (e.g., Kimonis et al., 2016). In the current sample, reliability of the total score and cognitive and affective scales were adequate ($\alpha = .84$, $.75$, and $.89$, respectively). The GEM was not

available in a Spanish language version. Thus, this measure was adapted using the translation–back-translation procedure (Van de Vijver & Hambleton, 1996).

Analysis Procedures

Multidimensional item response models—Multidimensional item response models are an extension of item response theory (IRT). Specifically, this study employed an extension of the Rasch model, which assumes that observed responses on a measure reflect one overarching major trait that accounts for item variance over and above a set of domain-specific latent constructs. ConQuest software (Adams, Wu, & Wilson, 2015) was utilized for the current study, using a multidimensional random coefficients multinomial logit model (Adams, Wilson, & Wang, 1997). Relative model fit was evaluated using the Akaike information criterion (AIC: Akaike, 1974): $AIC = d + 2p$. According to Kang and Cohen (2007), d refers to deviance and p refers to number of estimated parameters. AIC difference was calculated using a likelihood ratio chi-squared statistic (Myers, Wolfe, & Feltz, 2005). Because a chi-squared statistic is affected by sample size, this study also calculated the AIC proportionality constant ($AIC\text{-}pc = AIC/(\text{degrees of freedom})$). Lower AIC and AIC- pc values indicate better model fit.

Analyses—Before any statistical analyses were performed, ICU items reflecting prosocial behavior (1, 3, 5, 8, 13, 14, 15, 16, 17, 19, 23, and 24) were reverse-scored as recommended by ICU scoring procedures (ICU; Frick 2004). Then we compared multiple structural models using ConQuest software (Adams, Wu, & Wilson, 2015; Table 2). Specifically, we generated unidimensional, and 2-, and 3-factor models with all 24 items; we then generated the same models with the 20 items found to function well in past research. We then identified items that did not fit well with our data and generated models for an 18-item measure. The 20-item ICU excluded items 2, 10, 15, and 23, which were found to be misfitting or problematic in prior studies (e.g., Gao & Zhang, 2015; Hawes et al., 2014; Henry et al., 2016). In the current study, we used the criterion established by Linacre (2016) to identify misfitting items; by this standard the mean square fit statistic is considered out of range if not between 0.5 and 1.5 logits and if the corresponding t statistic has an absolute value that exceeds 2.0. The six items excluded had an absolute t -value and infit mean square fit over 2.0. Thus, consistent with Willoughby et al. (2015) and Hawes et al.'s (2014) methods, misfit was identified via poor contribution to the model.

The three subscales of the 3-factor ICU were termed Callous, Uncaring, and Unemotional. The 2-factor ICU models were based on the structure identified by Willoughby et al. (2015) and Waller et al. (2015). For the present study, a partial credit model (item+item*step), which allowed the response scale structure to vary from one item to the next, produced better model fit than a rating scale model (item+step), which constrained the step structure to be the same for all items. Thus, all analyses were based on a partial credit model.

Dimensionality: Unidimensionality is a primary assumption of the Rasch model and was checked using principle components analysis of residuals for each of the subscales identified as providing the best fitting model. As suggested by Linacre (2016), if the measure explains 40% or more of the total raw variance, with the first contrast (conceptual similar to a second

factor) having an eigenvalue of 3.0 or less with less than 5% variance due to the first contrast, the item set can be considered relatively unidimensional.

Item and person reliability: Rasch analysis provides an estimate of consistency (i.e., person reliability). The criterion for acceptable reliability was set at .70, as recommended by Nunnally and Bernstein (1994). Person separation was also examined to explore the ability of items to identify levels of the measure across persons on a less-to-more continuum (Bond & Fox, 2007). A separation of 2.0 is considered minimal with higher levels of separation indicating a wider range of items and persons (Linacre, 2015).

Scale use: Category probability curves were inspected in conjunction with step structure and Rasch-Andrich thresholds to assess the response category structure and determine if participants in our study were using the rating scale as intended. The Rasch-Andrich threshold represents the intersection between categories where there is a 50% chance of a participant being scored in either a higher or lower response category. Bond and Fox (2007) note that response category thresholds are inadequate if thresholds do not progress monotonically; moreover, Rasch-Andrich thresholds should progress by between 1.4 and 5.0 logits (Linacre, 2015). When violated, this means there might be significant conceptual or empirical overlap between response categories and the rating scale can potentially be improved by collapsing response categories.

Targeting and construct coverage: As previously noted, the Rasch model allows persons and items to be displayed simultaneously on a common logit scale. Thus, Rasch analysis allows researchers to see how items act on a continuum of ‘easiest’ to agree with to ‘hardest’ to agree with, thus creating a ruler of construct coverage; endorsability (difficulty) of items was assessed using item-person maps produced by the ConQuest software. The items at the top of the right side of the item-person map are questions that were hardest for participants to endorse, while those at the bottom were the easiest to endorse. In the item-person map (Figure 1) participants represented near the top of the left side of the item-person map have higher levels of the underlying trait (e.g., callous-unemotional) whereas participants represented near the bottom scored lower on the trait.

Differential item functioning: The Rasch model assumes that subgroups have equal probabilities of item response given trait level; thus, subgroups in the sample (e.g., Latina vs. non-Latina) should respond in a similar manner to items measuring the construct (Bond & Fox, 2007). Differential item functioning (DIF) was examined to ensure that the items were functioning in the same way across demographic groups. In the current study, we examined invariance of item function across sub-groups for the best fitting model using the software, Winsteps (Linacre, 2016) to test the multiple groups (age), as well as pairwise groups (child’s gender—male vs. female, Latina vs. non-Latina White, race/ethnicity majority (White non-Latina) vs. minority (all other racial/ethnic groups in this study), and language version—English vs. Spanish). For the multiple age groups, the study compared the item difficulty of each group to the average item difficulty of all groups. For the pairwise groups, the study compared the item difficulties of the first group to the second group.

Convergent validity—Convergent validity refers to how the ICU correlates with a known measure of a similar latent construct or diverges from a similar, but distinctly different latent construct (DeVellis, 2003). This was examined by calculating Pearson correlations between scores on the ICU subscales with the GEM total and subscale scores and the following CBCL *T* scores: Aggressive Behavior, Internalizing Behavior, Externalizing Behavior, Total Behavior Problems, OD Problems, and Conduct Problems. A series of comparisons of correlations between the scales were conducted using Eid, Gollwitzer and Schmidt's (2011) calculations via the Psychometrica online calculator (Lenhard & Lenhard, 2014).

Results

ICU Structure

Table 3 presents the fit statistics for model comparisons. Comparing the 24-item ICU, the 20-item ICU, and the 18-item ICU, the partial credit model indicated that the 18-item ICU provided the best fit to our data. The standardized infit mean square *t* statistics of the six removed items were over 2.0, indicating poor item fit. The AIC values of the 18-item ICU models were significantly lower ($p = .001$) than those of the 24-item ICU and the 20-item ICU. The AIC-pc values for the three models of the 18-item ICU also were the smallest. Comparing the three factor models for each ICU length, the 3-factor model fit our data better than the 1- and 2-factor models based on AIC values for the 24- and 20-item lengths; however, for the 18-item ICU, the 2-factor model fit best. With respect to the AIC-pc, the model fit of the three 2-factor models were the best for all three ICU lengths. Thus, we concluded that Model 8, the 18-item 2-factor model (without items 2, 3, 10, 15, 16, and 17), achieved the best fit. Consequently, this model was used for the remaining analyses.

Dimensionality, Item and Person Fit, Scale Use, and Targeting and Construct Coverage

Dimensionality was checked using a principle components analysis of residuals with Winsteps (Linacre, 2016). For the Callous-Unemotional (CU) factor, variance attributable to the measure was low at 29%, but with no evidence of a second dimension (first contrast eigenvalue of 1.48). Item logit positions ranged from $-.74$ to $.87$ with mean square infit values of $.81$ to 1.20 (mean square outfit of $.74$ to 1.64). Model person separation was $.99$ with reliability of person separation of $.49$, which was low, though with a Cronbach's alpha of $.78$. Reliability of person separation is computed excluding extreme scores, which typically results in a lower value than reliability computed using Cronbach's alpha. Reliability of person separation is also affected by the mismatch, or poor targeting, of items for the sample; this is shown in Figure 1 (map of person and item estimates for Model 8, the 2-factor ICU), as represented by the persons falling below where the items are located on the logit scale. Specifically, model item separation was 3.43 , indicating the ICU displayed a replicable item difficulty hierarchy. However, items were not well targeted for the sample with a person mean of -2.84 logit (model error = 1.12); in particular, items reflecting lower values of the trait are needed to improve construct coverage (Figure 1). The raw response distribution for all items was positively skewed, with infrequent use of higher response categories. Item thresholds varied by item with threshold inversions found for items 6, 18, and 21, indicating that for seven of ten items the response scale was used appropriately. A threshold inversion occurs when the logit position for a higher level category is lower than

the logit position for a lower level category (e.g., when it is “easier” to respond with ‘strongly agree’ than with ‘moderately agree’). Infrequent use of higher categories may have led to threshold inversions for these items. If adjacent response categories were collapsed, fit issues with item 18 would be resolved.

For the Empathic-Prosocial (EP) factor, variance attributable to the measure was adequate at 48%, with little evidence of a second dimension (first contrast eigenvalue of 2.19). Item logit positions ranged from -1.47 to $.60$ with mean square infit of $.80$ to 1.17 (mean square outfit of $.78$ to 1.12). Model person separation was 2.14, which was acceptable, with a reliability of person separation of $.82$ and Cronbach’s alpha of $.83$. Model item separation was 6.73, indicating the ICU displayed a replicable item difficulty hierarchy. Item thresholds varied by item, but no threshold inversions were found, indicating that the response scale was used appropriately for all items. As can be seen in Figure 1, items were fairly well-targeted for the sample on this factor with a person mean of -1.02 logit (model error = 0.61), though items could be added to reflect lower values of the trait. Generally, the Callous-Unemotional factor was relatively easier to agree with than the Empathic-Prosocial factor (Figure 1).

Differential Item Functioning

DIF analyses examined invariance of item function across sub-groups for the 2-factor, 18-item ICU. These analyses used Winsteps (Linacre, 2016) in order to test differences in logit item position across multiple groups for age as well as for gender, Latina vs. non-Latina White, ethnic/racial majority vs. minority, and language (English vs. Spanish). For the multiple age groups (ages 7, 8, 9, 10, 11, & 12), the study compared the item difficulty of each group to the average item difficulty of all groups and also compared item difficulty for younger (ages 7–9) and older (10–12) children. Further, DIF was assessed for measure presentation language (English/Spanish) for Latinas. For the pairwise groups, the study compared the item difficulties to each other. Substantial DIF was indicated by *both* statistical significance at $p < .01$, as well as an absolute value of the logit difference equal to or exceeding $.50$ (Smith, Rush, Fallowfield, Velikova, & Sharpe, 2009). Table 4 reports items with substantial DIF by group.

For the six age groups and two gender groups, no substantial DIF was found. When age was categorized, no substantial DIF was found for either scale nor was there DIF for language of measure presentation for Latinas. Seven items (1, 5, 7, 8, 12, 14, and 19) showed DIF across ethnic or language groups. Items 5, 8, and 12 were easier for Latinas to agree with while items 1, 4, 7, and 19 were easier for non-Latina White participants to agree with. Items 1, 4, 7, and 19 were easier to agree with when presented in English while items 5 and 12 were easier to agree with for respondents to the Spanish version of the ICU. Results for ethnic/racial majority vs. minority comparisons mirrored those for Latina/non-Latina White (significant DIF was found for items 5, 7, 12, 14). These results suggest there may be differences associated with language and culture that are reflected in a potential differential interpretation of items. It should also be noted that the comparisons across ethnic/racial categories, English/Spanish version of the ICU, and Latina/non-Latina White were not independent; that is, these demographic variables were associated.

Convergent Validity

Prior to computing Pearson correlations and linear regression analyses, the distribution of the ICU subscales were examined, as were the statistical significance of differences in subscale scores based on child gender and ethnicity (Latina vs. not). Analyses were performed using the ICU score (two subscales) logit person position, the CBCL *T* scores, and the GEM scores. The distribution of each ICU scale was reasonably normal (skewness = .225 and -.37 for CU and EP scales); the raw score item means ranged from 0.00 – 1.50 for CU and from 0.00 to 2.88 for EP.

Person logit position was used as the dependent variable in the analysis of effects of child gender and ethnicity (Latina vs. not) on the Callous-Unemotional scale using analysis of variance. The assumption of normality was met but homogeneity of variance was violated. The main effect of ethnicity was statistically significant, $F(1, 286) = 35.13, p < .001, \eta^2 = .11$, while the main effect of gender was not significant, $F(1, 286) = .08, p = .78, \eta^2 < .001$, nor was the interaction, $F(1, 286) = .83, p = .36, \eta^2 = .003$.

When the Empathic-Prosocial person logit position was the dependent variable, all ANOVA assumptions were met. The main effect of child gender was statistically significant, $F(1, 286) = 5.75, p = .017, \eta^2 = .02$, while the main effect of ethnicity was not significant, $F(1, 286) = .10, p = .76, \eta^2 < .001$, nor was the interaction, $F(1, 286) = .26, p = .61, \eta^2 = .001$. Table 5 provides the logit mean, standard deviation, and *n* per cell for the models with EP and CU scores as the dependent variable.

Variable correlations were computed and are provided in Table 6. According to Eid et al.'s (2011) comparison of correlations test, the correlations between the ICU CU scale and six CBCL *T* scores ($r = .46 \sim .63$) were statistically significantly higher than those between the ICU EP scale and the CBCL *T* scores ($r = .13 \sim .33$), $p < .001$ (Lenhard & Lenhard, 2014).

The correlation patterns between the two ICU subscales (CU and EP) and the two GEM subscales (Affective Empathy and Cognitive Empathy) differed. Specifically, the correlation between the CU subscale and GEM Affective subscale was positive and significant ($r = .25, p < .001$) whereas no statistically significant correlation was found between the EP subscale and GEM Affective Empathy score ($r = -.04, p > .05$). Meanwhile, both the CU and EP subscales were negatively associated with the Cognitive Empathy subscale ($r = -.45$ and $-.25$, respectively, $p < .001$); however, the correlation for the EP subscale was significantly lower than that between the CU subscale and GEM Cognitive Empathy subscale according to Eid et al.'s (2011) comparison of correlations test, $p < .01$ (Lenhard & Lenhard, 2014).

However, as there were significant effects of gender and ethnicity on ICU scores, a series of hierarchical regression analyses were conducted to control for those effects to determine the incremental predictive validity of ICU score. Six CBCL *T*-scores and GEM-Cognitive, GEM-Affective, and GEM-Total scores were the dependent variables. All regression analyses resulted in base (gender and ethnicity) and incremental models (adding either CU or EP) that were statistically significant at $p < .001$. Table 7 provides the *R*, *R*-square, incremental *R*-Square, and model standardized coefficients (Betas) with *p* values by dependent variable. CU score was a significant predictor at $p = .001$ for all dependent

variables except GEM-Affective and GEM-Total; EP was significant for all except GEM-Affective. The incremental R^2 for CU when it was significant ranged from .14 to .37 while for EP incremental R^2 ranged from .13 to .21. Both CU and EP were significantly and substantially predictive beyond effects of gender and ethnicity for all CBCL T -scores and less so for GEM scores.

Discussion

This study used MIRT to examine the structure and convergent validity of the parent-report ICU (Frick, 2004), a measure of children's and adolescents' callous and unemotional traits, in an ethnically diverse sample of 291 mother child-dyads recruited from community based IPV services. To our knowledge, this is the first study to report on dimensionality, item difficulty, response category use, and measurement invariance across Spanish and English translations of the ICU using MIRT. In consideration of recent work demonstrating that a 2-factor model of the parent-report version offers a more parsimonious model of CU traits, we were particularly interested in comparing unidimensional, 2-factor (EP and CU) and 3-factor (Callous, Unemotional, Uncaring) models. We used a partial credit model to examine and compare nine structural models, including unidimensional, 2-factor, and 3-factor models of 24-, 20-, and 18-item versions of the ICU.

ICU Structure

Our results indicated that the 18-item version (excluding items 2, 3, 10, 15, 16, and 17) with 2 subscales (EP and CU) provided the best fit to the observed ICU data based on the item responses of children's maternal caregivers. Our findings are consistent with recent work (Hawes et al., 2014; Waller et al., 2015; Willoughby et al., 2015) reporting that 2-factor models, comprised of CU and EP scale scores, best fit parent-report ICU data. Notably, the EP scale was comprised solely of positively worded items, whereas the CU scale was mostly comprised of negatively worded items. Thus, consistent with Ray et al.'s assertions (2016), the factor structure of the ICU could be an artifact emerging from different item properties (Henry et al., 2016; Ray et al., 2016) and suggests that future revisions to the ICU should aim for and test a better balance of negatively and positively worded items within each construct.

Our findings support Willoughby et al.'s (2015) assertions that an important distinction of the EP factor is that low ratings on some ICU items (i.e., EP items prior to reverse-scoring procedures) reflect the *absence* of prosocial and empathic behaviors rather than the *presence* of uncaring/callous behaviors. While this distinction seems semantic, our findings reinforce the authors' argument that these are correlated factors ($r = .45$, $p < .001$, in our study), but distinct dimensions of behavior. In other words, the *absence* of prosocial and empathic behaviors and the *presence* of uncaring behaviors represent related but distinct constructs. Thus, the absence of EP behaviors, while associated with CU behaviors, should not be considered indicative of CU traits in youth. This is of particular importance given recent changes to the DSM's diagnostic criteria for CD, which includes a limited prosocial emotions specifier for CD "that treats low EP and high CU behaviors as conceptually equivalent" (Willoughby et al., 2015, p. 83).

As CU traits seem to demarcate an important subgroup of antisocial youth who differ from other antisocial youth in terms of emotional, genetic, and social factors, by including the limited prosocial emotions specifier for CD, the most recent DSM attempted to integrate CU traits in the diagnostic criteria for CD (Kahn, Frick, Golmaryami, & Marsee, 2016). However, this specifier is comprised of both EP (i.e., lacks empathy) and CU (i.e., lacks guilt/remorse) behaviors. In light of the current findings (as well as those from Willoughby et al., 2015; Hawes et al., 2014; and Houghton et al., 2013) which suggest the incremental value of CU *and* EP behaviors in relation to behavioral outcomes, we support Willoughby et al.'s (2015) assertion that because the DSM specifier does not explicitly attend to the distinction between these constructs, the recent revisions to the DSM have not sufficiently integrated CU traits into the diagnostic criteria of CD.

This has several implications, including the fact that, as Frick and Nigg (2012) argued, youth with both CD and elevated levels of CU traits “exhibit a particularly severe and persistent form of the disorder” which should be captured by the diagnostic criteria of CD (p. 21). By not adequately including CU traits in the diagnostic criteria of CD, the predictive utility of CU traits is not being fully taken advantage of. Further, there is evidence suggesting that youth with and without CU traits have differential developmental pathways to their diagnosis of CD (Frick & White, 2008). The diagnostic criteria should, therefore, make distinctions between these developmental pathways to improve clinical practice and research. Finally, research suggests that antisocial youth with and without CU traits show differential responses in regards to treatment and psychological interventions (Frick & Dickens, 2006). In sum, as proposed by Frick and Nigg (2012), due to the fact that CU traits add to the clinical and predictive utility of CD, it becomes imperative that the DSM adequately integrates CU traits into the diagnostic criteria of CD.

Measure Invariance and Scale Function

CU scale—The Callous-Unemotional Scale showed no DIF on child's age or gender, and minimal DIF based on ethnicity (Latina/non-Latina white), majority/minority status, or language version. Further, no DIF was found for Latinas responding to the English vs. Spanish language versions of the measure. However, the items were not well-targeted for the sample (Figure 1), and a low reliability of person separation was found, although Cronbach's alpha was acceptable at .78. The utility of the response scale should be evaluated in future studies given infrequent use of higher categories in our sample. A more diverse and larger sample would be useful in determining whether the response scale should be revised.

EP scale—The Empathic-Prosocial Scale also showed no DIF for child age or gender, but DIF was found for five of eight items for Latina ethnicity, two items for majority/minority status, and for four items based on language version. This extent of DIF suggests there may be language or cultural issues involved in interpretation of the construct that are not resolved by careful translation of the measure. However, no significant DIF was found for Latinas responding to the English vs. Spanish versions of the measure. Thus, DIF more likely reflects a cultural difference in Latino culture that impacts the experience or expression of CU traits, and, relatedly, item endorsement. While reflecting on each individual test item goes beyond the scope of this article, some items included in the ICU are ethnocentric

that, while CU behaviors may be associated with EP behaviors, they are inherently different constructs. Future work is needed to explore whether the EP and CU behaviors are differentially related to parents' reports of children's cognitive and affective empathy, and how early adverse experiences, such as IPV exposure and related child maltreatment, shape the development of, and relationships between, these traits and behaviors.

Limitations

The generalizability of our findings is open to question due to the use of a convenience sample of mother-child dyads recruited from community-based IPV services. First, many families experiencing IPV do not seek out community-based services. Therefore, results of this study may not generalize to families in the larger community who are impacted by IPV. In addition, a more equally distributed representation of races and ethnicities would have improved the study and allowed for the examination of differences between specific subgroups. It is also important to note that inclusion criteria for the larger study included having had a pet in the home during the previous year. Although it is estimated that 75% of families with children over the age of 6 report pets in the home (American Veterinary Medical Association, 2007), our findings may not generalize to samples of children living in households without pets. Pets provide children with opportunities for empathy and social competence development and support resilience by serving as sources of comfort and companionship (DeGue, 2011; Melson, 2003); this, alongside the strong representation of Latino children, may help to explain the relatively low levels of CU traits in our sample, but we are unable to examine this hypothesis in the current study. Alternatively, as mothers were able to select which one of her eligible children participated in the study, data may reflect children with lower levels of behavioral challenges who caregivers thought would be more likely to comply with study procedures. Finally, our study could have been strengthened by conducting validity analyses with observational measures of behavioral and antisocial symptomology, in addition to caregiver and child self-reports of behaviors.

Conclusion and Future Directions

The assessment of CU traits may be particularly beneficial for mental health, domestic violence, and child welfare professionals who aim to effectively assess the needs of families and refer children and their caregivers to appropriate resources and/or tailored intervention services (McDonald et al., 2016). Children with CU traits are a particularly high-risk group who are resistant to many evidence-based intervention efforts that rely on educative techniques to reduce aggressive responses to conflict and promote prosocial behavior (Frick & White, 2008). A short, reliable and valid tool that can be administered by helping professionals during intake and assessment procedures is needed to identify children with CU traits who would likely be resistant to typical intervention approaches employed by IPV and family intervention services (McDonald et al., 2016). As a whole, the parent-report ICU is a promising tool for researchers and clinicians due to its inclusion of both CU and EP behaviors (Willoughby et al., 2015); however, our results do not support the use of the 24-item ICU, and recently proposed clinical cut-off scores (see Docherty, Boxer, Huesmann, O'Brien, & Bushman, 2016; Kimonis, Fanti, & Singh, 2014) to assess CU traits among children recruited from community-based IPV services. Our findings support retaining 18

items from the ICU, which comprise two distinct, but related, CU and EP scales identified in recent studies of the parent-report ICU.

The two shortened and revised CU and EP scales that were supported via our MIRT analyses offer a starting point to guide future revisions of the ICU in order to enhance its utility when employed to assess CU and EP traits in community-based samples of children with early adverse experiences. Future research is needed to expand the range of trait coverage of both scales and enhance the cross-cultural generalizability of the English and North American Spanish translations of the measure by removing, adding, and/or revising items in a culturally responsive manner (i.e., items displayed substantial DIF). The current study was only able to assess differences between Latinas and White non-Latina respondents and North American Spanish and English translations. As 42% of Latinas in our sample completed the Spanish-language version, our sample size was large enough to assess differences between the English language version for Latina respondents and Spanish language version for Latina respondents, with no resulting DIF. Thus, we conclude that DIF more likely resulted from cultural issues and not translation. Future studies would do well to address this finding. This is particularly important given that Latinos are the second largest, and second fastest growing, ethnic group in the United States (Ennis, Rios-Vargas, & Albert, 2011). Such cultural considerations have not been examined in the context of the parent-report version of the ICU, and items may operate differently according to these identities. With regard to the North American Spanish language version, assessing DIF across US born and recently immigrated Latinos may be particularly beneficial to enhancing knowledge of the psychometric properties of the ICU. Future studies would also do well by examining the predictive value of CU and EP traits in relation to ODD/CD within this group as prior studies have indicated that CU traits may be less associated with CD in some cultures (Kumsta, Sonuga-Barke, & Rutter, 2012).

Finally, given prior research indicating that recommended clinical cut-off scores for the ICU total score differ by caregiver gender (Kimonis, et al., 2014), research is needed to assess whether the EP/CU factor structure is stable across informant. While the current study and Waller et al. (2015) report on maternal caregivers, Willoughby et al. (2015) and Hawes (2014) did not indicate the gender of the caregiver respondent. Clearly, additional psychometric evaluations of the ICU are needed and our findings, alongside Hawes et al. (2014) and Ray et al. (2016), highlight the importance of IRT approaches in making meaningful improvements to the ICU in order to better assess CU and EP traits in vulnerable children at risk for psychosocial maladjustment.

Acknowledgments

This research was funded by Grant 5R01-HD-66503-4 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) and Grant 2015-0709 from the American Society for the Prevention of Cruelty to Animals (ASPCA). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or ASPCA. The authors would like to thank the community-based domestic violence advocates for their contribution to this work.

References

- Adams RJ, Wilson M, Wang WC. The multidimensional random coefficients multinomial logit model. *Applied psychological measurement*. 1997; 21(1):1–23.
- Adams, RJ., Wu, ML., Wilson, MR. ACER ConQuest: Generalised Item Response Modelling Software [Computer software] Version 4. Camberwell, Victoria: Australian Council for Educational Research; 2015.
- Achenbach, T., Rescorla, L. Manual for the ASEBA school-age forms & profiles. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families; 2001.
- Akaike H. A new look at the statistical model identification. *IEEE transactions on automatic control*. 1974; 19(6):716–723.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th. Washington, DC: Author; 2000. text rev.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th. Washington, DC: Author; 2013.
- American Veterinary Medical Association. Pet ownership and demographics sourcebook. American Veterinary Medical Association; Schaumburg, IL: 2007.
- Babcock JC, Green CE, Webb SA, Yerington TP. Psychophysiological profiles of batterers: Autonomic emotional reactivity as it predicts the antisocial spectrum of behavior among intimate partner abusers. *Journal of Abnormal Psychology*. 2005; 114(3):444–55. [PubMed: 16117581]
- Barker ED, Salekin RT. Irritable oppositional defiance and callous unemotional traits: Is the association partially explained by peer victimization? *Journal of Child Psychology and Psychiatry*. 2012; 53(11):1167–1175. [PubMed: 22783837]
- Benesch C, Görtz-Dorten A, Breuer D, Döpfner M. Assessment of callous-unemotional traits in 6 to 12 year-old children with oppositional defiant disorder/conduct disorder by parent ratings. *Journal of Psychopathology and Behavioral Assessment*. 2014; 36:519–529.
- Blair RJR. Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition*. 2005; 14(4):698–718. [PubMed: 16157488]
- Bond, TG., Fox, CM. Applying the Rasch model: Fundamental measurement in the human sciences. Mahwah: Erlbaum; 2007.
- Burke JD, Loeber R, Lahey BB. Adolescent conduct disorder and interpersonal callousness as predictors of psychopathy in young adults. *Journal of Clinical Child and Adolescent Psychology*. 2007; 36:334–346. [PubMed: 17658978]
- Carlson M, Oshri A, Kwon J. Child maltreatment and risk behaviors: The roles of callous/unemotional traits and conscientiousness. *Child Abuse & Neglect*. 2015; 50:234–243. [PubMed: 26233813]
- Ciucci E, Baroncelli A, Franchi M, Golmaryami FN, Frick PJ. The Association between callous-unemotional traits and behavioral and academic adjustment in children: Further validation of the Inventory of Callous-Unemotional Traits. *Journal of Psychopathology and Behavioral Assessment*. 2014; 36(2):189–200.
- Coyne SM, Nelson DA, Graham-Kevan N, Keister E, Grant DM. Mean on the screen: Psychopathy, relationship aggression, and aggression in the media. *Personality and Individual Differences*. 2010; 48:288–293.
- Dadds MR, Hawes DJ, Frost AD, Vassallo S, Bunn P, Hunter K, Merz S. Learning to ‘talk the talk’: The relationship of psychopathic traits to deficits in empathy across childhood. *Journal of Child Psychology and Psychiatry*. 2009; 50(5):599–606. [PubMed: 19445007]
- Dadds MR, Hunter K, Hawes DJ, Frost ADJ, Vassallo S, Bunn P, Merz S, Masry YE. A measure of cognitive and affective empathy in children using parent ratings. *Child Psychiatry and Human Development*. 2008a; 39(2):111–122. [PubMed: 17710538]
- Dadds MR, El Masry Y, Wimalaweera S, Guastella AJ. Reduced eye gaze explains “fear” blindness in childhood psychopathic traits. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2008b; 47:455–463. [PubMed: 18388767]

- DeGue, S. A triad of family violence: Examining overlap in the abuse of children, partners, and pets. In: Blazina, C. Boyraz, G., Shen-Miller, D.S., editors. *The psychology of the human-animal bond: A resource for clinicians and researchers*. New York: Springer; 2011. p. 245-262.
- DeVellis, R.F. *Scale development: Theory and applications*. Thousand Oaks, CA: SAGE Publications; 2003.
- Docherty M, Boxer P, Huesmann LR, O'Brien M, Bushman B. Assessing callous-unemotional traits in adolescents: Determining cutoff scores for the Inventory of Callous and Unemotional Traits. *Journal of Clinical Psychology*. 2016; doi: 10.1002/jclp.22313
- Ennis, SR., Ríos-Vargas, M., Albert, NG. *The Hispanic population: 2010*. US Department of Commerce, Economics and Statistics Administration, US Census Bureau; 2011. Retrieved from <http://www.census.gov/prod/cen2010/briefs/c2010br-04.pdf>
- Essau CA, Sasagawa S, Frick PJ. Callous-unemotional traits in a community sample of adolescents. *Assessment*. 2006; 13:454–469. [PubMed: 17050915]
- Fanti KA, Frick PJ, Georgiou S. Linking callous- unemotional traits to instrumental and noninstrumental forms of aggression. *Journal of Psychopathology and Behavioral Assessment*. 2009; 31:285–298.
- Forth, AE., Kosson, D., Hare, RD. *The Hare PCL: Youth version*. Toronto, ON: Multi-Health Systems; 2003.
- Frick, PJ. Unpublished rating scale. University of New Orleans; 2004. *The Inventory of callous-unemotional traits*.
- Frick PJ. Early Identification and Treatment of Antisocial Behavior. *Pediatric Clinics of North America*. 2016; 63(5):861–871. [PubMed: 27565364]
- Frick PJ, Dickens C. Current perspectives on conduct disorder. *Current Psychiatry Reports*. 2006; 8:59–72. [PubMed: 16513044]
- Frick, PJ., Hare, RD. *The antisocial processes screening device*. Toronto, ON: Multi-Health Systems; 2001.
- Frick PJ, Kimonis ER, Dandreaux DM, Farell JM. The 4 year stability of psychopathic traits in non-referred youth. *Behavioral Sciences & the Law*. 2003; 21:713–36. [PubMed: 14696028]
- Frick, PJ., Marsee, MA. Psychopathy and developmental pathways to antisocial behavior in youth. In: Patrick, C., editor. *Handbook of psychopathy*. New York: Guilford Press; 2006. p. 353-374.
- Frick PJ, Nigg JT. Current issues in the diagnosis of attention deficit hyperactivity disorder, oppositional defiant disorder, and conduct disorder. *Annual Review of Clinical Psychology*. 2012; 8:77–107.
- Frick PJ, White SF. Research review: The importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of Child Psychology and Psychiatry*. 2008; 49(4):359–375. [PubMed: 18221345]
- Gao Y, Zhang W. 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*. 2016; 38(3):331–340. [PubMed: 28255197]
- Gilbert R, Widom CS, Browne K, Fergusson D, Webb E, Janson S. Burden and consequences of child maltreatment in high-income countries. *Lancet*. 2009; 373(9657):68–81. [PubMed: 19056114]
- Hawes SW, Byrd AL, Henderson CE, Gazda RL, Burke JD, Loeber R, Pardini DA. Refining the parent-reported Inventory of Callous–Unemotional Traits in boys with conduct problems. *Psychological Assessment*. 2014; 26(1):256–266. [PubMed: 24188153]
- Henry J, Pingault JB, Boivin M, Rijdsdijk F, Viding E. Genetic and environmental aetiology of the dimensions of callous-unemotional traits. *Psychological Medicine*. 2016; 46(02):405–414. [PubMed: 26456336]
- Horan JM, Brown JL, Jones SM, Aber JL. Assessing invariance across sex and race/ethnicity in measures of youth psychopathic characteristics. *Psychological Assessment*. 2015; 27(2):657–668. [PubMed: 25383582]
- Houghton S, Hunter SC, Crow J. Assessing callous unemotional traits in children aged 7- to 12-years: A confirmatory factor analysis of the Inventory of Callous Unemotional Traits. *Journal of Psychopathology and Behavioral Assessment*. 2013; 35:215–222.

- Jolliffe D, Farrington DP. Examining the relationship between low empathy and bullying. *Aggressive Behavior*. 2006; 32(6):540–550.
- Jolliffe D, Farrington DP. Examining the relationship between low empathy and self-reported offending. *Legal and Criminological Psychology*. 2007; 12(2):265–286.
- Kang T, Cohen AS. IRT model selection methods for dichotomous items. *Applied Psychological Measurement*. 2007; 31(4):331–358.
- Kahn RE, Frick PJ, Golmaryami FN, Marsee MA. The moderating role of anxiety in the associations of callous-unemotional traits with self-report laboratory measures of affective and cognitive empathy. *Journal of Abnormal Child Psychology*. 2016:1–14. [PubMed: 26687502]
- Kimonis ER, Fanti KA, Anastassiou-Hadjicharalambous X, Mertan B, Goulter N, Katsimicha E. Can callous-unemotional traits be reliably measured in preschoolers? *Journal of Abnormal Child Psychology*. 2016; 44(4):625–638. [PubMed: 26344015]
- Kimonis ER, Fanti K, Singh JP. Establishing cut-off scores for the parent-reported Inventory of Callous-Unemotional Traits. *Archives of Forensic Psychology*. 2014; 1:27–48.
- Kimonis ER, Frick PJ, Skeem JL, Marsee MA, Cruise K, Munoz LC, Aucoin KJ, Morris AS. Assessing callous-unemotional traits in adolescent offenders: Validation of the Inventory of Callous-Unemotional Traits. *International Journal of Law and Psychiatry*. 2008; 31:241–252. [PubMed: 18514315]
- Kumsta R, Sonuga-Barke E, Rutter M. Adolescent callous–unemotional traits and conduct disorder in adoptees exposed to severe early deprivation. *The British Journal of Psychiatry*. 2012; 200(3):197–201. [PubMed: 22116980]
- Lenhard, W., Lenhard, A. Hypothesis tests for comparing correlations [online calculator]. Bibergau, Germany: Psychometrica; 2014. Available from <https://www.psychometrica.de/correlation.html>
- Linacre, JM. Winsteps® Rasch measurement computer program user's guide. Beaverton, Oregon: Winsteps.com; 2016. Available from <http://www.winsteps.com/>
- Linacre, JM. Winsteps® (Version 3.92.0) [Computer Software]. Beaverton, Oregon: Winsteps.com; 2016. Available from <http://www.winsteps.com/>
- Loeber R, Green SM, Lahey BB, Stouthamer-Loeber M. Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. *Journal of Abnormal Child Psychology*. 1991; 19(1):75–95. [PubMed: 2030249]
- Lopez, MH. Latinos and education: Explaining the attainment gap. Washington, DC: Pew Hispanic Center; 2009.
- Lynam DR. Pursuing the psychopath: Capturing the fledgling psychopath in a nomological net. *Journal of Abnormal Psychology*. 1997; 106:425–38. [PubMed: 9241944]
- Lynam DR, Caspi A, Moffitt TE, Loeber R, Stouthamer-Loeber M. Longitudinal evidence that psychopathy scores in early adolescence predict adult psychopathy. *Journal of Abnormal Psychology*. 2007; 116:155–65. [PubMed: 17324026]
- Lynam DR, Charnigo R, Moffitt TE, Raine A, Loeber R, Stouthamer-Loeber M. The stability of psychopathy across adolescence. *Development and Psychopathology*. 2009; 21:1133–1153. [PubMed: 19825261]
- Lynam DR, Derefinko KJ, Caspi A, Loeber R, Stouthamer-Loeber M. The content validity of juvenile psychopathy: An empirical examination. *Psychological Assessment*. 2007; 19(3):363–367. [PubMed: 17845128]
- MacMillan HL, Wathen CN, Varcoe CM. Intimate partner violence in the family: Considerations for children's safety. *Child Abuse & Neglect*. 2013; 37(12):1186–1191. [PubMed: 23830555]
- McDonald SE, Graham-Bermann SA, Maternick A, Ascione F, Williams JH. Patterns of adjustment among children exposed to intimate partner violence: A person-centered approach. *Journal of Child & Adolescent Trauma*. 2016; 9(2):1–16. [PubMed: 27152132]
- Melson GF. Child development and the human-companion animal bond. *American Behavioral Scientist*. 2003; 47:31–39.
- Myers ND, Wolfe EW, Feltz DL. An evaluation of the psychometric properties of the coaching efficacy scale for American coaches. *Measurement in Physical Education and Exercise Science*. 2005; 9(3): 135–160.

- O’Kearney R, Salmon K, Liwag M, Fortune CA, Dawel A. Emotional abilities in children with Oppositional Defiant Disorder (ODD): Impairments in perspective-taking and understanding mixed emotions are associated with high callous–unemotional traits. *Child Psychiatry & Human Development*, 1– 12. 2016; Advanced online publication. doi: 10.1007/s10578-016-0645-4
- Nunnally, JC., Bernstein, IH. *Psychometric theory*. 3rd. NY: McGraw-Hill; 1994.
- Ray JV, Frick PJ, Thornton LC, Steinberg L, Cauffman E. Positive and negative item wording and its influence on the assessment of callous-unemotional traits. *Psychological Assessment*. 2016; 28(4): 394. [PubMed: 26121386]
- Roose A, Bijttebier P, Decoene S, Claes L, Frick PJ. Assessing the affective features of psychopathy in adolescence: a further validation of the Inventory of Callous and Unemotional Traits. *Assessment*. 2010; 17(1):44–57. [PubMed: 19797326]
- Sable P. Pets, attachment, and well-being across the life cycle. *Social Work*. 1995; 40(3):334–341. [PubMed: 7761919]
- Shenk CE, Dorn LD, Kolko DJ, Rausch JR, Insana SP. Prior exposure to interpersonal violence and long-term treatment response for boys with a disruptive behavior disorder. *Journal of Traumatic Stress*. 2014; 27(5):585–592. [PubMed: 25270151]
- Slep AMS, O’leary SG. Parent and partner violence in families with young children: rates, patterns, and connections. *Journal of Consulting and Clinical Psychology*. 2005; 73(3):435. [PubMed: 15982141]
- Smith AB, Rush R, Fallowfield LJ, Velikova G, Sharpe M. Rasch fit statistics and sample size considerations for polytomous data. *BMC Medical Research Methodology*. 2009; 8:33–43.
- Swogger MT, Walsh Z, Kosson DS. Domestic violence and psychopathic traits: Distinguishing the antisocial batterer from other antisocial offenders. *Aggressive Behavior*. 2007; 33(3):253–260. [PubMed: 17444531]
- Van de Vijver F, Hambleton RK. Translating tests: Some practical guidelines. *European Psychologist*. 1996; 1(2):89–99.
- Waller R, Wright AG, Shaw DS, Gardner F, Dishion TJ, Wilson MN, Hyde LW. Factor structure and construct validity of the parent-reported Inventory of Callous-Unemotional Traits among high-risk 9-year-olds. *Assessment*. 2015; 22(5):561–580. [PubMed: 25371449]
- Weiler BL, Widom CS. Psychopathy and violent behaviour in abused and neglected young adults. *Criminal Behaviour and Mental Health*. 1996; 6:253–271.
- Willoughby MT, Mills-Koonce WR, Waschbusch DA, Gottfredson NC, Family Life Project Investigators. An examination of the parent report version of the Inventory of Callous-Unemotional Traits in a community sample of first-grade children. *Assessment*. 2015; 22(1):76–85. [PubMed: 24820529]
- Zayas LH. Hispanic family ecology and early childhood socialization: Health care implications. *Family Systems Medicine*. 1994; 12(3):315–325.
- Zayas LH, Solari F. Early childhood socialization in Hispanic families: Context, culture, and practice implications. *Professional Psychology: Research and Practice*. 1994; 25(3):200–206.

Figure 1.

Map of person and item distributions for the 2-factor ICU with 18 items. Dimension 1 = Callous-Unemotional; Dimension 2 = Empathic-Prosocial. Item numbers are from original ICU measure.

Table 1

Participant Demographic and Variable Summary

Variable	Mean/SD	Sk/Ku ^a	n	%
Child's Age	9.07/1.64	.25/−1.16		
Child's Gender				
Male			153	52.6%
Female			138	47.4%
Child's Race/Ethnicity				
American Indian or Alaska Native			3	1.0%
Asian			1	0.3%
African American			10	3.4%
White			64	22.0%
Latina			161	55.3%
Multiracial			52	17.9%
Mother's Race/Ethnicity				
American Indian, Alaska Native, Hawaiian			6	2.0%
Asian			1	0.3%
African American			10	3.4%
White			78	26.9%
Latina			176	60.7%
Multiracial			19	6.5%
Mother's Identification as Latina/Hispanic				
Yes			197	67.7%
No			93	32.0%
Interview in Spanish?				
Yes			98	33.7%
No			193	66.3%
Rule Breaking Behavior T-score	54.96/6.62	1.60/1.89		
Aggressive Behavior T-score	55.48/8.11	1.83/3.28		
Internalizing Problems T-score	53.22/12.30	.11/−.56		
Externalizing Problems T-score	50.00/12.21	.33/−.66		
Oppositional-Defiant T-score	54.88/8.76	1.50/1.35		
Conduct Problems T-score	54.81/7.44	1.77/2.54		
GEM-Affective	−2.29/17.94	−.28/−1.03		
GEM-Cognitive	11.73/9.39	−.55/−.30		
GEM-Total	20.55/26.65	−.30/−.65		
Person Logit ICU-Callous-Unemotional	−2.85/1.47	.23/−1.39		
Person Logit ICU-Empathic-Prosocal	−1.02/1.56	−.37/1.06		

^aSk = skewness; Ku = kurtosis

Table 2

Items of ICU Models by Factor

Factor	24-Item ICU (Original)		20-Item ICU (Literature)		18-Item ICU	
	# of Items	Items Included	# of Items	Items Excluded	# of Items	Items Excluded
2-Factor Model	11	2, 4, 6, 7, 9, 11, 12, 18, 20, 21, 22	10	2	10	2
	13	1, 3, 5, 8, 10, 13, 14, 15, 16, 17, 19, 23, 24	10	10, 15, 23	8	3, 10, 15, 16, 17
3-Factor Model	11	2, 4, 7, 8, 9, 10, 11, 12, 18, 20, 21	9	2, 10	9	2, 10
	8	3, 5, 13, 15, 16, 17, 23, 24	6	15, 23	4	3, 15, 16, 17
	5	1, 6, 14, 19, 22	5		5	

Table 3

Comparison of Model Fit for the ICU Models

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
	24-Item ICU (with all items)			20-Item ICU (without Items 2,10,15, 23)			18-Item ICU (without Items 2,3,10,15,16,17)		
	1-Factor	2-Factor	3-Factor	1-Factor	2-Factor	3-Factor	1-Factor	2-Factor	3-Factor
Sample Size (<i>n</i>)	291	291	291	291	291	291	291	291	291
# of Parameters (<i>p</i>)	72	74	77	60	62	65	54	56	59
Degree of Freedom (<i>df</i>)	219	217	214	231	229	226	237	235	232
Deviance (<i>d</i>)	11849.03	11655.15	11502.53	9511.96	9276.64	9224.47	8366.23	8174.40	8188.01
AIC	11993.03	11803.15	11656.53	9631.96	9400.64	9354.47	8474.23	8286.40	8306.01
AIC Difference	336.50	146.63		277.50	46.17		187.83		19.61
AIC-pc	54.76	54.39	54.47	41.70	41.05	41.39	35.76	35.26	35.80

Note. (1) AIC—Akaike information criterion; AIC Difference—difference between one of the three factor models (1-, 2-, and 3-Factor) and the model with the smallest AIC for each type of ICU; AIC-pc—AIC proportionality constant. (2) The bolded values of AIC or AIC-pc are the smallest among three factor models. (3) All AIC difference values are significant at $p = .001$. The model fit of Model 8 is the best.

Table 4

DIF Items and Groups with a Significant Lower Item Difficulty

Factor	Item	Race	Latina/White	Language
F1: Callous-Unemotional	Item 7	L-majority ***	L-White ***	L-English ***
	Item 12	L-minority **	L-Latina **	L-Spanish ***
F2: Empathic-Prosocial	Item 1		L-White **	L-English **
	Item 5	L-minority ***	L-Latina ***	L-Spanish ***
	Item 8		L-Latina ***	
	Item 14	L-majority ***	L-White ***	L-English ***
	Item 19		L-White **	L-English ***

Note. DIF assessed at $p = .01$ and logit position = .50 difference,

** $p = .01$,

*** $p = .001$, L—lower item difficulty/easier to agree with.

Table 5

Pearson Correlation among Key Study Variables (n = 291)

	1	2	3	4	5	6	7	8	9	10	11
1. ICU: Callous-Unemotional	1										
2. ICU: Empathic-Prosocial	.45***	1									
3. CBCL: Aggressive Behavior T-Score	.63***	.33***	1								
4. CBCL: Internalizing T-Score	.46***	.13*	.58***	1							
5. CBCL: Externalizing T-Score	.58***	.31***	.87***	.65***	1						
6. CBCL: Total Problems T-Score	.58***	.23***	.77***	.89***	.87***	1					
7. CBCL: Oppositional Defiant T-Score	.61***	.28***	.93***	.59***	.85***	.77***	1				
8. CBCL: Conduct Problems T-Score	.57***	.33***	.85***	.52***	.82***	.71***	.80***	1			
9. GEM Total	.01	-.28***	-.01	.19***	.03	.12*	.06	.02	1		
10. GEM: Affective Empathy	.25***	-.04	.20***	.29***	.22***	.28***	.25***	.22***	.88***	1	
11. GEM: Cognitive Empathy	-.45***	-.25***	-.38***	-.28***	-.37***	-.37***	-.39***	-.39***	-.18**	-.52***	1

Note. ICU—the Inventory of Callous and Unemotional Traits; CBCL—the Child Behavior Checklist; GEM—the Griffith Empathy Measure.

*
p .05;

**
p .01;

p .001

Table 6

Person Logit Position: Callous-Unemotional and Empathic/Prosocial Scale

Gender	Latina	# of Children (n)	F1: Callous-Unemotional		F2: Empathic-Prosocial	
			Mean	SD	Mean	SD
Male	No	48	-2.23	1.37	-0.89	1.46
	Yes	105	-3.13	1.43	-0.73	1.43
Female	No	45	-2.02	1.12	-1.26	1.35
	Yes	92	-3.24	1.48	-1.3	1.78
Total		290	-2.84	1.47	-1.02	1.56

Table 7

Base (gender and ethnicity) and Incremental Regression Models with Callous-Unemotional (CU) and Empathic-Prosocial (EP) Scales as the Independent Variables

	F1: Callous/Unemotional						F2: Empathic/Prosocial							
Dependent Variable	R	R ²	Incremental R ²	Adjusted R ²	Predictor	Beta	p	R	R ²	Incremental R ²	Adjusted R ²	Predictor	Beta	p
Rule-Breaking T-score	.49	.24	.164	.23	Gender	−.01	.85	.37	.14	.066	.13	Gender	.03	.62
					Ethnicity	−.16	.002					Ethnicity	−.28	<.001
					CU	.42	<.001					EP	.26	<.001
Aggressive T-score	.64	.41	.336	.40	Gender	−.06	.22	.43	.19	.114	.18	Gender	−.01	.91
					Ethnicity	−.11	.024					Ethnicity	−.27	<.001
					CU	.60	<.001					EP	.34	<.001
Internalizing T-score	.52	.27	.141	.27	Gender	.02	.70	.39	.15	.021	.15	Gender	.04	.46
					Ethnicity	−.27	<.001					Ethnicity	−.37	<.001
					CU	.39	<.001					EP	.15	.008
Externalizing T-score	.61	.38	.264	.37	Gender	−.07	.14	.47	.22	.103	.21	Gender	−.02	.67
					Ethnicity	−.19	<.001					Ethnicity	−.34	<.001
					CU	.53	<.001					EP	.33	<.001
ODD T-score	.63	.39	.309	.38	Gender	−.04	.35	.41	.16	.083	.16	Gender	−.01	.98
					Ethnicity	−.13	.005					Ethnicity	−.29	<.001
					CU	.58	<.001					EP	.29	<.001
CD T-score	.59	.35	.371	.34	Gender	.005	.91	.45	.20	.122	.19	Gender	.06	.29
					Ethnicity	−.15	.004					Ethnicity	−.29	<.001
					CU	.53	<.001					EP	.35	<.001
GEM-Affective	.36	.13	.034	.12	Gender	.12	.03	.31	.10	.00	.09	Gender	.12	.04
					Ethnicity	−.23	<.001					Ethnicity	−.28	<.001
					CU	.19	.001					EP	.02	.75
GEM-Cognitive	.47	.23	.156	.22	Gender	−.04	.43	.37	.14	.069	.13	Gender	−.08	.15

F2: Empathic/Prosocial									
F1: Callous/Unemotional									
Dependent Variable	R	R ²	Incremental R ²	Adjusted R ²	Predictor	Beta	p	R	R ²
GEM-Total	.29	.08	.002	.07	Ethnicity	.16	.004	.38	.14
					CU	-.41	<.001		
					Gender	.15	.007		
GEM-Total	.29	.08	.002	.07	Ethnicity	-.25	.001	.38	.14
					CU	-.05	.42		
					Gender	.11	.04		
GEM-Total	.29	.08	.002	.07	Ethnicity	-.23	<.001	.38	.14
					CU	-.05	.42		
					Gender	-.26	<.001		