June 2020

A Comprehensive Examination of Clinical Cutoff Scores for the Inventory of Callous-Unemotional Traits (ICU)

Emily C. Kemp

A COMPREHENSIVE EXAMINATION OF CLINICAL CUTOFF SCORES FOR THE INVENTORY OF CALLOUS-UNEMOTIONAL TRAITS (ICU)

A Thesis

Submitted to the Graduate Faculty of
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Arts

in

The Department of Psychology

by
Emily Claire Kemp
B.S., University of Texas at Austin, December 2014
August 2020

Table of Contents

List of Tables

List of Figures

Abstract

Youth with elevated callous-unemotional (CU) traits represent a clinically important subgroup of youth who display particularly severe conduct problems and antisocial behavior and thusly impose great costs to themselves, other individuals, and society. The recent addition of the specifier for CU traits, "with Limited Prosocial Emotions (LPE)," to major classification systems has prompted the need for comprehensive and valid assessment tools that aid in the identification of these traits. One such tool is a multi-informant questionnaire, the Inventory of Callous-Unemotional Traits (ICU). However, a major limitation of this measure is a lack of well-validated cutoff scores. With this, the present study compared the clinical utility of various proposed cutoff methods and scores (i.e., empirically derived cutoffs using receiver operating characteristic (ROC), normative cutoffs, and four-item approximations of LPE criteria) in both a longitudinal sample of justice-involved adolescents (N = 1,216; M_{age} = 15.29, SD = 1.29) and a cross-sectional sample of community youth (N = 289; M_{age} = 11.47 years; SD = 2.26). A series of chi-square and independent-samples T-tests were conducted with the use of clinically important validators (e.g., juvenile delinquency, social rejection), and average effect sizes were computed. As a result, comparisons of clinical utility are made and a range of optimal cutoff scores, for each ICU informant version, are provided for the detection of elevated CU traits. The outcomes of this study have important practical implications as they provide researchers and clinicians with guidance on choosing cutoffs that may aid in determining elevations of CU traits.

Keywords: callous-unemotional (CU) traits, inventory of callous-unemotional traits (ICU), normative cutoff scores, receiver operating characteristic (ROC), juvenile delinquency, conduct problems, peer rejection

Introduction

Callous-Unemotional Traits: Clinical Significance

Callous-unemotional (CU) traits are characterized by deficits in several affective, interpersonal, and motivational domains. More specifically, CU traits are defined by impairments in empathy, remorse or guilt, emotional depth (i.e., in the form of shallow or impoverished affect), and concern about performance in important activities (e.g., academic performance; Frick, Ray, Thornton, & Kahn, 2014a). Furthermore, CU traits are theorized to represent an affective dimension of the broader construct of psychopathy in adults (Hare & Neumann, 2008) and the affective components of conscience in children (Frick, Ray, Thornton, & Kahn, 2014b). CU traits have been found to be stable over time (Frick et al., 2014b), with youth high on these traits at an increased likelihood for continuing antisociality as well as displaying elevated psychopathic traits into adulthood (Barry et al., 2000; Feilhauer & Cima, 2013; Frick et al., 2014b; Frick & White, 2008; Longman, Hawes, & Kohlhoff, 2016; Lynam, Caspi, Moffitt, Loeber, & Stouthamer-Loeber, 2007; Viding, 2004). Even prior to adulthood, CU traits are associated with more stable and severe conduct problems and an increased risk of juvenile delinquency (Frick et al., 2003a; Frick, Cornell, Barry, Bodin, & Dane, 2003; Frick, Stickle, Dandreaux, Farrell, & Kimonis, 2005; McMahon, Witkiewitz, & Kotler, 2010). That is, prior studies have found youth with elevated CU traits to engage in more persistent, severe (i.e., violent), and harmful forms of antisocial behavior and criminal offending (Baskin-Sommers, Waller, Fish, & Hyde, 2015; Dadds, Whiting, & Hawes, 2006; Kahn, Byrd, & Pardini, 2013; Lawing, Frick, & Cruise, 2010; Pardini & Fite, 2010) and to show an increased risk for weapon use (e.g., gun carrying and gun use during crime; Robertson et al., in press; Saukkonen et al., 2016). More specifically, youth with elevated CU traits are more likely to display both reactive

and proactive aggression, including premeditated and instrumental forms of aggression used for personal gain, to display aggression in response to lower levels or a complete absence of provocation, and to engage in aggressive behaviors (e.g., direct bullying) that are more harmful to their victims (Crapanzano, Frick, Childs, & Terranova, 2011; Fanti, Frick, & Georgiou, 2009; Fanti & Kimonis, 2012; Frick et al., 2003b; Golmaryami et al., 2016; Kruh, Frick, & Clements, 2005; Lawing et al., 2010; Marsee & Frick, 2007; Muñoz, Frick, Kimonis, & Aucoin, 2008).

Youth with elevated CU traits, even in the absence of serious conduct problems, experience greater psychosocial impairment as well as academic and interpersonal difficulties (Barry, Barry, Deming, & Lochman, 2008; Ciucci, Baroncelli, Franchi, Golmaryami, & Frick, 2014; Graziano et al., 2016; Horan, Brown, Jones, & Aber, 2016; Waller, Baskin-Sommers, & Hyde, 2016). For example, CU traits have been associated with higher levels of peer rejection (Barnow, Lucht, & Freyberger, 2005; Barry et al., 2008; Matlasz, Frick, & Clark, under review; Piatigorsky & Hinshaw, 2004) as well as lower quality peer relationships, with findings that youth higher on CU traits have shorter lasting, less intimate, and less satisfying peer relationships that also include more interpersonal conflict (Haas, Becker, Epstein, & Frick, 2018; Miron, Satlof-Bedrick, & Waller, 2020; Muñoz, Kerr, & Besic, 2008). Furthermore, Graziano and colleagues (2016) found that, in addition to being less liked by peers, youth with elevated CU traits were more likely to be described as someone who "enjoys being mean." These findings were recently replicated by Matlasz and colleagues (under review) who found that CU traits were positively associated with peer-nominated levels of perceived "meanness" and are also supported by past findings that CU traits are positively associated with self-reported levels of meanness (Kyranides, Fanti, Sikki, & Patrick, 2017; Patrick & Drislane, 2015).

Lastly, unlike youth with conduct problems and non-elevated levels of CU traits, youth with elevated CU traits do not respond as highly to punishment or as well to many traditional approaches to mental health treatment (Hawes & Dadds, 2005; Hawes & Dadds, 2007; Hawes, Dadds, Brennan, Rhodes, & Cauchi, 2013; Wilkinson, Waller, & Viding, 2016). For example, Hawes & Dadds (2007) initially demonstrated poorer treatment outcomes in children with elevated CU traits, particularly boys with stable elevated CU traits, when implementing an evidence-based, manualized parent-training intervention over the course of ten weeks. Hawes and colleagues (2013) later replicated these findings in children being treated for conduct problems and found that CU traits uniquely moderated poorer treatment outcomes, even after controlling for autism spectrum disorder symptoms. More recently, Wilkinson and colleagues (2016) conducted a systematic review of treatment outcomes in youth with elevated CU traits and reported that, in at least half of treatment trials in forensic and clinical youth samples, CU traits were reported to moderate and be negatively related to treatment responsiveness (i.e., youth with elevated CU traits demonstrated less treatment sensitivity and smaller targeted reductions in antisocial behavior following treatment). However, it is important to note that these studies did not report that youth with elevated CU traits did not respond positively to intervention. Instead, it appears that youth with elevated CU traits often start treatment with the highest levels of problem behavior and, despite showing a reduction in conduct problems, still end treatment with the highest levels of problem behavior (White, Frick, Lawing, & Bauer, 2013).

In summary, it is clear that children and adolescents with elevated CU traits present a very important mental health concern. Their behavior is costly to others (e.g., causing monetary damage, inflicting emotional and physical pain and suffering), to themselves (e.g., reduced quality of interpersonal relationships and attained academic success), and to society more

globally (e.g., economic burden posed by costs to maintain juvenile justice systems and mental health services; Burt et al., 2018; Rivenbark et al., 2018), and they have been found to be less sensitive to traditional treatment approaches for youth with conduct problems.

Callous-Unemotional Traits and Developmental Pathways to Conduct Problems

In addition to the clinical utility of CU traits, there is also evidence that CU traits could be important for advancing causal theories of serious conduct problems. Research has made great strides to uncover the various risk factors that can lead to serious and chronic behavior problems (Dodge & Pettit, 2003; Fairchild et al., 2019; Frick & Viding, 2009; Moffitt, 2018) and, as such, has revealed that there are likely multiple causal pathways with a variety of dispositional (e.g., genetic, neural, physiological) and contextual (e.g., parenting) risk factors that interact with one another and lead to conduct disorder (CD; Frick, 2012, 2016). An example of one noteworthy distinction that research has uncovered is the clinical importance of specifying age of CD symptoms onset, a specifier that is included in both the Diagnostic and Statistical Manual of Mental Disorders (DSM-5; American Psychiatric Association, 2013) and the International Classification of Diseases (ICD-11; World Health Organization, 2018). That is, onset of CD symptoms in childhood (i.e., before age 10) denotes a significantly different and more severe trajectory with unique etiological and developmental mechanisms as compared to that of adolescent-onset (i.e., age 10 or after; Frick, 2012, 2016; Frick & Viding, 2009; Moffitt, 2018). More specifically, youth with childhood-onset CD are found to display conduct problems as early as preschool years and to continue to do so with increasing frequency and severity through childhood and adolescence, even showing higher rates of serious antisocial behavior and criminality into adulthood (Odgers et al., 2008). Contrastingly, youth with adolescent-onset CD are theorized to display an exaggerated rebelliousness (e.g., as demonstrated by reduced

bondedness with or respect for prosocial institutions and conventional values or rules) during a time that is normative for identity formation and exploration; therefore, serious behavior problems are more often confined to this particular developmental period (Dandreaux & Frick, 2009; Moffitt, 2018).

Importantly, research has suggested that elevated CU traits are more likely to be found within the childhood-onset group and, furthermore, seem to designate an etiologically distinct subgroup within the childhood-onset pathway (Frick et al., 2014a). For example, children with serious conduct problems and elevated CU traits show increased fearlessness and reduced emotional reactivity in response to various types of emotional stimuli, whereas children with conduct problems who are normative on CU traits show enhanced emotional reactivity to various types of emotional stimuli (Frick & White, 2008; Viding, Fontaine, & McCrory, 2012). These findings have led to theories that different causal mechanisms lead to the behavior problems in the two groups, with a) deficient emotional responsiveness leading to problems with conscience development and thusly promoting greater antisociality in those with elevated CU traits, and b) problems regulating emotions leading to conduct problems in those without elevated CU traits (Frick et al., 2014b).

Callous-Unemotional Traits and Diagnostic Classification

As a result of this research that supports both the clinical and etiological importance of CU traits, the most recent editions of major diagnostic and classification systems (i.e., the DSM-5 and ICD-11), have added the specifier, "with Limited Prosocial Emotions (LPE)," to refer to a subgroup of youth with severe behavior problems (20-30% of youth with CD) who also display significant or elevated levels of CU traits (Kahn, Frick, Youngstrom, Findling, & Youngstrom, 2012; Pardini & Fite, 2010; Pardini, Frick, & Moffitt, 2010; Rowe et al., 2010). Diagnostic

criteria for the LPE specifier consist of four behavioral or emotional indicators of CU traits, at least two of which must be present for the previous 12 months and displayed across multiple relationships and settings to meet criteria: 1) lack of remorse or guilt (e.g., only apologizing when prompted, lack of concern about negative consequences unless they inconvenience the youth); 2) callousness or lack of empathy (e.g., little concern for the effects of one's actions on others, displays of teasing or hurtful behaviors); 3) lack of caring about performance in important activities (e.g., showing little concern about or effort toward academic performance, sports, or extracurricular activities); and 4) shallow or deficient emotions (e.g., shallow affect, insincere displays of emotion (e.g., sadness) used only to obtain a desired outcome or to get out of trouble; American Psychiatric Association, 2013; World Health Organization, 2018). According to the DSM-5, the LPE specifier may be given for a diagnosis of CD only (American Psychiatric Association, 2013), while the ICD-11 allows the specifier be given for both conduct-dissocial and oppositional defiant disorder (ODD; World Health Organization, 2018).

Assessment of Callous-Unemotional Traits

With the inclusion of the LPE specifier in the latest versions of major classification systems for mental health disorders, there is an enhanced need to develop and refine assessments for CU traits. In research, the most common way that CU traits are assessed is with rating scales. For example, the Antisocial Process Screening Device (APSD; Frick & Hare, 2001) is a 20-item multi-informant (i.e., self-, parent-, and teacher-report) scale used in samples of children and adolescents. The APSD includes a 6-item CU subscale and is historically one of the most widely researched measurement devices for the assessment of CU traits and other dimensions associated with psychopathy (Kimonis et al., 2008; Sharp & Kine, 2008). Overall, total scores from the APSD have been found to successfully designate more severe subgroups of juvenile offenders,

show associations with earlier onset (i.e., childhood) offending, and be predictive of future antisocial behavior in justice-involved adolescents (Kimonis et al., 2008). Thus, the APSD has demonstrated utility for designating more behaviorally severe groups of high-risk youth. Additional measures that assess CU traits have included the Child Psychopathy Scale (CPS; Lynam, 1997), a parent-rated scale originally consisting of 41 items (a modified version allows for self-report and removes all questions regarding explicit antisocial behavior), and the Youth Psychopathic Traits Inventory (YPI; Andershed, Kerr, Stattin, & Levander, 2002), a 50-item self-report scale. Both of these scales contain items (i.e., 10 in the CPS and 5 in the YPI) that specifically assess CU traits, in addition to other items or subscales that assess interpersonal traits (e.g., grandiosity, manipulativeness), impulsivity, and thrill-seeking (Sharp & Kine, 2008). Importantly, these scales contain a number of limitations that reduce their utility as targeted and holistic measures of CU traits. First, all three scales include a very limited number of items (i.e., ranging from 5 to 10 items and comprising only one-tenth to one-third of the total scale) that directly assess CU traits. The restricted range of scores that result from this limited number of CU-specific items, in addition to the knowledge that CU traits are often not normally distributed in most samples, have likely contributed to reports of moderate internal consistency across many samples (Loney, Frick, Clements, Ellis, & Kerlin, 2003; Sharp & Kine, 2008). Second, items contained in the APSD and CPS are rated with limited response options (i.e., a 3-point Likert-type scale for the APSD and dichotomous yes/no response options for the CPS) that further restrict the range and variability of scores (Frick & Hare, 2001; Kimonis et al., 2008; Lynam, 1997). Third, the majority, if not all, of the CU-items found within all three scales contain wording in the same direction (e.g., all items are negatively worded in the CPS and all are positively worded in the YPI), thusly making response sets more likely (Andershed et al., 2002;

Kimonis et al., 2008; Lynam, 1997). Finally, the CPS and YPI were either developed for or have been primarily utilized in samples of older children and adolescents, ages 11-12 years and older, which has reduced the utility of such scales in research or clinical assessments involving younger children (Andershed et al., 2002; Lynam, 1997; Sharp & Kine, 2008).

The Inventory of Callous-Unemotional Traits (ICU)

Development. In order to overcome these limitations, the Inventory of Callous-Unemotional Traits (ICU) was developed to provide a comprehensive assessment of CU traits, as they are operationalized in the LPE specifier. More specifically, the ICU is a short but exhaustive (24-item) multi-informant questionnaire designed to extensively and efficiently assess CU traits in youth ranging in developmental stage (i.e., young children as young as 3 years to adults), demographics, and risk level (e.g., ranging from normative, large-scale community samples to high-risk samples of juvenile offenders; Kimonis et al., 2008). To first develop this measure, four items from the CU subscale of the APSD (i.e., "I care about how well I do at school or work;" "I feel bad or guilty when I do something wrong;" "I do not show my emotions to others;" "I am concerned about the feelings of others"; Frick & Hare, 2001) were used as the basis of the item content, as these items were found, across samples, to most consistently best measure CU traits (Frick, Bodin, & Barry, 2000). For each of these four items, three positively (i.e., higher scores indicating higher levels of CU traits) and three negatively worded (i.e., reversal) items were written to reduce potential response sets. Further, ICU item response options are based on a 4-point Likert-type scale that ranges from 0 (*Not at all true*) to 3 (*Definitely true*), thusly improving the range and variability of responses and disallowing an exact middle response rating. In order to allow for CU assessment across a diversity of participants, including a wide age range from young children to adults, there are five versions of the scale, including the youth

self-report scale and teacher- and parent-report scales that are separated by preschool and non-preschool youth versions (see Appendices A-C for items found in the ICU self-, parent-, and teacher-report versions; Frick, 2004). In addition, there are 28 approved language translations, making this scale especially valuable for cross-cultural and international research regarding CU traits.

The ICU has been used in over 250 published studies that have generally provided strong support for its ability to measure CU traits continuously (Ray & Frick, 2018). Across various factor analyses, there is support for an overarching CU dimension that is captured well by a unit weighting (i.e., summing) of the ICU items (Kliem et al., 2020; Ray & Frick, 2018). In a meta-analysis of 115 samples (N = 27,947), the average internal consistency of the 24-item total score from the ICU self-report version was α = .83, thus showing marked improvements from its predecessors (Cardinale & Marsh, 2017). Also reported in this meta meta-analysis, the mean correlation between ICU self-report total score and measures of the affective dimension of psychopathy was .40, and the mean correlation between ICU self-report and measures of empathy was -.25 (Cardinale & Marsh, 2017). Further, the mean correlations with measures of general externalizing problems, delinquency, and proactive aggression were .35, .30, and .44, respectively (Cardinale & Marsh, 2017). There have been fewer total studies validating the parent and teacher-report versions of the ICU, however, a number of studies have reported similar estimates of internal consistency for the 24-item ICU parent-report version (average α = .82; Ueno, Ackermann, Freitag, & Schwenck, 2019; Yoshida et al., 2019). In addition, the mean correlation between ICU parent-report total score and measures of the affective dimension of psychopathy was r = .27, and the mean correlations with measures of aggression and rule-breaking behavior were r = .51 and .55, respectively (Ueno et al., 2019). Two studies that

assessed the internal consistency of the 24-item ICU teacher-report version also reported good internal consistency (average α = .88; see Ueno et al., 2019). Ueno and colleagues (2019) also reported on the correlations between ICU teacher-report total score and measures of the affective dimension of psychopathy (*r* = .24) and of externalizing and behavior problems (*r* = .14 and .23, respectively; Ueno et al., 2019).

Need for clinical cutoff scores. In summary, there is strong support for the ICU as a continuous measure of CU traits. However, a major limitation in the use of the ICU for diagnostic purposes is the absence of well-validated cutoff scores for designating elevated or significant levels of CU traits. One way to make such designations is to use normative cutoff scores. Recently, one source of data for such cutoffs has become available. Kemp and colleagues (2019) combined data from five large-scale, non-referred community datasets from Belgium (Roose, Bijttebier, Decoene, Claes, & Frick, 2010), Cyprus (Fanti, Demetriou, & Kimonis, 2013), Germany (Essau, Sasagawa, & Frick, 2006), and Italy (Baroncelli, Roti, & Ciucci, 2018; Ciucci et al., 2014) to provide norms for the ICU self-report total score (24-item). They provided norms for the self-report version of the ICU based on a total sample of 4,683 boys and girls ages 11-17, with separate data provided for girls ages 11-14 (*n* = 1,475), boys 11-14 (*n* = 1,444), girls 15-17 (*n* = 833), and boys 15-17 (*n* = 931; see Appendix D). Unfortunately, such a large and multi-site normative base has not been established for either of the informant versions (i.e., parent- and teacher-report) of the ICU.

Another method for developing ICU cutoff scores and forming clinically distinct groups has been to use items from the ICU that approximate the four symptoms from the LPE specifier. Specifically, one method established by Kimonis and colleagues (2015) uses a four-item set, derived from Item Response Theory (IRT) analyses of the total ICU scale, to approximate DSM-

5 criteria for the LPE specifier. That is, the following four items, 1) “I care about how well I do at school or work”; 2) “I feel bad or guilty when I do something wrong”; 3) “I do not show my emotions to others”; and 4) “I am concerned about the feelings of others,” are used to approximate each symptom from the LPE specifier. Further, Kimonis and colleagues (2015) compared two scoring methods (i.e., split and extreme) for designating symptom presence. The more stringent “extreme” scoring method consists of accepting only extreme responses (i.e., “definitely true” for positively worded items or “not at all true” for negatively worded items) as an indicator of symptom presence, and extreme scores on any two or more items designates a clinically elevated group for which DSM-5 LPE criteria are said to be met. The less stringent “split” method dictates that positively worded items endorsed as either “very true” or “definitely true” and that negatively worded items scored as either “not at all true” or “somewhat true” meet criteria for symptom presence and that at least two of the four items must be scored as such to meet criteria for clinically elevated CU traits. These authors reported that, in a combined sample of community-based and incarcerated youth, the extreme scoring method provided the largest average weighted effect size (Cohen’s *d*) across external validity tests (ES_w = .83; Kimonis et al., 2015). More specifically, however, the extreme scoring method best discriminated amongst community youth (ES_w = .98), while the split method outperformed extreme scoring in discrimination amongst incarcerated adolescents (ES_w = .47; Kimonis et al., 2015).

A final approach for forming clinically distinct groups has been to use empirically derived cutoff scores from the ICU. For example, one study by Kimonis and colleagues (2014) derived cutoff scores for the ICU parent-report (both mother- and father-report) version using prospective longitudinal data (18 months) in boys and girls. With the use of group-based trajectory modeling, Kimonis et al. (2014) reported that ICU total score cutoffs of 24 for mother-

report and 27 for father-report distinguished groups of youth who were significantly more likely to engage in future self-reported bullying. In addition, Docherty and colleagues (2017) conducted a study that derived cutoff scores for the ICU self-, parent-, and teacher-report versions in a combined sample of community high school students and incarcerated males and girls. With the use of receiver operating characteristic (ROC), area under the curve (AUC), and predicted probabilities, Docherty et al. (2017) found that ICU total score cutoffs of 28 for youth self-report, 30 for parent-report, and 33 for teacher-report were all associated with multi-informant composite scores of aggression and violence as well as detained status; however, parent-report cutoff scores outperformed self- and teacher-report versions when comparing AUC values (AUCs = .75, .63, and .64, respectively).

Important issues in developing cutoff scores. Thus, while there have been several attempts to test methods for using the ICU to form clinically important groups, there has not been a comprehensive test of multiple approaches within the same sample. Such a comprehensive test would allow for comparisons of these cutoff scoring methods and their utility in research and clinical practice. Upon preparing an exhaustive examination of ICU cutoff scores and scoring methods, several issues should be considered.

The first issue to consider is what criteria should be used as validators. That is, the development and test of various ICU cutoff scores will be informative only if theoretically important and empirically supported external validators are used as a test of construct validity (Cronbach, & Meehl, 1955). Furthermore, the external validators need to be valid in and of themselves and uniquely related to the clinical construct that the scale was developed to assess; that is, ICU total scores should be tested for their ability to distinguish groups on important outcomes related to CU traits (Cronbach, & Meehl, 1955). Specifically, in the case of CU traits

and their unique associations with worse behavioral outcomes in youth with serious conduct problems, and based on previous literature establishing the validity of the ICU in high-risk or offender samples, external validators of aggression, antisocial behavior, and juvenile delinquency (including gun use) are particularly important tests of the clinical utility and construct validity of ICU cutoff scores (Cardinale & Marsh, 2017). In addition, due to the associations found between CU traits and worse conduct problems and psychosocial impairment, particularly as it relates to difficulties with interpersonal relations, external validators of informant-reported conduct problems and interpersonal difficulties (i.e., peer rejection and peer-perceived "meanness") are particularly important tests of the ICU's ability to distinguish clinically elevated groups among normative or community youth (Matlasz et al., under review).

A second issue to consider is the assessment of both concurrent and predictive validity (i.e., criterion validity) of cutoff scores. For a scale to provide clinical utility in detecting clinically significant groups of youth with elevated CU traits, it is important that both concurrent and predictive validity are tested. That is, it is important to test whether ICU cutoff scores can identify clinically important groups both based on current levels of behavioral problems or psychosocial functioning (i.e., concurrent validity) and based on their predictive ability to designate youth more likely to show future symptomology related to CU traits (i.e., increased rates of aggressive, antisocial, and delinquent behavior over time; Frick et al., 2014b).

A third issue to consider is the role of method variance. Much of the previous research testing ICU total scores and their associations with relevant behavioral outcomes has been influenced by shared method variance. That is, a majority of studies have evaluated associations between ICU self-report total scores and self-reported outcomes. This may be problematic as research has found associations between both self-reported CU traits and self-reported behavioral

problems and parent-reported CU traits and parent-reported behavioral problems may be inflated by using the same informant for both predictor and outcome (Docherty, Boxer, Huesmann, O'Brien, & Bushman, 2017; Gao & Zhang, 2016). For example, one of the few studies to include all three informant versions of the ICU, conducted by Roose and colleagues (2010), reported that both ICU self-report total scores and composites of ICU parent- and teacher-report total scores were associated with self-reported measures of antisocial behavior, empathy, and prosocial beliefs, however, there was evidence for the effects of shared method variance as the self-reported outcomes showed higher correlations with ICU self-report scores (r = .45,-.55, -.47, respectively; all ps < .01) than with composite informant-report scores (i.e., r = .26, -.31, -.27; all ps < .01).

A final important issue in developing such cutoff scores is the need to weigh sensitivity and specificity. Sensitivity refers to a test's ability to accurately screen in, identify, or diagnose true disorder/disease or symptom presence (i.e., true positive), whereas, specificity refers to a test's ability to accurately screen out or identify individuals who do not have the true disorder/disease or symptom present (i.e., the ability to screen out or reduce false positives; Youngstrom, 2013). With this, it should be noted that cutoff scores that optimize diagnostic sensitivity typically also lead to reduced specificity (Youngstrom, 2013). That is, lower cutoff scores typically lead to increased rates of accurate classification of individuals with a true disorder (i.e., increased sensitivity), but also to increased rates of misclassification of individuals who do not have a true disorder (i.e., false alarms; leading to decreased specificity). Thus, it is critical to consider what the optimal ratio of sensitivity-to-specificity is when choosing cutoff scores.

One statistical method that can be used to evaluate the tradeoffs between sensitivity and specificity for particular cutoff scores is ROC analysis. Specifically, ROC has been found to be an effective and efficient statistical tool that provides information regarding diagnostic sensitivity and specificity for scale cutoff scores, based on a particular dataset and the rates of true positives and false positives observed within it (Youngstrom, 2013). Such analyses are based upon signal detection theory and allow clinicians and researchers the statistical ability to test the cutoff score sensitivity weighted against the probability of falsely detecting a disorder/disease or other condition that does not truly exist (i.e., a false positive), and the inverse of this probability represents the specificity of the diagnostic threshold or cutoff score (Fawcett, 2006; Youngstrom, 2013). More specifically, ROC analysis produces a graphical ROC curve that shows the rate of true positives plotted against false positives at various thresholds or cutoff scores. The area under the ROC curve (i.e., AUC) that results from ROC analysis provides a measurement of the overall model classification accuracy (i.e., the percentage of participants correctly classified across all possible cutoff scores or thresholds; Fawcett, 2006). ROC analysis has been found to be particularly promising and effective because it is a simple statistical test that also provides informative and rigorous visualization of classifier model performance and accuracy (Fawcett, 2006; Youngstrom, 2013). Furthermore, with the use of ROC, Youden's index (i.e., Youden's *J*-statistic = sensitivity + (specificity – 1)) can be calculated and provides an easily comprehensible score that helps to inform or derive optimal cutoff scores (i.e., scores with maximized sensitivity and specificity, represented by Youden's $J \cong 1$).

Statement of the Problem

In summary, prior research has shown that children and adolescents with elevated CU traits represent an etiologically distinct and clinically important subgroup of youth who display

significant behavior problems. This has led to their inclusion as the LPE specifier in the latest editions of major diagnostic and classification systems for psychiatric disorders (i.e., the DSM-5 and ICD-11). Subsequently, there has been an increased focus on enhancing ways to measure these traits for both research and clinical purposes. While the ICU has been proven to be a reliable and valid measure of these traits continuously, little work has focused on developing cutoff scores for its use in detecting elevated levels and potentially helping to make diagnostic decisions.

Thus, the present study examined four distinct methods for determining cutoff scores for the ICU. The first method consisted of statistically deriving optimal cutoffs (i.e., optimized both for diagnostic sensitivity and specificity) via an empirical approach using ROC and Youden's index. The second approach developed cutoffs by using norms provided by two separate normative distributions, and the third and fourth methods used the four-item DSM-5 LPE criteria approximation, as developed and proposed by Kimonis et al. (2015), by using extreme and split scoring methods. The clinical utility (i.e., concurrent and predictive validity) of these four separate cutoff types (i.e., empirical cutoffs, normative cutoffs, and four-item approximations using extreme and split scoring) were then tested and compared. Furthermore, this study tested cutoff scores for youth self-, parent-, and teacher-report versions of the ICU and included clinically important validators that were measured in ways to reduce the effects of shared method variance. Finally, the present study tested these cutoff scores using both longitudinal and cross-sectional methods in two samples that differ on a number of important characteristics. Specifically, the first sample included adolescent boys (ages 13-17) followed prospectively for 5 years after their first official arrest, and the second sample included boys and girls, ages 8-15, recruited from a school system in the southeastern United States for a cross-sectional study.

Study Aims

In summary, the present study tested the following aims:

1. With the use of two distinctive samples (i.e., high-risk juvenile offenders and representative community youth) and clinically relevant validators (e.g., self-reported aggression and offending, informant-reported conduct problems, peer rejection), optimal total score cutoffs for each version of the ICU (i.e., self-, parent-, and teacher-report) were derived with the use of ROC and Youden's index scores (i.e., Youden's $J \cong 1$ represents maximized sensitivity and specificity). Furthermore, classification accuracy and overall model performance (as indicated by AUC values) of each ICU version, within the respective sample's outcomes data, were assessed and reported.
2. Within each sample and for each informant version of the ICU, four separate sets of cutoff scores and classification methods were established. That is, groups classified as either high or low on CU traits were established by 1) empirically derived total score cutoffs from aim one; 2) age- and gender-based total score cutoffs using normative distributions; 3) four-item LPE approximation using more stringent "extreme" scoring; and 4) four-item LPE approximation using less stringent "split" scoring.
3. Finally, the criterion (i.e., predictive and concurrent) validity of these cutoff scores and classification methods were compared with the use of meta-analytic methods (i.e., average effect sizes, *r*). Furthermore, interpretations regarding the comparative clinical utility of these cutoff scores and scoring methods are provided along with proposed ranges of optimal total score cutoffs across each ICU informant version.

Methods

Participants

Sample 1: high-risk sample of first-time juvenile offenders. Participants were 1,216 male adolescents who were first-time juvenile offenders. These participants were recruited as part of the Crossroads study, an ongoing, multi-site study with a prospective longitudinal design. Participants were recruited from three regions in the United States, including Jefferson Parish, Louisiana (n = 151), Orange County, California (n = 532), and Philadelphia, Pennsylvania (n = 533), and were considered eligible if they were male, English-speaking, between the ages of 13 and 17 years (M = 15.29, SD = 1.29), and recently arrested for an eligible (i.e., midrange) offense of low-to-moderate severity, such as theft of goods or simple battery (see Ray et al., 2017 for more information regarding eligibility criteria). At baseline, self-reported race and ethnicity data showed the sample was primarily Hispanic (45.8%), followed by African American (36.9%), Caucasian (14.8%), and self-identified other (2.5%). At baseline, the highest level of education either parent obtained included 27.2% with less than a high school degree, 34.1% with a GED or high school diploma, 20.4% with trade school or some college, 13.5% with a 4-year college degree, and 4.8% with a graduate level education. Follow-up data was collected in multiple waves over the 5 years following initial arrest and study enrollment.

Sample 2: normative sample of community youth. Participants were 289 boys and girls, ages 8 to 15 years (M = 11.47, SD = 2.26), who were recruited as part of a cross-sectional study (Matlasz et al., under review) from two school systems in the southeastern United States. Participants were in the 3rd (n = 93), 6th (n = 69), and 8th (n = 127) grades at the time of data collection. The sample was comprised 59.9% of girls, and, according to parent-report, was primarily Black, Afro-Caribbean, or African American (40.1%), followed by Caucasian (35.3%),

Biracial (12.1%), Hispanic (4.5%), and other ethnic minorities (2.4% East Asian or Asian American, 0.7% Middle Eastern or Arab American, 0.7% Native American or Alaskan Native, and 0.7% self-identified other), while 3.5% of the sample chose not to report on race or ethnicity. The majority (86.2%) of participants' parents had an education level less than a college diploma, and the majority (71.6%) of the sample's reported total household annual income was less than $40,000.

Procedures

Sample 1. The Institutional Review Boards at Crossroads study site institutions (i.e., University of California, Irvine, Temple University, and Louisiana State University) approved the study procedures. Parental informed consent and youth assent were obtained for all participants at each timepoint, before interviews were conducted. After youth participants turned 18 years old, only youth consent was obtained. Participants and their parents were informed that participation was completely voluntary, would not influence the youth's relationship with the juvenile justice system or court, and that they were able to withdraw from the study at any time without penalty. Further, the participants and their parents were informed that the research study had obtained a Privacy Certificate from the Department of Justice, which protected their data from being subpoenaed for use in legal proceedings. Crossroads study interviewers were trained extensively on the study design, safety procedures, participant recruitment and retention, consent/assent procedures, maintenance of confidentiality, and interview administration and rapport building. Prior to being authorized to conduct interviews independently, interviewers were required to pass tests on the study training, observe two interviews with the site study coordinator, and, finally, successfully complete a check-out interview supervised by the study coordinator.

Youth were enrolled in the study and completed the baseline interview within six weeks of the disposition date for their first arrest. Participants were then reassessed at follow-up timepoints that occurred twice per year (every 6 months) during the first 3 years and then once per year during years 4-5, providing eight total waves of follow-up data. Participants were able to select their preferred location to complete the interviews, often at the youth's home, a local restaurant, public library, or in a secure facility if the youth was incarcerated at the time of the follow-up interview. Finally, if participants moved too far to conduct in-person interviews, phone interviews were completed. Interviews lasted, on average, approximately 2-3 hours and were administered using a secure computer-based program on a laptop. To control for differences in participants' reading ability, interviewers read all items aloud. Participants were compensated $50 for the baseline interview and payments increased by $15 for each subsequent interview, up to a total of $140 at year 4.

Sample 2. Study approvals were received from the Institutional Review Board at Louisiana State University, the superintendent of the two schools systems, and the principals of the participating schools. Teachers in the 3rd, 6th, and 8th grades were approached to help with youth participant recruitment and study data collection. More specifically, participating teachers agreed to send a description of the study home with the children in their class, along with parental consent forms and parent-report measures that took no more than an average of 5 minutes to complete. For all participating children who returned signed parental consent forms, teachers were given measures to complete, which took an average of 5 minutes per child. Lastly, upon receiving parental consent forms, study coordinators visited classrooms to obtain assent from participating youth. All child self-report measures were administered during the school day, in a group setting, and on school computers. Together, all measures took an average of 45

minutes, or about the equivalent of one class period, to complete. To compensate teachers for their time and effort, and to encourage participation generally, schools were compensated $10 per participating child to go toward the purchase of classroom supplies.

Measures

Callous-unemotional traits. CU traits were assessed with the ICU total scale in both samples. In the first sample of high-risk youth, the ICU youth self-report version was collected during baseline interviews and showed acceptable internal consistency (M = 26.28, SD = 8.08; Cronbach's α = .76). In the second sample of community youth, ICU self- and informant-report versions (i.e., parent- and teacher-report) were administered. In this sample, ICU self-report showed acceptable internal consistency (M = 18.93, SD = 8.04; Cronbach's α = .76); parent-report showed good internal consistency (M = 17.26, SD = 9.92; Cronbach's α = .86); and teacher-report showed excellent internal consistency (M = 21.31, SD = 13.54; Cronbach's α = .93). In both samples, participants missing more than two-thirds of the total ICU items were removed (i.e., listwise deletion) from analyses; for those with more than two-thirds of items present, mean-item proration was utilized to address item-level missing data.

Sample 1: Outcome variables. In sample one, self-reported aggression, self-reported gun use (carrying and use during a crime), and juvenile delinquency (self-reported offending and official arrests) were used as external validators and assessed prospectively over time.

Self-reported aggression. Levels of self-reported aggression were measured at all follow-up timepoints with the Peer Conflict Scale (PCS; Marsee et al., 2011; Marsee et al., 2014), a 40-item self-report scale that was designed to comprehensively assess the different forms and functions of aggressive behavior. The 20 PCS-items for physical aggression, both reactive and proactive (10 items each), were used. Items were rated on a four-point Likert-type scale from 0

(*Not at all true*) to 3 (*Definitely true*) and summed to form continuous total scores. Across 8 follow-up time points, the PCS physical aggression score showed good to excellent internal consistency (Cronbach's α = .85-.90), the PCS physical proactive subscale showed acceptable to good internal consistency (Cronbach's α = .72-.83), and the PCS physical reactive subscale showed good internal consistency (Cronbach's α = .83-.86). For ROC analyses, a cutoff z-score of 1.5 was applied to form dichotomized variables of high versus low physical aggression (both for total and subscale scores).

Self-reported delinquency. Levels of self-reported delinquency were assessed at all follow-up timepoints with the 24-item revised version of the Self-Report of Offending Scale (SRO; Huizinga, Esbensen, & Weiher, 1991) which assessed drug and property offenses and crimes against persons. Drug offense items included: "driven while drunk or high"; "sold marijuana"; and "sold other drug." Property offense items included: "destroy property"; "set fire"; "broke in to steal something"; "shoplift"; "receive stolen property"; "use credit card illegally"; "stole car"; and "enter car to steal something." Finally, items describing crimes against persons included: "killed someone"; "forced someone to have sex with you"; "shot someone (where the bullet hit the victim)"; "shot at someone (where you pulled the trigger)"; "taken something from another person by force, using a weapon"; "taken something from another person by force, without a weapon"; "beaten up or physically attacked someone so badly that they probably needed a doctor"; "beaten up, threatened, or physically attacked someone as part of a gang"; and "been in a fight." For each item, participants were asked (yes or no) if, in the last 6 months, they engaged in each crime and, if yes, how many times. The SRO variety score was calculated and used to assess the number of different offenses participants endorsed over each assessment period. This method is preferred over a frequency score as it is less prone to

recall errors, especially when the offense (e.g., fighting) is committed frequently (Hindelang, Hirschi, & Weis, 1981; Thornberry & Krohn, 2000). Further, frequency scores are often weighted toward less severe offenses, like fighting, that occur more frequently (Thornberry & Krohn, 2000). Higher scores thus represent a greater variety of crimes committed and have been shown to correlate with official arrest records (Monahan & Piquero, 2009). Total offending and violent offending (i.e., crimes against persons) variety scores were calculated. Across follow-up, the total variety score showed acceptable to good internal consistency (Cronbach's α = .75-.83), while the violent offending variety score showed poor to questionable internal consistency (Cronbach's α = .51-.64), however, due to low base rates of endorsements on this subscale, mean inter-item correlations are also reported ($M_{correlation}$ = 0.16-0.29). For ROC analyses, a cutoff z-score of 1.5 was applied to form dichotomized variables of high versus low total self-reported offending and violent self-reported offending.

Self-reported gun use. Self-reported gun carrying was assessed using gun-related items of the SRO at each of the eight follow-up timepoints. Specifically, each item asked participants (yes or no) if they carried a gun at any point since the last interview and, if yes, how many times. A total gun carrying variable was created for the purpose of this study by summing the number of times participants endorsed carrying a gun across all follow-up time points. For ROC analyses, and due to a relatively low base rate (n = 221; 19.3%), a dichotomized variable of any endorsed gun carrying over the entire follow-up (i.e., yes = 1; no = 0) was used.

Self-reported gun use during the commission of a crime was also assessed using items from the SRO at each follow-up timepoint. Specifically, participants were asked (yes or no) if, since the last interview, they had carjacked someone, shot someone, shot at someone, committed armed robbery, committed acts of gang violence, or killed someone. If participants endorsed

engagement in any of these offenses, they were then asked if they used a gun (i.e., yes = 1; no = 0) during the commission of that crime. Due to a low base rate of gun use during crime (*n* = 102; 8.9%), a dichotomous variable was used to differentiate participants who endorsed gun use during a crime at any point during follow-up from those who did not (i.e., yes = 1; no = 0) in all analyses.

Official arrests. In order to avoid issues due to shared method variance, an additional measure including official arrest records was included. Official rearrest data was obtained for both juvenile and adult arrests within the jurisdictions in which the participants were initially arrested. Only new charges during the follow-up periods were included (i.e., probation and technical violations were excluded). Official rearrest data for all crimes as well as violent crime (i.e., crimes against persons) in particular were assessed. Due to a low base rate of rearrest for violent crime (*n* = 217; 18.3%), a dichotomous variable (i.e., at least one rearrest for violent crime at any point during follow-up = 1; no violent rearrests = 0) was used in all analyses. For ROC analyses in total rearrest data, a dichotomous variable was created such that any participant with two or more rearrests (*n* = 330; 27.9%) were coded as 1 and those with one or fewer were coded as 0.

Sample 2: Outcome variables. In sample two, informant-reported conduct problems, peer-rejection (i.e., social preference), and peer-nominated levels of "meanness" were used as external validators and measured cross-sectionally via multiple informant reports (i.e., parent- and teacher-reports, peer-nominations).

Informant-reported conduct problems. Levels of informant-reported conduct problems were measured with the Disruptive Behavior Disorders Scale (DBD; Pelham, Gnagy, Greenslade, & Milich, 1992), a 39-item measure of symptoms consistent with DSM criteria for

ADHD, ODD, and CD diagnoses. Items responses are on a 4-point Likert-type scale from 1 (*Not at all*) to 4 (*Very much*). For the purposes of this study, only items from the ODD and CD subscales were included in analyses, and ratings from the DBD have been shown to correlate with diagnoses of ODD and CD based on structured interviews (Waschbusch, Walsh, Andrade, King, & Carrey; 2007). The DBD was completed by both parent and teacher, and both showed excellent internal consistency (parent-report Cronbach's α = .95; teacher-report Cronbach's α = .96) and were correlated with one another (r = .29, p < .001). With this and based on the recommendation of Piacentini and colleagues (1992), the highest rating on each item was taken to yield a resolved score for each item. These resolved scores were then summed and averaged to create a composite score of conduct problems that also showed excellent internal consistency (Cronbach's α = .93). For ROC analyses, criteria set forth by Pelham et al. (1992) to approximate diagnostic criteria were used to create a dichotomous variable of high versus low conduct problems. Based on these criteria, 50 youth (17.3%) met criteria for either elevated ODD or CD symptoms, or both.

Peer rejection and social preference. Levels of peer rejection and social preference were measured with two sociometric items in the form of peer-nominations that asked, "who do you like the most?" and "who do you like the least?". Nominations for both of these sociometric items were converted to proportion scores, which controls for differences in the number of students and potential nominations in each school and grade (Cillessen, 2009). These scores were then standardized within grade and "liked least" scores were subtracted from "liked most" scores to yield a total peer social preference score (Coie & Dodge, 1998). This social preference score is the most common method for assessing peer rejection (Newcomb, Bukowski, & Pattee, 1993), and past research has shown such social status scores to be positively related to measures of

aggression and negatively related to prosocial behaviors towards peers (Dodge, Coie, & Brakke, 1982). For ROC analyses, participants were classified as being rejected by their peers (i.e., rejected by peers = 1; not rejected = 0) if their social preference score z-score was less than −1, liked-most z-score was less than 0, and liked-least z-score was greater than 0 (Matlasz et al., under review). Based on these criteria, 28 youth (9.7%) were found to be rejected by their peers.

Peer-nominated "meanness." Levels of peer-reported "meanness" were assessed with peer nominations that asked participants to name: "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?" (Matlasz et al., under review). Participating youth were given unlimited nominations and were instructed to write in the names of classmates in their same grade who they felt best fit each description. These scores were again converted to proportion scores and standardized within grade. For ROC analyses, a cutoff z-score of 1.5 was used to form a dichotomous variable of high versus low peer-nominated meanness.

Data Analytic Plan

All analyses were conducted using SPSS Statistics version 25. There was minimal missing data across both samples. In the first sample, no participants were missing ICU self-report data at baseline, 69 (5.7%) were missing outcomes data for self-reported aggression, self-reported offending, and self-reported gun use, and 31 (2.5%) were missing follow-up arrest record data and were thusly excluded from analyses. In the second sample, 28 participants (9.7%) were missing two-thirds or more of ICU self-report data, and thus, mean proration was not appropriate, while seven participants (2.4%) were missing ICU parent-report and 18 (6.2%) were missing ICU teacher-report. In this sample, there were no missing outcomes data for informant-reported conduct problems, peer rejection, and peer-nominated meanness.

Preliminary data analyses. First, zero-order correlations were performed to test associations between raw ICU total scores and respective outcomes in each sample. Specifically, in the first sample, correlations were conducted between ICU self-report total scores and outcomes of self-reported aggression, self-reported offending, gun use, and official arrests. In the second sample, correlations were conducted between ICU self-, parent-, and teacher-report total scores and outcomes of informant-reported conduct problems, peer rejection, and peer-nominated "meanness." These initial tests were used to indicate the extent to which the present study's ICU scores were related to their respective outcomes.

Primary data analyses: Aim 1. The first study aim was to form optimal cutoff scores (i.e., those with maximized sensitivity and specificity) for each ICU version (i.e., self-, parent-, and teacher-report total scores) within each sample through empirical methods. To accomplish this, ROC analyses were performed using dichotomized versions of each outcome variable (described above) within each sample, and resulting measures of sensitivity and specificity were used to calculate Youden's index scores (i.e., J = sensitivity + (specificity – 1)) for all possible ICU cutoffs observed in the sample. For each sample and respective dichotomized outcome, the ICU total score with the largest Youden's index (i.e., $J \cong 1$) was selected as the optimal cutoff score. Based on these analyses, the average of these cutoff scores, across all outcomes, was selected as the empirically derived cutoff score for each version of the ICU and each sample. Furthermore, the classification accuracies (i.e., AUCs) of ICU versions across dichotomized outcomes were reported.

Aim 2. The second study aim was to classify participants in each sample as having either elevated or non-elevated CU traits based on four distinct cutoff methods and scores for the ICU: 1) empirically derived optimal total score cutoffs using methods described in aim one; 2) total

score cutoffs based on age- and gender-specific norms (i.e., total scores at the 95th percentile); 3) cutoffs formed from Kimonis and colleagues' (2015) four-item DSM-5 LPE criteria approximation using extreme scoring; and 4) Kimonis et al.'s (2015) four-item LPE approximation using split scoring.

Using normative data to form ICU cutoffs and clinically distinct groups. In order to assess the criterion validity and clinical utility of normative cutoff scores, separate ICU self-, parent-, and teacher-report norms were applied to each sample. Participants in the first sample were classified as having clinically elevated CU traits based on the following age- and gender-based multinational normative cutoff scores at the 95th percentile: ICU self-report total score ≥ 37 for boys ages 11-14 and ≥ 41 for boys ages 15-17 (Kemp et al., 2019; see Appendix D).

In the second sample, the same age- and gender-based multinational norms provided by Kemp et al. (2019) were applied to form cutoff scores at the 95th percentile for ICU self-report, that is: ICU self-report total score ≥ 37 for boys ages 14 and younger; ≥ 41 for boys age 15; ≥ 32 for girls ages 14 and younger; and ≥ 35 for girls age 15. In addition, gender-based local normative cutoffs using the 95th percentile were used: ICU parent-report total score ≥ 38 for boys and ≥ 33 for girls (see Appendix E) and ICU teacher-report total score ≥ 50 for boys and ≥ 38 for girls (see Appendix F).

Using the four-item DSM-5 LPE criteria approximation to form ICU cutoffs and clinically distinct groups. In order to assess the criterion validity and clinical utility of an alternative, brief scoring method that uses a four-item LPE approximation, participants were classified in both samples as having either elevated or non-elevated CU traits based on responses (i.e., self-, parent-, and teacher-reported) to the four items proposed by Kimonis and colleagues (i.e., ICU items 3, 5, 6, and 8; 2015). For the first, "extreme," scoring method, only participants

with extreme ratings (i.e., responses of "definitely true" endorsed for positively worded items and "not at all true" for negatively worded items) on two or more of the four items met approximated DSM-5 LPE criteria and were thusly classified as having elevated CU traits. Finally, for the "split" scoring method, participants with responses of either "very true" or "definitely true" (for positively worded items) and "somewhat true" or "not at all true" (for negatively worded items) on two or more items met approximated LPE criteria and were classified as having elevated CU traits.

Finally, in both samples, intercorrelations among the four cutoff methods (i.e., associations between binary grouping variables) were tested with the phi coefficient. In addition, the degree of classification overlap, or percentage agreement, by these various scoring methods were reported.

Aim 3. The final study aim was to compare the criterion validity (i.e., predictive in sample one and concurrent in sample two) resulting from each ICU cutoff method (i.e., optimal cutoffs derived from ROC analysis, age- and gender-based normative cutoffs, and the four-item LPE approximation using both extreme and split scoring) in each sample. To accomplish this aim, group-means difference tests in the forms of chi-square and independent-samples *T*-tests were performed across ICU cutoff methods and outcomes. Finally, comparisons of these cutoff methods' overall validity and clinical utility were conducted with the use of meta-analytic methods by converting these inferential statistics into a common effect size (i.e., *r)* and then examining average effect sizes across outcomes for each cutoff method in each sample.

Results

Preliminary Analyses

The distributions of main study variables are described in Table 1. In addition, all zero-order correlations between predictor ICU variables (self-, parent-, and teacher-report) and continuous outcome variables in each sample are reported in Table 1. Further, because all outcome variables were significantly correlated with ICU total scores and in the directions expected (e.g., positive associations with measures of juvenile delinquency and informant-reported conduct problems and a negative association with social preference score, indicating greater peer rejection), all were kept and included in primary analyses.

Test of Main Study Aims

Aim 1: computation of optimal cutoff scores. To accomplish this first aim, a series of ROC analyses were performed in each sample, with total scores for each ICU version coded as the test probability variable and each dichotomized version of outcomes coded as the state variable, to produce ROC curves, estimates of models' overall classification accuracy, and sensitivity-specificity metrics.

Sample 1. Across all ROC analyses for each outcome variable, ICU self-report total scores at baseline performed significantly better than chance at distinguishing participants based on dichotomized outcomes (see Figure 1 for all ROC curves). Specifically, the ICU performed significantly better than chance at distinguishing participants with high versus low levels of total physical aggression (AUC = 0.77, *SE* = 0.03, *z* = 10.80, 95% CI = 0.72, 0.82), proactive physical aggression (AUC = 0.74, *SE* = 0.03, *z* = 7.53, 95% CI = 0.68, 0.80), and reactive physical aggression (AUC = 0.73, *SE* = 0.03, *z* = 8.96, 95% CI = 0.68, 0.78). Furthermore, Youden's index scores of 0.41, 0.38, and .36 indicated optimal cutoff scores of 30, 34, and 28, respectively,

for distinguishing participants with high versus low total aggression, proactive aggression, and reactive aggression, respectively.

Second, the ICU performed significantly better than chance at distinguishing participants on high versus low total self-reported offending (AUC = 0.73, *SE* = 0.03, *z* = 9.12, 95% CI = 0.68, 0.78) and violent self-reported offending (AUC = 0.62, *SE* = 0.02, *z* = 7.75, 95% CI = 0.59, 0.66). Youden's index scores of 0.37 and 0.19 indicated optimal cutoff scores of 28 and 24, respectively, for distinguishing high versus low total offending and violent offending. The ICU performed significantly better than chance at distinguishing participants who reported any gun carrying over follow-up from those who did not (AUC = 0.64, *SE* = 0.02, *z* = 6.43, 95% CI = 0.59, 0.68) and those who reported any gun use during crime over follow-up from those who did not (AUC = 0.70, *SE* = 0.03, *z* = 7.41, 95% CI = 0.65, 0.75). Further, Youden's index scores of 0.20 and 0.29 indicated optimal cutoff scores of 30 and 28, respectively, for distinguishing participants who report any gun carrying and any gun use during crime over follow-up.

Finally, the ICU performed significantly better than chance at distinguishing participants with two or more rearrests from those with one or none (AUC = 0.58, *SE* = 0.02, *z* = 4.33, 95% CI = 0.54, 0.61) and those with any violent rearrests from those with none (AUC = 0.55, *SE* = 0.02, *z* = 2.32, 95% CI = 0.51, 0.59). Further, Youden's index scores of 0.13 and 0.08 indicated optimal cutoff scores of 27 and 33, respectively, for distinguishing participants with more rearrests and those with any violent rearrests. As a result of all of these analyses, an average cutoff score of 29 was used as the empirically derived optimal cutoff in this sample.

Table 1.

Main Study Variables: Distributions and Zero-Order Correlations

Sample One: High-Risk Juvenile Offenders

Variables	Mean (*SD*)	Range	Skewness	Kurtosis	Correlation with ICU-SR
ICU-SR	26.28 (8.08)	0 – 55	0.07	0.08	–
Total physical aggression	4.57 (4.94)	0 – 40	2.30	7.66	0.35**
Proactive aggression	0.98 (1.70)	0 – 17	3.78	19.51	0.31**
Reactive aggression	3.59 (3.54)	0 – 23	1.68	3.68	0.34**
Total SR offending	8.02 (11.44)	0 – 85	2.85	10.58	0.31**
Violent SR offending	3.06 (4.30)	0 – 33	2.69	9.37	0.28**
Gun carrying	0.43 (1.12)	0 – 8	3.31	11.86	0.21**
Gun use during crime	0.26 (1.10)	0 – 13	6.32	48.99	0.18**
Total rearrests	1.24 (2.02)	0 – 17	2.55	9.19	0.14**
Violent rearrests	0.29 (0.77)	0 – 11	5.19	46.88	0.07*

Sample Two: Normative Community Youth

Variables	Mean (*SD*)	Range	Skewness	Kurtosis	Correlation with ICU-SR	Correlation with ICU-PR	Correlation with ICU-TR
ICU-SR	18.93 (8.04)	3 – 50	0.71	0.71	–	–	–
ICU-PR	17.26 (9.92)	1 – 47	0.76	0.12	0.24**	–	–
ICU-TR	21.31 (13.54)	0 – 61	0.51	-0.50	0.29**	0.24**	–
Conduct problems	25.65 (8.71)	18 – 67	1.77	3.39	0.29**	0.49**	0.52**
Social preference	0.03 (1.47)	-6 – 4	-0.36	1.45	-0.26**	-0.15*	-0.24**
Peer-nominated meanness	0.002 (0.72)	-1 – 5	2.95	12.60	0.33**	0.17**	0.21**

Note. ICU = Inventory of Callous-Unemotional Traits; SR = self-report; PR = parent-report; TR = teacher-report; SD = standard dev
Correlations reported with Pearson's *r*. Self-reported offending reported as frequency scores across follow-up; gun use and rearrest r
as frequency scores (i.e., number of times endorsed or documented across follow-up). Peer rejection represented by social preferenc

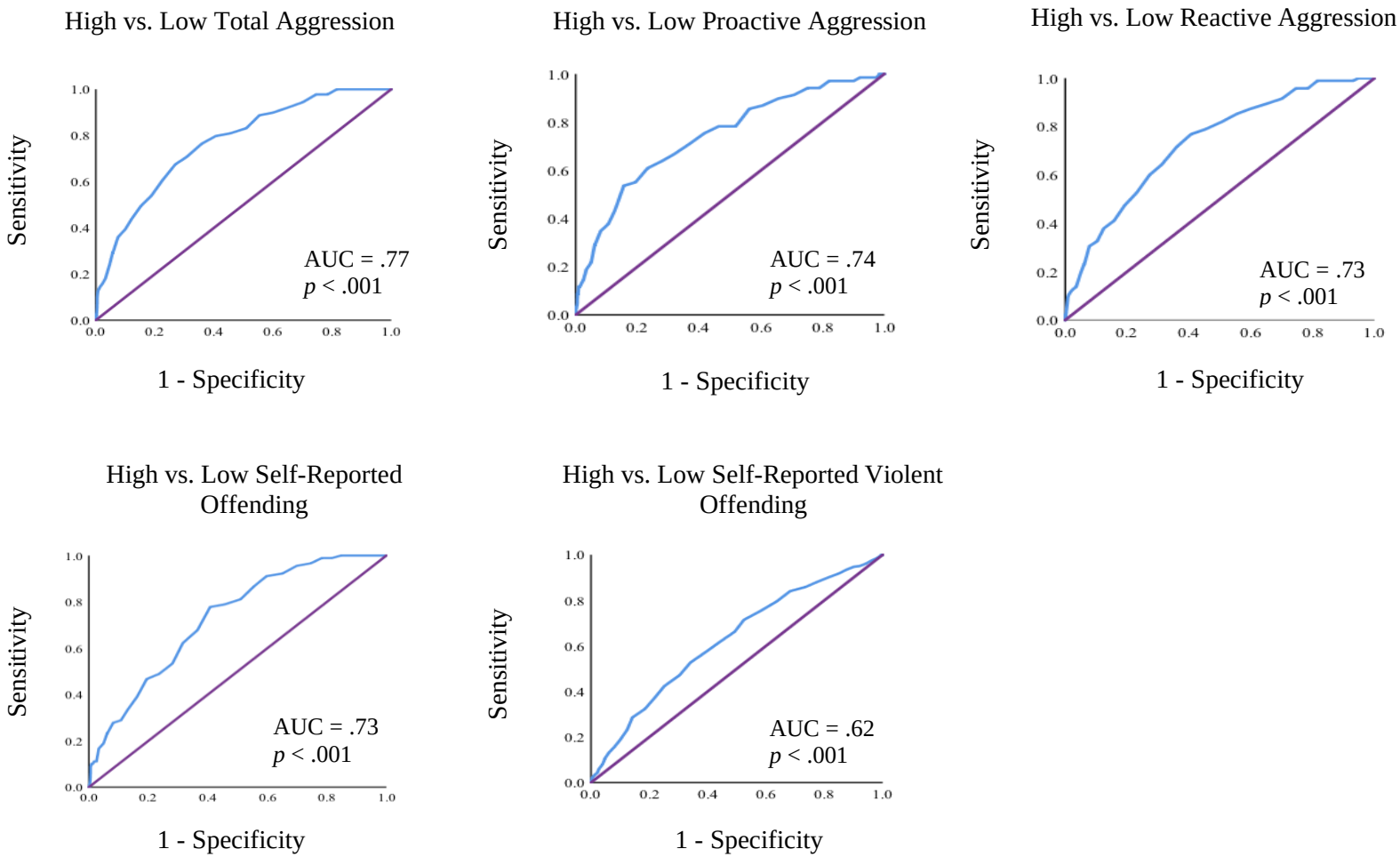
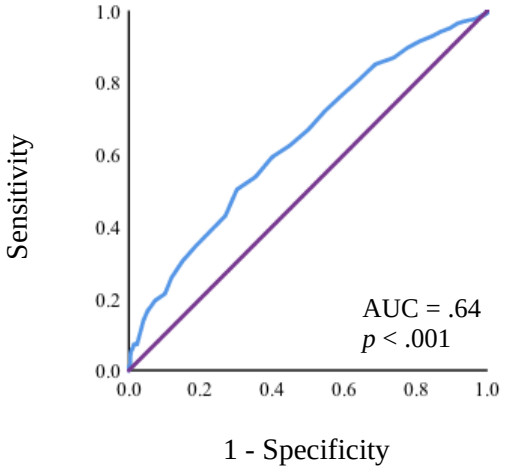


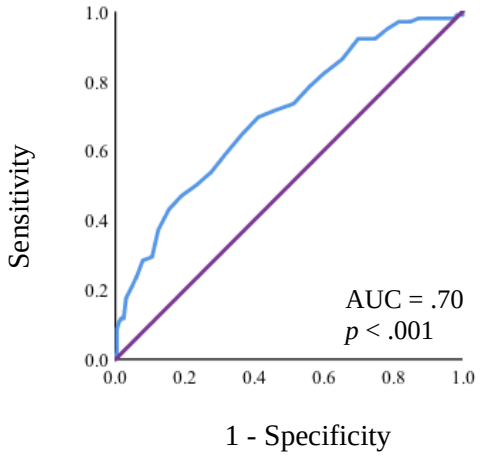
Figure 1. *ROC Plots of Baseline ICU Self-Report Total Scores and Dichotomized Outcomes in Sample One.*

Figure cont'd.

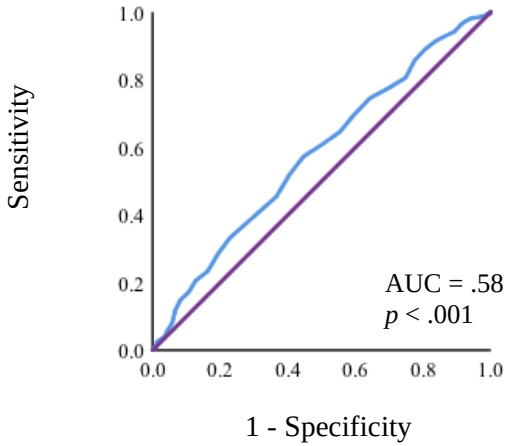
Any Gun Carrying



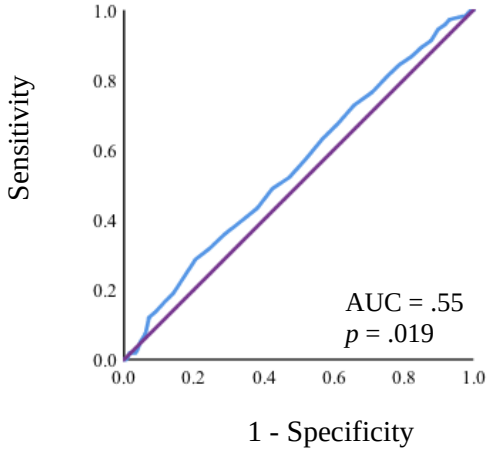
Any Gun Use During Crime



High vs. Low Rearrests



Any Violent Rearrests



Sample 2. Across all ROC analyses, total scores for all ICU versions (i.e., self-, parent-, and teacher-report) did not perform significantly better than chance at distinguishing participants who were peer-rejected from those who were not (see Figures 2-4 for ROC curves). As such, Youden's index and optimal cutoff scores for the outcome of peer rejection were not computed or factored into computations of average optimal cutoff scores.

First, ICU self-report scores performed significantly better than chance at distinguishing participants with high versus low informant-reported conduct problems (AUC = 0.64, *SE* = 0.05, *z* = 2.88, 95% CI = 0.55, 0.74) and participants with high versus low levels of peer-nominated "meanness" (AUC = 0.67, *SE* = 0.08, *z* = 2.06, 95% CI = 0.51, 0.83). Further, Youden's index scores of 0.25 and 0.31 indicated optimal cutoff scores of 22 and 26, respectively, for distinguishing participants with significant conduct problems and high levels of peer-nominated meanness. As a result of these analyses, an average cutoff score of 24 was used as the empirically derived optimal self-report cutoff in this sample.

Second, ICU parent-report scores performed significantly better than chance at distinguishing participants with high versus low conduct problems (AUC = 0.70, *SE* = 0.05, *z* = 4.32, 95% CI = 0.61, 0.79) and participants with high versus low levels of peer-nominated "meanness" (AUC = 0.64, *SE* = 0.07, *z* = 2.18, 95% CI = 0.51, 0.77). Further, Youden's index scores of 0.41 and 0.28 both indicated an optimal cutoff score of 23 for distinguishing participants with significant conduct problems and high levels of peer-nominated meanness. Thus, a cutoff score of 23 was used as the empirically derived optimal parent-report cutoff in this sample.

Finally, ICU teacher-report scores performed significantly better than chance at distinguishing participants with high versus low conduct problems (AUC = 0.80, *SE* = 0.04, *z* =

7.24, 95% CI = 0.72, 0.88) and participants with high versus low levels of peer-nominated "meanness" (AUC = 0.64, *SE* = 0.09, *z* = 1.59, 95% CI = 0.47, 0.82). Further, Youden's index scores of 0.57 and 0.47 indicated optimal cutoff scores of 32 and 37, respectively, for distinguishing participants with significant conduct problems and high levels of peer-nominated meanness. As a result of these analyses, an average cutoff score of 35 was used as the empirically derived optimal teacher-report cutoff in this sample.

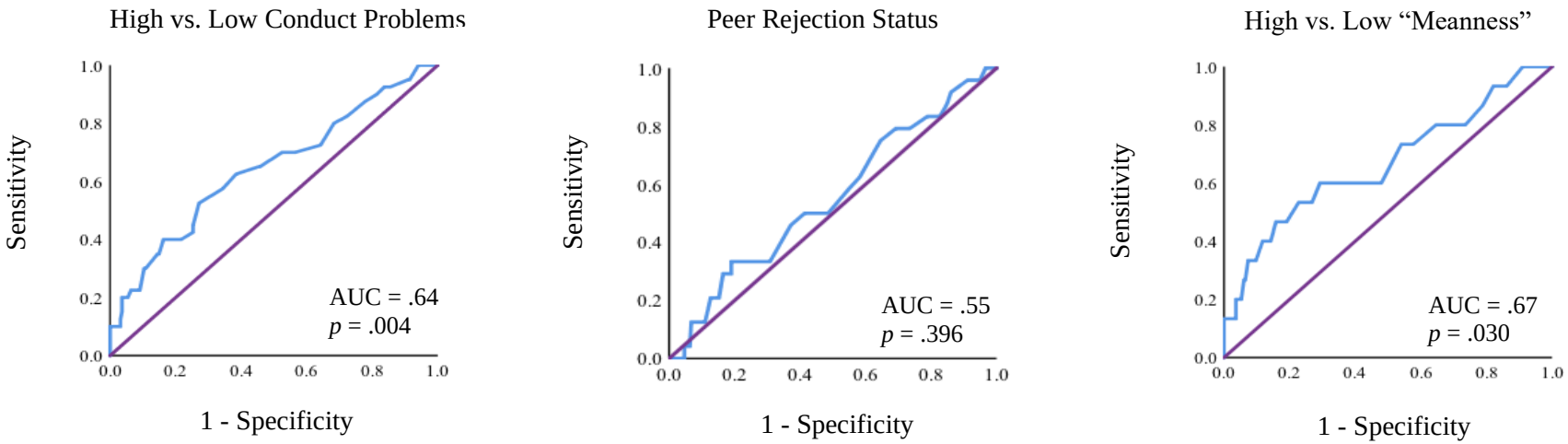


Figure 2. *ROC Plots of ICU Self-Report Total Scores and Dichotomized Outcomes in Sample Two.*

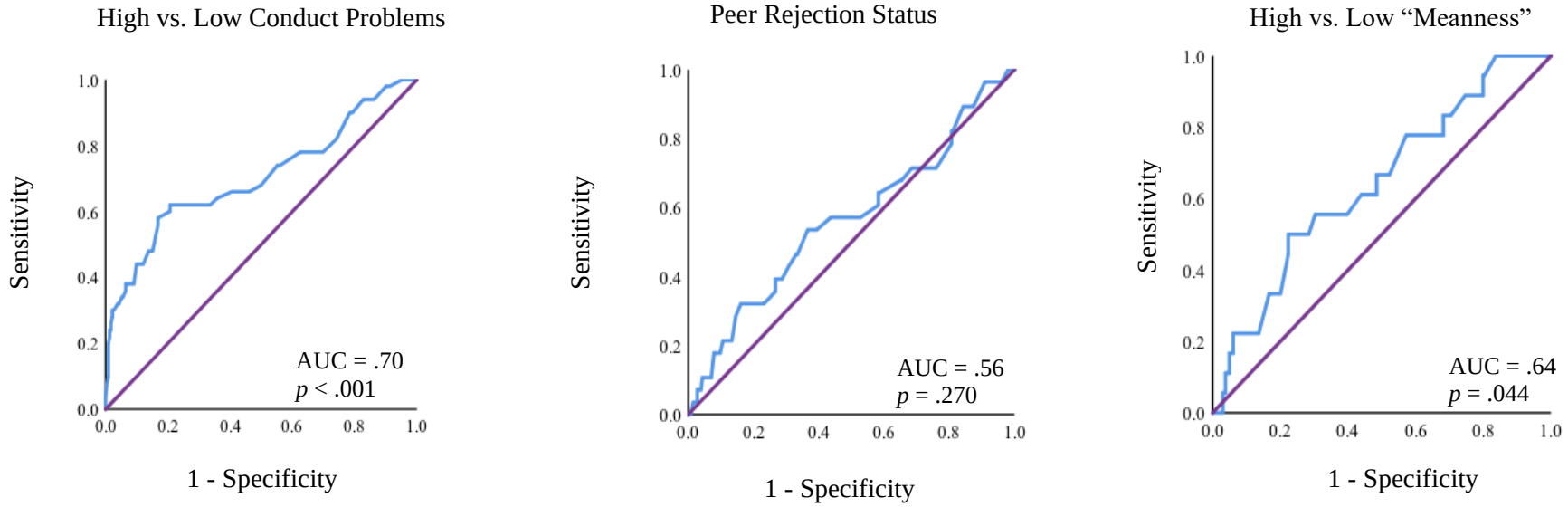


Figure 3. *ROC Plots of ICU Parent-Report Total Scores and Dichotomized Outcomes in Sample Two.*

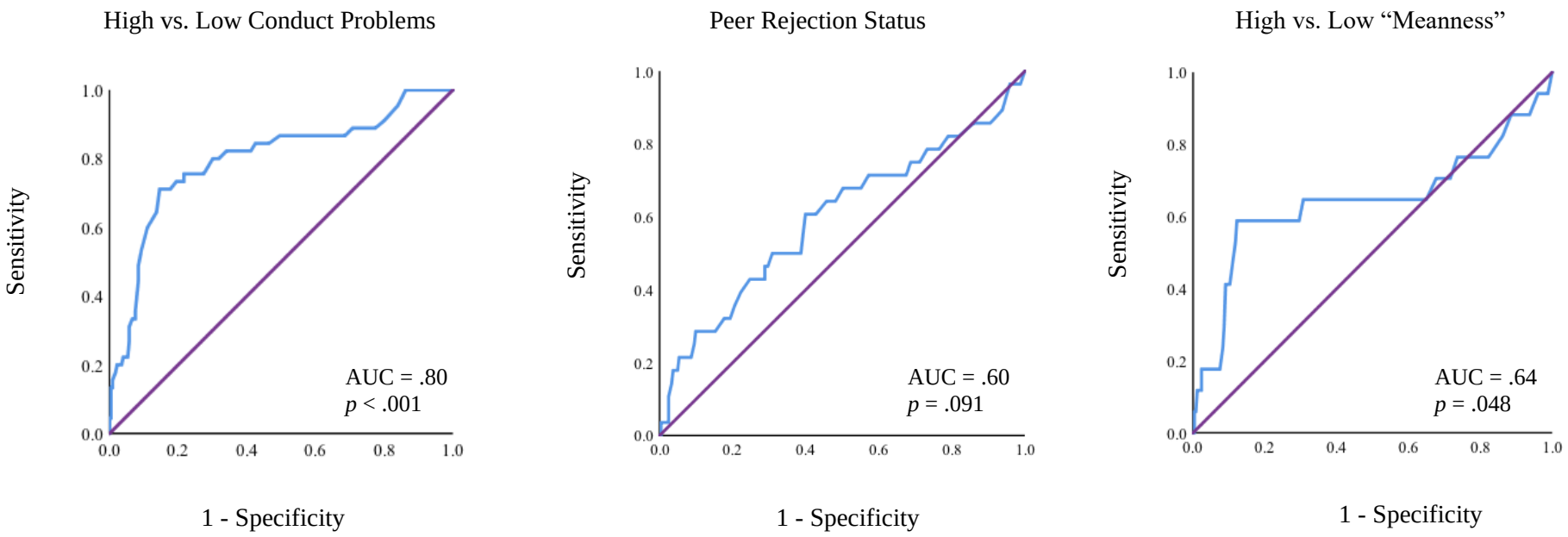


Figure 4. *ROC Plots of ICU Teacher-Report Total Scores and Dichotomized Outcomes in Sample Two.*

Aim 2: classification of participants as either high or low on CU traits. Participants in both samples were classified as having either elevated or non-elevated CU traits based on four methods for forming cutoff scores: 1) empirically derived cutoffs using ROC analyses as part of aim one; 2) normative cutoffs at the 95th percentile provided by age- and gender-based multinational norms (sample one) and age- and gender-based multinational norms or gender-based local norms (sample two); 3) a four-item LPE criteria approximation using "extreme" scoring; 4) a four-item LPE criteria approximation using "split" scoring. All classification methods, in both samples, were significantly correlated with one another (see Tables 2-3).

Sample 1. In the first sample of high-risk male adolescent offenders, the first cutoff method, an empirically derived optimal cutoff score of 29, classified 39% (n = 479) of participants as having elevated CU traits. The second cutoff method, age- and gender-based normative cutoff scores at the 95th percentile, classified 5% (n = 64) of participants as having elevated CU traits. The third cutoff method, four-item LPE approximation with extreme scoring, classified 7% (n = 86) of participants as having elevated CU traits. Finally, the fourth cutoff method, four-item LPE approximation with split scoring, classified 50% (n = 604) of participants as having elevated CU traits.

The intercorrelations and shared classification agreement (%s) across ICU cutoff types are reported in Table 2. Across classification methods, the empirical cutoff (i.e., ICU total score = 29) and four-item LPE approximation with split scoring were the most highly correlated (φ = 0.51), followed by correlations between normative cutoffs and the four-item LPE approximation with extreme scoring (φ = 0.35), the empirical cutoff and normative cutoffs (φ = 0.29), the empirical cutoff and the four-item LPE approximation with extreme scoring (φ = 0.28), four-item LPE approximation methods (i.e., extreme vs. split scoring; φ = 0.28), and the normative cutoff

and the four-item LPE approximation with split scoring (φ = 0.21). In terms of classification agreement (i.e., the proportion of participants classified by both methods as having either high or low CU traits), normative cutoffs and the four-item LPE approximation with extreme scoring resulted in the greatest classification overlap (i.e., 92%). The empirical cutoff and the four-item LPE approximation with split scoring resulted in the second highest classification overlap (75%), followed by 66% agreement both between the empirical cutoff and the four-item LPE approximation with extreme scoring and between empirical and normative cutoffs. Finally, classification overlap shared between four-item LPE approximation methods (i.e., extreme vs. split scoring; 57%) and normative cutoffs and the four-item LPE approximation with split scoring (55%) provided the least amount of agreement.

Table 2.
Associations and Classification Agreement between ICU Cutoff Methods in Sample One

	Intercorrelations (φ)			**Classification Agreement (%)**		
ICU Self-Report Cutoff Method	**Empirical**	**Norms**	**Extreme LPE**	**Empirical**	**Norms**	**Extreme LPE**
Empirical Cutoff	–	–	–	–	–	–
Normative Cutoffs	0.29***	–	–	66	–	–
Extreme LPE Approx.	0.28***	0.35***	–	66	92	–
Split LPE Approx.	0.51***	0.21***	0.28***	75	55	57

Note. LPE = "Limited Prosocial Emotions" 4-item approximation. Intercorrelations between binary cutoffs reported with phi coefficient (φ). ***$p < .001$.

Sample 2. In the second sample of normative community youth, the first cutoff method, empirically derived average optimal cutoffs of 24, 23, and 35 for ICU self-, parent-, and teacher-report, respectively, classified 25% (n = 64), 28% (n = 80), and 19% (n = 52), respectively, of participants as having elevated CU traits. The second cutoff method used multinational self-

report norms (95th percentile) and classified 5.0% (*n* = 13) of participants as having elevated CU traits, in addition to local ICU parent- and teacher-report norms (95th percentile) that classified 6% (*n* = 18) and 7% (*n* = 19), respectively, of participants as having elevated CU traits. The third cutoff method, four-item LPE approximation with extreme scoring, classified 3% (*n* = 9) of participants as having elevated CU traits based on ICU self-report, 1% (*n* = 4) based on parent-report, and 10% (*n* = 26) based on teacher-report. Finally, the fourth cutoff method, four-item LPE approximation with split scoring, classified 19% (*n* = 50) of participants as having elevated CU traits based on ICU self-report, 21% (*n* = 59) based on parent-report, and 41% (*n* = 112) based on teacher-report.

The intercorrelations and shared classification agreement (%s) across ICU cutoff types are reported in Table 3. In ICU self-report, the empirical cutoff (i.e., ICU total score = 24) and the four-item LPE approximation with split scoring were most highly correlated ($\varphi = 0.49$), followed by correlations between empirical and normative cutoffs ($\varphi = 0.40$), four-item LPE approximation methods (i.e., extreme vs. split scoring; $\varphi = 0.38$), normative cutoffs and the four-item LPE approximation with split scoring ($\varphi = 0.33$), the empirical cutoff and the four-item LPE approximation with extreme scoring ($\varphi = 0.28$), and normative cutoffs and the four-item LPE approximation with extreme scoring ($\varphi = 0.24$). In terms of classification agreement, normative cutoffs and the four-item LPE approximation with extreme scoring resulted in the greatest classification overlap (94%), followed by four-item LPE approximation methods (84%), normative cutoffs and the four-item LPE approximation with split scoring (84%), the empirical cutoff and the four-item LPE approximation with split scoring (82%), and empirical and normative cutoffs (81%). The empirical cutoff and the four-item LPE approximation with extreme scoring provided the lowest level of agreement (78%).

In ICU parent-report, the empirical cutoff and the four-item LPE approximation with split scoring were most highly correlated (φ = 0.57), followed by correlations between empirical and normative cutoffs (φ = 0.42), normative cutoffs and the four-item LPE approximation with split scoring (φ = 0.37), normative cutoffs and the four-item LPE approximation with extreme scoring (φ = 0.34). Correlations between four-item approximation methods (i.e., extreme vs. split scoring; φ = 0.23) and the empirical cutoff and the four-item LPE approximation with extreme scoring (φ = 0.19) were the lowest. In terms of classification overlap, normative cutoffs and the four-item LPE approximation with extreme scoring resulted in the greatest classification agreement (94%), followed by the empirical cutoff and the four-item LPE approximation with split scoring (83%), normative cutoffs and the four-item LPE approximation with split scoring (83%), and four-item LPE approximation methods (81%). Classification overlap between empirical and normative cutoffs (78%) and the empirical cutoff and the four-item LPE approximation with extreme scoring (73%) were the lowest.

In ICU teacher-report, correlations between the empirical cutoff and the four-item LPE approximation with extreme scoring as well as the empirical cutoff and the four-item LPE approximation with split scoring were both most highly correlated (φs = 0.57), followed by correlations between empirical and normative cutoffs (φ = 0.56), normative cutoffs and the four-item LPE approximation with extreme scoring (φ =0.40), four-item approximation methods (i.e., extreme vs. split scoring; φ = 0.39), and normative cutoffs and the four-item LPE approximation with split scoring (φ = 0.33). In terms of classification overlap, again, normative cutoffs and the four-item LPE approximation with extreme scoring resulted in the greatest agreement (91%). Second-highest levels of classification agreement (88%) were provided both by the empirical cutoff and the four-item LPE approximation with extreme scoring and empirical and normative

cutoffs, followed by the empirical cutoff and the four-item LPE approximation with split scoring (77%), four-item approximation methods (68%), and normative cutoffs and the four-item LPE approximation with split scoring (66%).

Table 3.
Associations and Classification Agreement between ICU Cutoff Methods in Sample Two

	Intercorrelations (φ)			Classification Agreement (%)		
ICU Self-Report Cutoff Method	**Empirical**	**Norms**	**Extreme LPE**	**Empirical**	**Norms**	**Extreme LPE**
Empirical Cutoff	–	–	–	–	–	–
Normative Cutoffs	0.40***	–	–	80	–	–
Extreme LPE Approx.	0.28***	0.25***	–	78	94	–
Split LPE Approx.	0.49***	0.34***	0.39***	82	84	84
ICU Parent-Report Cutoff Method	**Empirical**	**Norms**	**Extreme LPE**	**Empirical**	**Norms**	**Extreme LPE**
Empirical Cutoff	–	–	–	–	–	–
Normative Cutoffs	0.42***	–	–	78	–	–
Extreme LPE Approx.	0.19**	0.34***	–	73	94	–
Split LPE Approx.	0.57***	0.37***	0.23***	83	83	81
ICU Teacher-Report Cutoff Method	**Empirical**	**Norms**	**Extreme LPE**	**Empirical**	**Norms**	**Extreme LPE**
Empirical Cutoff	–	–	–	–	–	–
Normative Cutoffs	0.56***	–	–	88	–	–
Extreme LPE Approx.	0.57***	0.40***	–	88	91	–
Split LPE Approx.	0.57***	0.33***	0.39***	77	66	68

Note. LPE = "Limited Prosocial Emotions" 4-item approximation. Intercorrelations between binary cutoffs reported with phi coefficient (φ). ***$p < .001$.

Aim 3: examination and comparison of ICU cutoff methods.

Sample 1. Results from validity tests for each cutoff method (i.e., chi-square and independent-samples *T*-tests) are reported in Table 4. Using empirical cutoff methods, significant differences were observed between high- and low-CU groups on all outcomes except for the

binary outcome of any violent rearrests. Second, using normative cutoff methods, significant differences were observed between high- and low-CU groups on all outcomes except for total rearrests and any violent rearrests. Third, using the four-item LPE criteria approximation with extreme scoring, significant differences were observed between high- and low-CU groups on all outcomes except for total rearrests and any violent rearrests. Fourth, using the four-item LPE criteria approximation with split scoring, significant differences were observed between high- and low-CU groups on all outcomes except for the binary outcome of any violent rearrests.

Finally, average effect sizes produced by cutoff methods, along with exact total score cutoffs resulting from empirical and normative cutoff methods, are reported in Table 5. The empirical cutoff method produced the largest average effect size (r = 0.19) across all outcomes. In addition, the four-item LPE approximation with split scoring provided the next highest effect size (r = 0.15), followed by normative cutoffs (r = 0.11) and the LPE approximation with extreme scoring (r = 0.10).

Table 4.
Group-Means Difference Tests across ICU Cutoffs in Sample One

	Low-CU	**High-CU**	**Means-Difference Tests**		
Empirical Cutoff	**Mean (*SD*) or % (*n*)**	**Mean (*SD*) or % (*n*)**	**df**	***t* or χ2**	**ES *(r)***
Total physical aggression	3.44 (3.69)	6.35 (6.01)	1145	9.20***	0.26
Proactive aggression	0.63 (1.17)	1.52 (2.19)	1145	7.85***	0.23
Reactive aggression	2.80 (2.83)	4.83 (4.15)	1145	9.07***	0.26
Total SR offending	5.72 (8.48)	11.62 (14.24)	1143	7.91***	0.23
Violent SR offending	2.30 (3.32)	4.24 (5.28)	1143	6.92***	0.20
Gun carrying	0.30 (0.92)	0.64 (1.35)	1143	4.73***	0.14
Any gun use during crime	5% (*n*=36)	15% (*n*=66)	1	31.23***	0.17
Total rearrests	1.04 (1.80)	1.56 (2.28)	1182	4.18***	0.12
Any violent rearrests	17% (*n*=123)	20% (*n*=94)	1	1.98	0.04

Table cont'd.

	Low-CU	High-CU	Means-Difference Tests		
Normative Cutoffs	**Mean (*SD*) or % (*n*)**	**Mean (*SD*) or % (*n*)**	**df**	***t* or χ2**	**ES *(r)***
Total physical aggression	4.31 (4.53)	9.88 (8.57)	1145	4.74***	0.14
Proactive aggression	0.89 (1.51)	2.82 (3.47)	1145	4.07***	0.12
Reactive aggression	3.42 (3.33)	7.06 (5.50)	1145	4.82***	0.14
Total SR offending	7.63 (10.98)	15.89 (16.88)	1143	3.56**	0.11
Violent SR offending	2.89 (4.08)	6.44 (6.63)	1143	3.90***	0.12
Gun carrying	0.39 (1.07)	1.20 (1.67)	1143	3.52**	0.10
Any gun use during crime	8% (*n*=85)	31% (*n*=17)	1	35.59***	0.18
Total rearrests	1.22 (1.99)	1.66 (2.47)	1182	1.67	0.05
Any violent rearrests	18% (*n*=204)	21% (*n*=13)	1	0.31	0.02

	Low-CU	High-CU	Means-Difference Tests		
Extreme LPE Approx.	**Mean (*SD*) or % (*n*)**	**Mean (*SD*) or % (*n*)**	**df**	***t* or χ2**	**ES *(r)***
Total physical aggression	4.32 (4.70)	8.00 (6.60)	1145	4.84***	0.14
Proactive aggression	0.90 (1.58)	2.10 (2.69)	1145	3.88***	0.11
Reactive aggression	3.42 (3.41)	5.91 (4.44)	1145	4.83***	0.14
Total SR offending	7.55 (10.83)	14.38 (16.68)	1143	3.56**	0.11
Violent SR offending	2.89 (4.09)	5.41 (6.08)	1143	3.60**	0.11
Gun carrying	0.39 (1.07)	0.97 (1.57)	1143	3.18**	0.09
Any gun use during crime	8% (*n*=83)	24% (*n*=19)	1	24.62***	0.15
Total rearrests	1.22 (1.98)	1.45 (2.44)	1182	1.17	0.03
Any violent rearrests	18% (*n*=202)	19% (*n*=15)	1	0.002	0.001

	Low-CU	High-CU	Means-Difference Tests		
Split LPE Approx.	**Mean (*SD*) or % (*n*)**	**Mean (*SD*) or % (*n*)**	**df**	***t* or χ2**	**ES *(r)***
Total physical aggression	3.56 (3.80)	5.61 (5.70)	1145	7.13***	0.21
Proactive aggression	0.67 (1.19)	1.29 (2.06)	1145	6.20***	0.18
Reactive aggression	2.89 (2.91)	4.31 (3.96)	1145	6.94***	0.20
Total SR offending	5.63 (8.30)	10.45 (13.53)	1143	7.24***	0.21
Violent SR offending	2.31 (3.31)	3.83 (5.00)	1143	6.05***	0.18
Gun carrying	0.29 (0.85)	0.58 (1.32)	1143	4.42***	0.13
Any gun use during crime	6% (*n*=34)	12% (*n*=68)	1	13.17***	0.11
Total rearrests	1.09 (1.89)	1.40 (2.13)	1182	2.66**	0.08
Any violent rearrests	18% (*n*=106)	19% (*n*=111)	1	0.29	0.02

Note. CU = callous-unemotional traits; LPE = "Limited Prosocial Emotions" 4-item approximation; SR = self-reported; SD = standard deviation; df = degrees of freedom; ES = effect size in the form of correlation coefficient (Pearson's *r*). χ2 values and *t*-scores from chi-square and independent-samples *T*-tests. ***$p < .001$, **$p < .01$.

Table 5.
Comparisons of Effect Size across ICU Cutoff Methods in Sample One

ICU Self-Report Cutoff Method	ES (*r*)	*SE*	95% CI
Empirical Cutoff: ICU total = 29	0.19	0.03	0.14 – 0.23
Normative Cutoffs: Male age ≤14 = 37; Male age ≥ 15 = 41	0.11	0.01	0.08 – 0.14
Extreme LPE Approximation	0.10	0.02	0.07 – 0.13
Split LPE Approximation	0.15	0.02	0.10 – 0.19

Note. ICU = Inventory of Callous-Unemotional Traits; LPE = "Limited Prosocial Emotions" 4-item approximation; ES = average effect size (*r*) across 9 total outcomes; SE = standard error; CI = confidence interval.

Sample 2. Results from chi-square and independent-samples *T*-tests are reported in Table 6. Using empirical cutoff methods in all ICU versions, significant differences were observed between high- and low-CU groups across all outcomes. Using both multinational normative cutoffs in ICU self-report data and local normative cutoffs in ICU teacher-report data, significant group differences were observed across all outcomes. Using local normative cutoffs in ICU parent-report data, significant group differences were observed only for informant-reported conduct problems. Third, using the four-item LPE approximation with extreme scoring, ICU self-report data provided no significant group differences across outcomes, ICU parent-report data provided significant group differences on informant-reported conduct problems only, and ICU teacher-report data provided significant group differences across all outcomes. Fourth, using the four-item LPE approximation with split scoring in all ICU versions, significant group differences were observed across all outcomes.

Finally, average effect sizes produced by cutoff methods, along with exact total score cutoffs resulting from empirical and normative cutoff methods, are reported in Table 7. Again, the empirical cutoff method produced the largest average effect size for each ICU version (r_{self-}

$_{report}$ = 0.19, $r_{parent\text{-}report}$ = 0.23, $r_{teacher\text{-}report}$ = 0.29) across all outcomes. For ICU self-report, normative cutoffs provided the next highest effect size (r = 0.15), followed by the four-item LPE approximation with split scoring (r = 0.14) and LPE approximation with extreme scoring (r = 0.08). For ICU parent-report, the four-item LPE approximation with split scoring provided the next highest effect size (r = 0.20), followed by local normative cutoffs (r = 0.13) and the LPE approximation with extreme scoring (r = 0.09). Finally, for ICU teacher-report, the four-item LPE approximation with split scoring provided the second highest effect size (r = 0.26), followed by local normative cutoffs (r = 0.23), and LPE approximation with extreme scoring (r = 0.22).

Table 6.
Group-Means Difference Tests across ICU Cutoffs in Sample Two

ICU Self-Report	**Low-CU**	**High-CU**	**Means-Difference Tests**		
Empirical Cutoff	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	24.45 (7.93)	27.10 (8.61)	259	2.27*	0.14
Social preference	0.26 (1.30)	-0.45 (1.70)	259	-3.52**	0.21
Peer-nominated meanness	-0.12 (0.51)	0.33 (1.03)	259	3.41**	0.21
Normative Cutoffs	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	24.81 (7.94)	30.62 (10.56)	259	2.52*	0.16
Social preference	0.16 (1.33)	-1.42 (2.34)	259	-2.41*	0.15
Peer-nominated meanness	-0.07 (0.57)	1.05 (1.59)	259	2.51*	0.15
Extreme LPE Approx.	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	24.96 (7.97)	29.11	259	1.00	0.06
Social preference	0.11 (1.43)	-0.54	259	-1.33	0.08
Peer-nominated meanness	-0.02 (0.70)	0.28	259	1.30	0.08
Split LPE Approx.	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t* or χ2**	**ES *(r)***
Conduct problems	24.46 (7.81)	27.83 (9.09)	259	2.66**	0.16
Social preference	0.18 (1.40)	-0.31 (1.54)	259	-2.18*	0.13
Peer-nominated meanness	-0.07 (0.60)	0.23 (0.99)	259	2.02*	0.13

Table cont'd.

ICU Parent-Report	**Low-CU**	**High-CU**	**Means-Difference Tests**		
Empirical Cutoff	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	23.35 (6.32)	31.94 (10.79)	280	6.68***	0.37
Social preference	0.16 (1.40)	-0.27 (1.66)	280	-2.22*	0.13
Peer-nominated meanness	-0.07 (0.61)	0.22 (0.93)	280	2.58*	0.15
Normative Cutoffs	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	24.76 (7.37)	40.89 (12.84)	280	5.27***	0.30
Social preference	0.04 (1.45)	-0.08 (2.02)	280	-0.34	0.02
Peer-nominated meanness	0.003 (0.72)	0.14 (0.84)	280	0.77	0.05
Extreme LPE Approx.	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	25.59 (8.60)	39.50 (8.50)	280	3.21**	0.19
Social preference	0.04 (1.49)	0.04 (0.92)	280	-0.001	0.0001
Peer-nominated meanness	0.02 (0.73)	-0.37 (0.10)	280	-1.05	0.06
Split LPE Approx.	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	24.34 (7.36)	31.27 (11.13)	280	4.53***	0.26
Social preference	0.17 (1.41)	-0.46 (1.68)	280	-2.92**	0.17
Peer-nominated meanness	-0.07 (0.56)	0.31 (1.11)	280	2.55*	0.15

ICU Teacher-Report	**Low-CU**	**High-CU**	**Means-Difference Tests**		
Empirical Cutoff	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	23.39 (5.88)	34.44 (11.27)	269	6.85***	0.39
Social preference	0.21 (1.41)	-0.74 (1.54)	269	-4.29***	0.25
Peer-nominated meanness	-0.10 (0.54)	0.43 (1.11)	269	3.30**	0.20
Normative Cutoffs	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	24.54 (7.40)	38.42 (10.57)	269	5.62***	0.32
Social preference	0.11 (1.43)	-1.06 (1.71)	269	-3.41**	0.20
Peer-nominated meanness	-0.04 (0.65)	0.62 (1.12)	269	2.52*	0.15
Extreme LPE Approx.	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	24.46 (7.16)	35.31 (12.38)	270	4.39***	0.26
Social preference	0.12 (1.41)	-0.92 (1.95)	270	-3.43**	0.20
Peer-nominated meanness	-0.06 (0.62)	0.75 (1.24)	270	3.27**	0.20
Split LPE Approx.	**Mean (*SD*)**	**Mean (*SD*)**	**df**	***t***	**ES *(r)***
Conduct problems	22.38 (4.97)	29.96 (10.15)	270	7.31***	0.41
Social preference	0.24 (1.42)	-0.30 (1.56)	270	-2.97**	0.18
Peer-nominated meanness	-0.10 (0.57)	0.18 (0.90)	270	2.87**	0.17

Note. CU = callous-unemotional traits; SD = standard deviation; df = degrees of freedom; ES = effect size in the form of correlation coefficient (Pearson's *r*); LPE = "Limited Prosocial Emotions" 4-item approximation. $\chi 2$ values and *t*-scores from chi-square and independent-samples *T*-tests. ***$p < .001$, **$p < .01$, *$p < .05$.

Table 7.
Comparisons of Effect Size across ICU Cutoff Methods in Sample Two

ICU Self-Report Cutoff Method	**ES (*r*)**	***SE***	**95% CI**
Empirical Cutoff: ICU total = 24	0.19	0.04	0.12 – 0.26
Normative Cutoffs: Male age ≤14 = 37; Male age ≥ 15 = 41 Female age ≤14 = 32; Female age ≥ 15 = 35	0.15	0.04	0.08 – 0.22
Extreme LPE Approximation	0.08	0.04	0.01 – 0.15
Split LPE Approximation	0.14	0.04	0.07 – 0.21
ICU Parent-Report Cutoff Method	**ES (*r*)**	***SE***	**95% CI**
Empirical Cutoff: ICU total = 23	0.23	0.08	0.06 – 0.39
Normative Cutoffs: Male = 38; Female = 33	0.13	0.09	-0.06 – 0.31
Extreme LPE Approximation	0.09	0.06	-0.03 – 0.19
Split LPE Approximation	0.20	0.04	0.13 – 0.27
ICU Teacher-Report Cutoff Method	**ES (*r*)**	***SE***	**95% CI**
Empirical Cutoff: ICU total = 35	0.29	0.06	0.17 – 0.41
Normative Cutoffs: Male = 50; Female = 38	0.23	0.05	0.13 – 0.34
Extreme LPE Approximation	0.22	0.04	0.15 – 0.29
Split LPE Approximation	0.26	0.09	0.10 – 0.43

Note. ICU = Inventory of Callous-Unemotional Traits; LPE = "Limited Prosocial Emotions" 4-item approximation; ES = average effect size (*r*) across 3 total outcomes. SE = standard error; CI = confidence interval.

Discussion

Youth with elevated CU traits present an important mental health concern for a number of reasons. Namely, research shows youth with elevated CU traits display more severe conduct problems (Frick et al., 2003a), engage in more persistent and severe antisocial behavior (Baskin-Sommers et al., 2015; Dadds et al., 2006; Kahn et al., 2013; Lawing et al., 2010; Pardini & Fite, 2010), respond differentially or less sensitively to traditional treatment methods (Wilkinson et al., 2016), and signal the presence of distinct etiological mechanisms (i.e., reduced emotional responsiveness; Frick et al., 2014b). As a result of these important findings, the LPE specifier for CU traits has been added to the latest editions of major systems for the diagnoses of conduct problems in children (American Psychiatric Association, 2013; World Health Organization, 2018). As a result, now more than ever before, the availability of valid and comprehensive assessment tools for the detection of elevated CU traits is important both for researchers and clinicians. One such tool is the ICU, a brief (24-item) multi-informant questionnaire, which has been used extensively by researchers and has strong support for its use as a continuous total score (Cardinale & Marsh, 2017). Importantly, the ICU is the one of the few measures that directly and comprehensively assesses symptom criteria for the LPE specifier. However, one major limitation is its lack of well validated clinical cutoff scores. Thus, in the present study, we examined the clinical utility of four distinct methods for forming cutoff scores for the ICU and detecting elevated CU traits in youth.

In doing this, our first study aim was to use ROC analyses to develop empirically based cutoff scores for the ICU in two samples of children and adolescents: an ethnically and racially diverse sample of adolescents followed for 5 years after their first arrest (N = 1,216) and a school-based sample of children in the 3rd, 6th, and 8th grades (N = 289). Results of ROC analyses

showed that, total scores on all versions of the ICU (self-report in sample one; self-, parent-, and teacher-report in sample two) performed significantly better than chance at differentiating youth who did and did not show a number of important behavioral outcomes: self-reported aggression, self-reported offending, self-reported gun use, and official rearrests in sample one (see Figure 1) and informant-reported conduct problems and peer-nominated "meanness" in sample two (see Figures 2-4). These findings are consistent with past research supporting the validity of total scores from the ICU (Cardinale & Marsh 2017), as well as studies that used ROC analyses, in particular, to demonstrate the clinical utility provided by the ICU total scale in differentiating youth on prospective levels of self-reported bullying (Kimonis et al., 2014) and concurrent levels of multi-informant reported aggression and violence (Docherty et al., 2017). It should be noted, however, that while efforts were taken to include a variety of sources of information, both for the ICU and for external validators (i.e., objective arrest data in sample one; parent- and teacher-reported conduct problems, peer-nominated levels of rejection and perceived meanness in sample two), there was evidence of shared method variance in both samples. Specifically, classification accuracies (i.e., AUCs) were highest in analyses with ICU self-report scores and self-reported outcomes in sample one and ICU parent- and teacher-report scores and informant-reported conduct problems in sample two (see Figures 1-4). These findings are again consistent with past research (Docherty et al., 2017; Gao & Zhang, 2016; Roose et al., 2010).

Of particular note were the high levels of classification accuracy provided by the teacher report version of the ICU across outcomes of informant-reported conduct problems and peer-nominated meanness (i.e., AUCs = .80 and .64, respectively). Unfortunately, the vast majority of studies using the ICU have tended to use the self-report version. Our findings, along with those

of Ueno et al. (2019), suggest that increased use of the ICU teacher-report version in research is warranted, at least in pre-adolescent samples.

Our second aim was to form groups either high or low on CU traits by using the four proposed cutoff methods for the ICU. Results from this step showed that, in the first sample of justice-involved adolescents using ICU self-report scores, the four-item LPE approximation with split scoring classified the largest proportion (50%) of youth as having elevated CU traits, while the normative cutoff classified the smallest proportion of youth (5%). The empirical cutoff classified the next highest proportion of youth (39%) and the four-item LPE approximation with extreme scoring classified the next lowest (7%). Further, all of these classification methods were significantly correlated with one another (see Table 2); however, the four-item LPE approximation with split scoring and the empirical cutoff were most highly correlated ($\varphi = 0.51$) in who they classified. In contrast, in the second sample of school children, the empirical cutoff classified the largest proportion of youth (25%) as having elevated CU traits, using ICU self-report scores, while the four-item LPE approximation with extreme scoring classified the smallest proportion (3%). The four-item LPE approximation with split scoring classified the second highest proportion (19%) and the multinational normative cutoff classified the next lowest (5%). Again, while all classification methods were significantly correlated (see Table 3), the empirical cutoff and four-item LPE approximation with split scoring were most highly correlated ($\varphi = 0.49$). Interestingly, multinational age- and gender-based normative cutoffs at the 95% percentile classified equivalent rates (5%) of participants as elevated in both samples, supporting the representative utility and application of these cutoffs across a variety of sample types (Kemp et al., 2019).

In the second sample using ICU parent-report scores, results showed that the empirical cutoff classified the largest proportion (28%) of youth as having elevated CU traits, while the four-item LPE approximation with extreme scoring classified the smallest proportion of youth (1%). The four-item LPE approximation with split scoring classified the next highest proportion (21%), and the local normative cutoff classified the next lowest (6%). Again, the empirical cutoff and the four-item LPE approximation with split scoring were most highly correlated ($\varphi = 0.57$). In contrast, classification methods using ICU teacher-report scores in this sample showed that the four-item LPE approximation with split scoring classified the largest proportion of youth (41%) as having elevated CU traits, while the local normative cutoff classified the smallest proportion (7%). The empirical cutoff classified the next highest proportion (19%), and the four-item LPE approximation with extreme scoring classified the second lowest (10%), with the empirical cutoff and the four-item LPE approximation with extreme scoring being most highly correlated ($\varphi = 0.57$). Across most of these cutoff methods, youth were more likely to be classified as having elevated CU traits based on ICU teacher-report scores, rather than ICU parent-report. This may be a function of higher endorsement of CU traits by teachers ($M = 21.31$, $SD = 13.54$), rather than parents ($M = 17.26$, $SD = 9.92$).

Finally, our third aim was to compare the validity of these methods by comparing groups formed by these cutoffs on the carefully selected criteria in each sample. Average effect sizes from validity tests revealed that all tested cutoff methods resulted in small to medium effect sizes (*r*s ranging from = 0.075-0.288, broadly; Tables 5 and 7). As would be expected, given that they were derived from these validators and in these samples, the cutoffs derived from empirical methods outperformed all other cutoff methods in both samples and across all versions of the ICU (i.e., average *r*s = 0.19 for self-report in both samples; 0.23 for parent-report and 0.29 for

teacher-report in sample two; Tables 5 and 7). Thus, empirical cutoffs appeared to result in both good predictive (sample one) and concurrent validity (sample two), providing the highest estimates of overall criterion validity across relevant validators.

As noted above, the distributions of participant classification according to these empirical cutoffs (i.e., 19-34% classified as having elevated CU traits) suggest that these cutoffs are less stringent (i.e., more sensitive but less specific) when identifying youth with elevated CU traits. Thus, the validity results described above must be weighed with tradeoffs between improved sensitivity and reduced specificity when evaluating their clinical utility. Situations in which sensitivity may be particularly important include the classification of participants for research purposes, where one wants to minimize the number of false negatives (i.e., at-risk children who would be placed in the non-elevated group) and where the potential for stigma for a research diagnosis is low. Also, increased sensitivity may be particularly important when screening children at risk for negative outcomes, so that they can receive targeted and effective prevention programs that are of low risk for harm (Kimonis et al., 2019; Wilkinson et al., 2016).

Average effect sizes resulting from normative cutoffs at the 95th percentile provided lower but similarly good estimates of criterion validity (i.e., average *r*s = 0.11 for self-report in sample one; 0.15 for self-report, 0.13 for parent-report, and 0.23 for teacher-report in sample two; Tables 5 and 7). These normative cutoffs set at the 95th percentile provided more stringent (i.e., less sensitive but more specific) classification of participants. Situations in which specificity may be particularly important include forensic or justice contexts, in which, the assigned label of LPE or descriptions of elevated CU traits, similar to that of the psychopathy label in adults, may have negative consequences for the youth, such as leading to harsher sentencing decisions for

youth in the juvenile justice system (Berryessa & Wohlstetter, 2019; Prasad & Kimonis, 2018; Tyrer, 2013)

Finally, the four-item LPE approximations resulted in varying degrees of comparative validity. First, the four-item LPE approximation with split scoring generally identified the second highest number of youth as having elevated CU traits and was thus, the second most sensitive method after empirical methods. It also generally showed the second highest effect sizes after the empirical method. Conversely, the four-item LPE approximation with extreme scoring consistently resulted in the lowest estimates of criterion validity. While our findings for the detained sample are similar to those reported by Kimonis et al. (2015), who found that the split scoring method performed better in incarcerated youth, they differ from her findings that the extreme method performed better in community children and adolescents. Thus, this method for scoring the ICU requires further examination before conclusive determinations about its validity can be made. Of concern is the fact that its cutoff scores are based on only 4 items, which could lead to great variability in results across samples.

Practical Implications

In summary, our analyses suggest that cutoffs based on full-scale ICU total scores (i.e., cutoffs derived by empirical and normative methods) may be optimal over that of the brief four-item LPE criteria approximation methods. Furthermore, because both empirical and normative cutoff methods were found to provide similarly high levels of criterion validity, the choice of cutoff scores should be determined based on the purpose for which the scores are being used and by weighing the importance of false positives and false negatives. It is suggested that optimal ranges of total scores for which youth may be deemed at "at-risk" for clinically elevated CU traits include the following: ICU youth self-report = 24 – 41 in boys and 24 – 35 in girls; ICU

parent-report = 23 – 38 in boys and 23 – 33 in girls; ICU teacher-report = 35 – 50 in boys and 35 – 38 in girls. These ranges, based largely on empirical methods, are recommended for use in contexts in which sensitivity is most important (i.e., false negatives are most detrimental). In contrast, it is recommended that scores above these respective upper limits (e.g., scores above 41 for ICU self-report in boys) may be considered clinically significant for the detection of elevated CU traits when specificity is paramount.

Interestingly, our suggested ICU total score cutoff ranges are similar to cutoffs published previously, both by Kimonis et al. (2014) who reported optimal cutoff scores of 24 and 27 for ICU mother- and father-report total scores, respectively, and by Docherty et al. (2017) who reported cutoff scores of 28, 30, and 33 for ICU self-, parent-, and teacher-report total scores, respectively. Thus, research is beginning to define scores for the ICU that can be used to identify groups elevated on CU traits. However, it should be noted that the contexts in which these cutoff scores are optimal will vary. That is, for research purposes, use of the ICU as a continuous total score may still be preferred, as this typically provides increased variability and thus, increased power to detect significant associations.

Limitations

All of these interpretations should be considered in light of a number of limitations. First, in sample one, findings may have limited generalizability to populations of high-risk girls, such as female juvenile offenders, or more serious juvenile offenders as eligibility criteria for the sample included only adolescent boys who had been arrested for the first time for low-to-moderate severity offenses. In regards to generalizability to girls, research has shown that girls tend to show lower levels of CU traits (Kemp et al., 2019). Thus, using the same cutoff score for boys and girls will likely lead to more boys being designated as elevated compared to girls. For

example, in our school sample of ICU self-report data, boys were more likely to be designated as having elevated CU traits across cutoff methods that were not gender-based (i.e., empirical methods: 28% of boys designated vs. 22% of girls; LPE approximation with extreme scoring: 5% of boys vs. 3% of girls; and LPE approximation with split scoring: 27% of boys vs. 14% of girls). Future research needs to test whether gender-specific cutoff scores improve their validity for designating girls at risk for problematic outcomes. Also related to generalizability, although low base rates do not computationally influence outcomes related to sensitivity and specificity in ROC analyses, they do affect overall classification accuracy of the model or test (i.e., ICU in this case; Youngstrom, 2013). As shown in these analyses, higher base rates of youth designated as having elevated CU traits led to higher effect sizes for the comparisons using our various external criteria. Thus, samples with higher base rates of elevations may result in differences in the strength of associations with external criteria.

Second, there were also a number issues that limited the school-based sample. First, it was a relatively small sample (N = 289). This could have limited our power to detect significant associations or significant classification accuracy by ICU total scores in predicting peer rejection status. Further, this means that the normative cutoffs for the ICU parent- and teacher-report scores were based on a relatively limited normative base. Lastly, the school district from which the sample was recruited was ethnically and racially diverse but also from a relatively low-income, rural region of the southeastern United States. This may limit both the applicability of the cutoffs assessed, as well as the generalizability of the validity tests to other samples with different characteristics.

Third, although this study included a number of clinically relevant external outcomes (e.g., aggression, self-reported offending and rearrests, informant-rated conduct problems), it

also lacked a number of other external validators that research has found to be related to CU traits. For instance, numerous studies have reported that CU traits are related to a number of temperamental risk factors (e.g., fearlessness, reduced emotional reactivity; Frick & White, 2008; Viding et al., 2012), as well as reduced responsiveness to punishment and to typical mental health treatments for conduct problems (Hawes & Dadds, 2005; Hawes & Dadds, 2007; Hawes et al., 2013; Wilkinson et al., 2016). Thus, future research exploring the clinical utility of cutoffs for the ICU would benefit from the inclusion of a wider array of external criteria for the tests of validity.

Conclusions

The present study used a number of approaches for testing potential cutoff scores for the ICU, both for use in research and clinical contexts when diagnostic cutoffs are needed. It used two distinct samples that varied on age, ethnicity, and range of behavior problems, as well as a number of clinically relevant external validators measured both prospectively and concurrently and using multiple sources. Results from validity tests suggest that total score cutoffs resulting from both empirical and normative methods show substantial validity. Further, these cutoff scores were similar to those reported in other samples. Thus, research is starting to uncover scores that can aid in clinical decision making. Importantly, while such cutoffs may be relied on in some research contexts, no single score or test should be used in isolation for making clinical diagnoses (Frick, Barry, & Kamphaus, 2010). As a result, even these clinical elevations as measured by the ICU should only be one source of information used by clinicians for making a diagnosis of an LPE specifier. Other important sources of information may include the Clinical Assessment of Prosocial Emotions (CAPE; Frick, 2013), a clinical guide for the assessment of

CU traits and diagnostic criteria for the LPE specifier that allows for clinician ratings utilizing multiple sources of information.

While this study begins to establish a research base for choosing valid cutoff scores for the ICU, future research is needed to replicate and test the validity of these cutoff scores in other, more diverse samples (e.g., more severe samples of youth offenders, high-risk samples of girls, larger representative samples of community youth, including youth from urban regions) and with the use of other or additional clinically relevant external validators, such as measures of emotional reactivity, fearlessness, punishment responsivity, and responsiveness to treatment.

Appendix A. The Inventory of Callous-Unemotional Traits (ICU) – Self-Report

Name: _____

Date Completed: _____

Instructions: Please read each statement and decide how well it describes you. Mark your answer by circling the appropriate number (0-3) for each statement. Do not leave any statement unrated.

	Not at all true	Somewhat true	Very true	Definitely True
1. I express my feelings openly.	0	1	2	3
2. What I think is "right" and "wrong" is different from what other people think.	0	1	2	3
3. I care about how well I do at school or work.	0	1	2	3
4. I do not care who I hurt to get what I want.	0	1	2	3
5. I feel bad or guilty when I do something wrong.	0	1	2	3
6. I do not show my emotions to others.	0	1	2	3
7. I do not care about being on time.	0	1	2	3
8. I am concerned about the feelings of others.	0	1	2	3
9. I do not care if I get into trouble.	0	1	2	3
10. I do not let my feelings control me.	0	1	2	3
11. I do not care about doing things well.	0	1	2	3
12. I seem very cold and uncaring to others.	0	1	2	3
13. I easily admit to being wrong.	0	1	2	3
14. It is easy for others to tell how I am feeling.	0	1	2	3
15. I always try my best.	0	1	2	3
16. I apologize ("say I am sorry") to persons I hurt.	0	1	2	3
17. I try not to hurt others' feelings.	0	1	2	3
18. I do not feel remorseful when I do something wrong.	0	1	2	3
19. I am very expressive and emotional.	0	1	2	3
20. I do not like to put the time into doing things well.	0	1	2	3

21. The feelings of others are unimportant to me.	0	1	2	3
22. I hide my feelings from others.	0	1	2	3
23. I work hard on everything I do.	0	1	2	3
24. I do things to make others feel good.	0	1	2	3

Appendix B. The Inventory of Callous-Unemotional Traits (ICU) – Parent-Report

Name of Child: _____ Date of Birth: _____

Completed by: Mother Father Other: _____

Date completed: _____

Instructions: *Please complete the background information above. Then read each statement and decide how well it describes your child. Mark your answer by circling the appropriate number (0-3) for each statement. Do not leave any statement unrated.*

	Not at all true	Somewhat true	Very true	Definitely True
1. Expresses his/her feelings openly.	0	1	2	3
2. Does not seem to know "right" from "wrong".	0	1	2	3
3. Is concerned about schoolwork.	0	1	2	3
4. Does not care who he/she hurts to get what he/she wants.	0	1	2	3
5. Feels bad or guilty when he/she has done something wrong.	0	1	2	3
6. Does not show emotions.	0	1	2	3
7. Does not care about being on time.	0	1	2	3
8. Is concerned about the feelings of others.	0	1	2	3
9. Does not care if he/she is in trouble.	0	1	2	3
10. Does not let feelings control him/her.	0	1	2	3
11. Does not care about doing things well.	0	1	2	3
12. Seems very cold and uncaring.	0	1	2	3
13. Easily admits to being wrong.	0	1	2	3
14. It is easy to tell how he/she is feeling.	0	1	2	3
15. Always tries his/her best.	0	1	2	3
16. Apologizes ("says he/she is sorry") to persons he/she has hurt.	0	1	2	3
17. Tries not to hurt others' feelings.	0	1	2	3

18. Shows no remorse when he/she has done something wrong.	0	1	2	3
19. Is very expressive and emotional.	0	1	2	3
20. Does not like to put the time into doing things well.	0	1	2	3
21. The feelings of others are unimportant to him/her.	0	1	2	3
22. Hides his/her feelings from others.	0	1	2	3
23. Works hard on everything.	0	1	2	3
24. Does things to make others feel good.	0	1	2	3

Appendix C. The Inventory of Callous-Unemotional Traits (ICU) – Teacher-Report

Name of child: _____ Date completed: _____

Teacher's name: _____ Length of time you have known the student: _____

Subjects you teach the child: _____

Instructions: *Please complete the background information above. Then read each statement and decide how well it describes the student. Mark your answer by circling the appropriate number (0-3) for each statement. Do not leave any statement unrated.*

	Not at all true	Somewhat true	Very true	Definitely True
1. Expresses his/her feelings openly.	0	1	2	3
2. Does not seem to know "right" from "wrong".	0	1	2	3
3. Is concerned about schoolwork.	0	1	2	3
4. Does not care who he/she hurts to get what he/she wants.	0	1	2	3
5. Feels bad or guilty when he/she has done something wrong.	0	1	2	3
6. Does not show emotions.	0	1	2	3
7. Does not care about being on time.	0	1	2	3
8. Is concerned about the feelings of others.	0	1	2	3
9. Does not care if he/she is in trouble.	0	1	2	3
10. Does not let feelings control him/her.	0	1	2	3
11. Does not care about doing things well.	0	1	2	3
12. Seems very cold and uncaring.	0	1	2	3
13. Easily admits to being wrong.	0	1	2	3
14. It is easy to tell how he/she is feeling.	0	1	2	3
15. Always tries his/her best.	0	1	2	3
16. Apologizes ("says he/she is sorry") to persons he/she has hurt.	0	1	2	3
17. Tries not to hurt others' feelings.	0	1	2	3

18. Shows no remorse when he/she has done something wrong.	0	1	2	3
19. Is very expressive and emotional.	0	1	2	3
20. Does not like to put the time into doing things well.	0	1	2	3
21. The feelings of others are unimportant to him/her.	0	1	2	3
22. Hides his/her feelings from others.	0	1	2	3
23. Works hard on everything.	0	1	2	3
24. Does things to make others feel good.	0	1	2	3

Appendix D. Multinational Normative Database for ICU – Self-Report

Inventory of Callous-Unemotional Traits (ICU) - Self-Report Youth Version: *T*-scores & Percentiles

Raw Score	Girls				Boys				Raw Score
	Ages 11-14		Ages 15-17		Ages 11-14		Ages 15-17		
	T-score	%ile	*T*-score	%ile	*T*-score	%ile	*T*-score	%ile	
59	-	-	-	-	-	-	87	100	59
58	-	-	-	-	91	100	-	-	58
57	-	-	-	-	-	-	-	-	57
56	-	-	-	-	-	-	-	-	56
55	-	-	-	-	-	-	83	100	55
54	-	-	-	-	86	100	81	100	54
53	-	-	-	-	-	-	-	-	53
52	-	-	-	-	84	100	-	-	52
51	-	-	-	-	83	100	-	-	51
50	-	-	-	-	-	-	77	99	50
49	89	100	-	-	81	100	76	99	49
48	-	-	-	-	79	100	75	99	48
47	-	-	83	100	78	100	73	98	47
46	-	-	81	100	77	99	72	98	46
45	84	100	-	-	76	99	71	97	45
44	82	100	79	100	75	99	70	97	44
43	81	100	78	100	74	99	69	96	43
42	80	99	76	99	72	99	68	96	42
41	79	99	75	99	71	98	66	95	41
40	77	99	74	99	70	98	65	94	40
39	76	99	73	99	69	97	64	93	39
38	75	98	71	98	68	96	63	92	38
37	73	97	70	97	66	95	62	90	37
36	72	97	69	95	65	93	61	88	36
35	71	96	68	95	64	91	59	86	35
34	70	96	66	93	63	90	58	83	34
33	68	96	65	92	62	88	57	79	33
32	67	95	64	90	60	85	56	75	32
31	66	94	62	88	59	82	55	72	31
30	64	92	61	87	58	80	54	68	30
29	63	90	60	85	57	77	53	64	29
28	62	89	59	83	56	74	51	59	28
27	61	86	57	79	55	70	50	55	27
26	59	84	56	78	53	66	49	50	26
25	58	82	55	73	52	63	48	44	25
24	56	78	54	70	51	59	47	40	24
23	55	75	52	64	50	55	46	36	23
22	54	72	51	59	49	51	44	33	22
21	53	68	50	54	47	45	43	27	21
20	52	64	48	49	46	40	42	24	20
19	50	59	47	44	45	35	41	19	19
18	49	55	46	39	44	30	40	17	18
17	48	49	45	34	43	25	39	14	17
16	46	43	43	30	42	21	38	12	16
15	45	37	42	23	40	18	36	10	15
14	44	31	41	19	39	15	35	6	14
13	43	27	40	17	38	12	34	5	13
12	41	21	38	14	37	10	33	4	12
11	40	16	37	10	36	7	32	2	11
10	39	11	36	8	34	5	31	2	10
9	37	8	35	6	33	4	29	1	9
8	36	5	33	3	32	3	28	1	8
7	35	3	-	-	31	2	27	1	7
6	33	2	31	2	30	2	-	-	6
5	32	1	29	1	28	1	25	0	5
4	31	1	28	1	27	0	24	0	4
3	30	1	27	0	26	0	23	0	3
2	28	0	26	0	-	-	-	-	2
1	27	0	24	0	-	-	-	-	1
0	26	0	-	-	-	-	-	-	0
N	**1,475**		**833**		**1,444**		**931**		***N***
Raw Mean	**18.81**		**21.20**		**23.20**		**26.79**		**Raw Mean**
Raw *SD*	**7.76**		**7.88**		**8.45**		**8.66**		**Raw *SD***

Appendix E. Local Normative Database for ICU – Parent-Report

Inventory of Callous-Unemotional Traits (ICU) - Parent-Report: *T*-scores & Percentiles

Raw Score	Girls		Boys		Raw Score
	T-score	%ile	*T*-score	%ile	
55	-	-	-	-	55
54	-	-	-	-	54
53	-	-	-	-	53
52	-	-	-	-	52
51	-	-	-	-	51
50	-	-	-	-	50
49	-	-	-	-	49
48	-	-	-	-	48
47	-	-	77	100	47
46	-	-	76	99	46
45	80	100	75	99	45
44	79	99	74	99	44
43	78	98	73	98	43
42	77	98	72	97	42
41	76	98	71	97	41
40	75	98	70	97	40
39	74	98	69	96	39
38	73	97	68	95	38
37	72	97	67	93	37
36	71	96	66	92	36
35	70	96	66	91	35
34	69	96	65	90	34
33	68	95	64	89	33
32	67	93	63	88	32
31	66	92	62	85	31
30	65	90	61	84	30
29	64	89	60	83	29
28	62	88	59	82	28
27	61	86	58	81	27
26	60	85	57	78	26
25	59	84	56	74	25
24	58	83	55	73	24
23	57	78	54	71	23
22	56	75	53	67	22
21	55	73	52	66	21
20	54	72	51	62	20
19	53	69	50	59	19
18	52	67	49	54	18
17	51	64	48	50	17
16	50	60	47	47	16
15	49	54	46	44	15
14	48	50	45	39	14
13	47	44	44	35	13
12	46	37	43	31	12
11	44	33	42	28	11
10	43	32	42	25	10
9	42	28	41	18	9
8	41	23	40	13	8
7	40	18	39	11	7
6	39	15	38	9	6
5	38	10	37	6	5
4	37	9	36	5	4
3	36	4	35	4	3
2	35	1	34	3	2
1	34	1	33	2	1
0	-	-	-	-	0
N	**169**		**113**		***N***
Raw Mean	**16.21**		**18.83**		**Raw Mean**
Raw *SD*	**9.45**		**10.41**		**Raw *SD***

Appendix F. Local Normative Database for ICU – Teacher-Report

Inventory of Callous-Unemotional Traits (ICU) - Teacher-Report: *T*-scores & Percentiles

Raw Score	Girls		Boys		Raw Score
	T-score	%ile	*T*-score	%ile	
61	-	-	74	100	61
60	-	-	-	-	60
59	-	-	-	-	59
58	-	-	-	99	58
57	-	-	-	-	57
56	-	-	-	-	56
55	-	-	-	-	55
54	80	100	-	-	54
53	-	-	69	98	53
52	-	-	68	96	52
51	-	-	67	96	51
50	-	-	67	95	50
49	-	-	66	94	49
48	-	-	65	94	48
47	-	-	64	93	47
46	-	-	64	90	46
45	-	-	63	88	45
44	72	99	62	87	44
43	71	98	62	86	43
42	70	98	61	83	42
41	69	97	60	82	41
40	68	96	60	81	40
39	67	95	59	79	39
38	67	95	58	78	38
37	66	92	58	77	37
36	65	91	57	76	36
35	64	90	56	74	35
34	63	87	55	72	34
33	62	85	55	70	33
32	61	85	54	68	32
31	61	84	53	64	31
30	60	81	53	62	30
29	59	78	52	60	29
28	58	76	51	58	28
27	57	75	51	54	27
26	57	74	50	53	26
25	56	73	49	52	25
24	55	69	49	51	24
23	54	67	48	50	23
22	53	65	47	46	22
21	52	63	46	42	21
20	52	62	46	37	20
19	51	59	45	36	19
18	50	58	44	34	18
17	49	57	44	31	17
16	48	55	43	26	16
15	47	52	42	25	15
14	47	45	42	21	14
13	46	41	41	20	13
12	45	38	40	19	12
11	44	35	40	18	11
10	43	34	39	15	10
9	42	31	38	12	9
8	41	27	37	11	8
7	41	23	37	10	7
6	40	18	36	9	6
5	39	13	35	8	5
4	38	12	35	8	4
3	37	7	34	6	3
2	36	4	33	4	2
1	36	2	33	3	1
0	35	1	32	2	0
N	**164**		**107**		***N***
Raw Mean	**18.18**		**26.11**		**Raw Mean**
Raw *SD*	**11.96**		**14.43**		**Raw *SD***

References

Andershed, H. A., Kerr, M., Stattin, H., & Levander, S. (2002). Psychopathic traits in non-referred youths: A new assessment tool.

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

Barnow, S., Lucht, M., & Freyberger, H. J. (2005). Correlates of aggressive and delinquent conduct problems in adolescence. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression*, *31*(1), 24-39. doi:10.1002/ab.20033

Baroncelli, A., Roti, B., & Ciucci, E. (2018). The associations between callous-unemotional traits and emotional awareness in youth. *Personality and Individual Differences*, *120*, 247-252. doi:10.1016/j.paid.2017.07.036

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. doi:10.1177/0093854807310508

Barry, C. T., Frick, P. J., DeShazo, T. M., McCoy, M., Ellis, M., & Loney, B. R. (2000). The importance of callous–unemotional traits for extending the concept of psychopathy to children. *Journal of Abnormal Psychology*, *109*(2), 335. doi:10.1037/0021-843X.109.2.335

Baskin-Sommers, A. R., Waller, R., Fish, A. M., & Hyde, L. W. (2015). Callous-Unemotional Traits Trajectories Interact with Earlier Conduct Problems and Executive Control to Predict Violence and Substance Use Among High Risk Male Adolescents. *Journal of Abnormal Child Psychology, 43*(8), 1529-1541. doi:10.1007/s10802-015-0041-8

Berryessa, C. M., & Wohlstetter, B. (2019). The psychopathic "label" and effects on punishment outcomes: A meta-analysis. *Law and human behavior*, *43*(1), 9.

Bujang, M. A., & Adnan, T. H. (2016). Requirements for minimum sample size for sensitivity and specificity analysis. *Journal of clinical and diagnostic research: JCDR*, *10*(10), YE01.

Burt, S. A., Hyde, L. W., Frick, P. J., Jaffee, S. R., Shaw, D. S., & Tremblay, R. (2018). Commentary: Childhood conduct problems are a public health crisis and require resources: a commentary on Rivenbark et al. *Journal of Child Psychology and Psychiatry, 59*(6), 711-713. doi:10.1111/jcpp.12930

Cardinale, E. M., & Marsh, A. A. (2017). The reliability and validity of the Inventory of Callous Unemotional Traits: a meta-analytic review. *Assessment*. doi: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). New York, NY, US: Guilford Press.

Ciucci, E., Baroncelli, A., Franchi, M., Golmaryami, F. N., & Frick, P. J. (2014). The association between callous-unemotional traits and behavioral and academic adjustment in children: Further validation of the Inventory of Callous-Unemotional Traits. *Journal of Psychopathology and Behavioral Assessment*, *36*(2), 189-200. doi:10.1007/s10862-013-9384-z

Coie, J. D., & Dodge, K. A. (1998). Aggression and antisocial behavior. In W. Damon (Series Ed.) & N. Eisenberg (Vol. Ed.), *Handbook of child psychology: Vol. 3. Social, emotional, and personality development* (5th ed., pp. 779–862). New York: Wiley.

Crapanzano, A. M., Frick, P. J., Childs, K., & Terranova, A. M. (2011). Gender differences in the assessment, stability, and correlates to bullying roles in middle school children. *Behavioral Sciences & the Law*, *29*(5), 677-694. doi:10.1002/bsl.1000

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281. doi:10.1037/h0040957

Dadds, M. R., Whiting, C., & Hawes, D. J. (2006). Associations among cruelty to animals, family conflict, and psychopathic traits in childhood. *Journal of Interpersonal Violence, 21*(3), 411-429. doi:10.1177/0886260505283341

Dandreaux, D. M. & Frick, P. J. (2009). Developmental pathways to conduct problems: A further test of the childhood and adolescent-onset distinction. *Journal of Abnormal Child Psychology, 37*, 375-385. doi:10.1007/s10802-008-9261-5

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. doi:10.1002/jclp.22313

Dodge, K. A., Coie, J. D., & Brakke, N. P. (1982). Behavior patterns of socially rejected and neglected preadolescents: The roles of social approach and aggression. *Journal of Abnormal Child Psychology*, *10*(3), 389-409. doi:10.1007/BF00912329

Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology*, *39*, 349-371. doi:10.1037/0012-1649.39.2.349

Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Callous-unemotional traits in a community sample of adolescents. *Assessment*, *13*(4), 454-469. doi:10.1177/1073191106287354

Fairchild, G., Hawes, D. J., Frick, P. J., Copeland, W. E., Odgers, C. L., Franke, B., ... & De Brito, S. A. (2019). Conduct disorder. *Nature Reviews Disease Primers*, *5*(1), 43. doi:10.1038/s41572-019-0095-y

Fanti, K. A., Demetriou, C. A., & Kimonis, E. R. (2013). Variants of callous-unemotional conduct problems in a community sample of adolescents. *Journal of Youth and Adolescence*, *42*(7), 964-979. doi:10.1007/s10964-013-9958-9

Fanti, K. A., Frick, P. J., & Georgiou, S. (2009). Linking callous-unemotional traits to instrumental and non-instrumental forms of aggression. *Journal of Psychopathology and Behavioral Assessment*, *31*(4), 285. doi:10.1007/s10862-008-9111-3

Fanti, K. A., & Kimonis, E. R. (2012). Bullying and victimization: The role of conduct problems and psychopathic traits. *Journal of Research on Adolescence*, *22*(4), 617-631. doi:10.1111/j.1532-7795.2012.00809.x

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern recognition letters*, *27*(8), 861-874. doi:10.1016/j.patrec.2005.10.010

Feilhauer, J., & Cima, M. (2013). Youth psychopathy: Differential correlates of callous-unemotional traits, narcissism, and impulsivity. *Forensic Science International*, *224*(1-3), 1-7. doi:10.1016/j.forsciint.2012.10.016

Frick, P.J. (2004). The Inventory of Callous-Unemotional Traits. *Unpublished rating scale*. Retrieved from https://sites01.lsu.edu/faculty/pfricklab/icu/

Frick, P. J. (2012). Developmental pathways to Conduct Disorder: Implications for future directions in research, assessment, and treatment. *Journal of Clinical Child & Adolescent Psychology*, *41*, 378-389. doi:10.1080/15374416.2012.664815

Frick, P. J. (2013). Clinical assessment of prosocial emotions (CAPE). *Unpublished test manual*.

Frick, P. J. (2016). Current research on conduct disorder in children and adolescents. *South African Journal of Psychology*, *46*(2), 160-174. doi:10.1177/0081246316628455

Frick, P. J., Barry, C. T., & Kamphaus, R. W. (2010). Clinical assessment of child and adolescent personality and behavior . New York, NY, US: Springer Science+ Business Media.

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. doi:10.1037/1040-3590.12.4.382

Frick, P. J., Cornell, A. H., Bodin, S. D., Dane, H. E., Barry, C. T., & Loney, B. R. (2003)a. Callous-unemotional traits and developmental pathways to severe conduct problems. *Developmental Psychology, 39*(2), 246-260. doi:10.1037/0012-1649.39.2.246

Frick, P. J., Cornell, A. H., Barry, C. T., Bodin, S. D., & Dane, H. E. (2003)b. Callous-unemotional traits and conduct problems in the prediction of conduct problem severity, aggression, and self-report of delinquency. *Journal of Abnormal Child Psychology, 31*(4), 457-470. doi:10.1023/A:1023899703866

Frick, P.J. & Hare, R.D. (2001). *The Antisocial Process Screening Device (APSD)*. Toronto: Multi-Health Systems.

Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014)a. 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. doi:10.1037/a0033076

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

Frick, P. J., Stickle, T. R., Dandreaux, D. M., Farrell, J. M., & Kimonis, E. R. (2005). Callous-unemotional traits in predicting the severity and stability of conduct problems and delinquency. *Journal of Abnormal Child Psychology, 33*(4), 471-487. doi:10.1007/s10648-005-5728-9

Frick, P. J., & Viding, E. M. (2009). Antisocial behavior from a developmental psychopathology perspective. *Development and Psychopathology, 21*, 1111-1131. doi:10.1017/S0954579409990071

Frick, P. J., & White, S. F. (2008). Research review: the importance of callous-unemotional traits for developmental models of aggressive and antisocial behavior. *Journal of Child Psychology and Psychiatry, 49*(4), 359-375. doi:10.1111/j.1469-7610.2007.01862.x

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. doi:10.1007/s10862-015-9527-5

Golmaryami, F. N., Frick, P. J., Hemphill, S. A., Kahn, R. E., Crapanzano, A. M., & Terranova, A. M. (2016). The social, behavioral, and emotional correlates of bullying and victimization in a school-based sample. *Journal of Abnormal Child Psychology, 44*(2), 381-391. doi:10.1007/s10802-015-9994-x

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. doi: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. doi:10.1007/s10802-017-0330-5

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

Hawes, D. J., & Dadds, M. R. (2005). The treatment of conduct problems in children with callous-unemotional traits. *Journal of consulting and clinical psychology*, *73*(4), 737.

Hawes, D. J., & Dadds, M. R. (2007). Stability and malleability of callous-unemotional traits during treatment for childhood conduct problems. *Journal of Clinical Child and Adolescent Psychology*, *36*(3), 347-355. doi:10.1080/15374410701444298

Hawes, D. J., Dadds, M. R., Brennan, J., Rhodes, T., & Cauchi, A. (2013). Revisiting the treatment of conduct problems in children with callous-unemotional traits. *Australian & New Zealand Journal of Psychiatry, 47*(7), 646-653. doi:10.1177/0004867413484092

Hindelang, M., Hirschi, T., & Weis, J. (1981). *Measuring delinquency*: Beverly Hills, CA: Sage.

Horan, J. M., Brown, J. L., Jones, S. M., & Aber, J. L. (2016). The influence of conduct problems and callous-unemotional traits on academic development among youth. *Journal of Youth and Adolescence, 45*(6), 1245-1260. doi:10.1007/s10964-015-0349-2

Huizinga, D., Esbensen, F. A., & Weiher, A. W. (1991). Are there multiple paths to delinquency?. *The Journal of Criminal Law & Criminology*, *82*, 83–118. doi:10.2307/1143790

Kahn, R. E., Byrd, A. L., & Pardini, D. A. (2013). Callous-unemotional traits robustly predict future criminal offending in young men. *Law and Human Behavior, 37*(2), 87-97. doi:10.1037/b0000003

Kahn, R. E., Frick, P. J., Youngstrom, E., Findling, R. L., & Youngstrom, J. K. (2012). The effects of including a callous–unemotional specifier for the diagnosis of conduct disorder. *Journal of Child Psychology and Psychiatry*, *53*(3), 271-282. doi:10.1111/j.1469-7610.2011.02463.x

Kemp, E. C., Frick, P. J., & Robertson, E. L. (2019). *Inventory of Callous-Unemotional Traits (ICU) - Self-Report Youth Version: T-scores & Percentiles* [Normative dataset]. Retrieved from https://sites01.lsu.edu/faculty/pfricklab/icu/

Kimonis, E. R., Fanti, K. A., Frick, P. J., Moffitt, T. E., Essau, C., Bijttebier, P., & Marsee, M. A. (2015). Using self-reported callous-unemotional traits to cross-nationally assess the DSM-5 'With Limited Prosocial Emotions' specifier. *Journal of Child Psychology and Psychiatry*, *56*(11), 1249-1261. doi:10.1111/jcpp.12357

Kimonis, E. R., Fanti, K. A., & Singh, J. P. (2014). Establishing cut-off scores for the parent-reported inventory of callous-unemotional traits. *Archives of Forensic Psychology*, *1*(1), 27-48.

Kimonis, E. R., Fleming, G., Briggs, N., Brouwer-French, L., Frick, P. J., Hawes, D. J., ... & Dadds, M. (2019). Parent-Child Interaction Therapy adapted for preschoolers with callous-unemotional traits: An open trial pilot study. *Journal of Clinical Child & Adolescent Psychology*, *48*(sup1), S347-S361.

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Muñoz, L. C., . . . Morris, A. S. (2008). Assessing callous-unemotional traits in adolescent offenders: validation of the Inventory of Callous-Unemotional Traits. *International Journal of Law Psychiatry*, *31*(3), 241-252. doi: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.

Kruh, I.P., Frick, P.J., & Clements, C.B. (2005). Historical and personality correlates to the violence patterns of juveniles tried as adults. *Criminal Justice and Behavior*, *32*(1), 69-96. doi:10.1177/0093854804270629

Kyranides, M. N., Fanti, K. A., Sikki, M., & Patrick, C. J. (2017). Triarchic dimensions of psychopathy in young adulthood: Associations with clinical and physiological measures after accounting for adolescent psychopathic traits. *Personality Disorders: Theory, Research, and Treatment*, *8*(2), 140. doi:10.1037/per0000193

Lawing, K., Frick, P. J., & Cruise, K. R. (2010). Differences in offending patterns between adolescent sex offenders high or low in callous—unemotional traits. *Psychological Assessment, 22*(2), 298. doi:10.1037/a0018707

Loney, B. R., Frick, P. J., Clements, C. B., Ellis, M. L., & Kerlin, K. (2003). Callous-unemotional traits, impulsivity, and emotional processing in adolescents with antisocial behavior problems. *Journal of Clinical Child and Adolescent Psychology, 32*(1), 66-80. doi:10.1207/S15374424JCCP3201_07

Longman, T., Hawes, D. J., & Kohlhoff, J. (2016). Callous–unemotional traits as markers for conduct problem severity in early childhood: A meta-analysis. *Child Psychiatry & Human Development*, *47*(2), 326-334. doi:10.1007/s10578-015-0564-9

Lynam, D. R. (1997). Pursuing the psychopath: Capturing the fledgling psychopath in a nomological net. *Journal of Abnormal Psychology*, *106*(3), 425. doi:10.1037/0021-843X.106.3.425

Lynam, D. R., Caspi, A., Moffitt, T. E., Loeber, R., & Stouthamer-Loeber, M. (2007). Longitudinal evidence that psychopathy scores in early adolescence predict adult psychopathy. *Journal of Abnormal Psychology*, *116*(1), 155. doi:10.1037/0021-843X.116.1.155

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.

Marsee, M. A., Barry, C. T., Childs, K. K., Frick, P. J., Kimonis, E. R., Muñoz, L. C., ... & Lau, K. S. (2011). Assessing the forms and functions of aggression using self-report: factor structure and invariance of the Peer Conflict Scale in youths. *Psychological Assessment*, *23*(3), 792-804. doi:10.1037/a0023369

Marsee, M. A., & Frick, P. J. (2007). Exploring the cognitive and emotional correlates to proactive and reactive aggression in a sample of detained girls. *Journal of Abnormal Child Psychology*, *35*(6), 969–981. doi:10.1007/s10802-007-9147-y

Marsee, M. A., Frick, P. J., Barry, C. T., Kimonis, E. R., Centifanti, L. C. M., & Aucoin, K. J. (2014). Profiles of the forms and functions of self-reported aggression in three adolescent

samples. *Development and Psychopathology*, *26*(3), 705-720. doi:10.1017/S0954579414000339

Matlasz, T. M., Frick, P. J., & Clark, J. E. (2019). Understanding the social relationships of youth with callous-unemotional traits using peer nominations. *Manuscript under review for publication.*

McMahon, R. J., Witkiewitz, K., & Kotler, J. S. (2010). Predictive validity of callous–unemotional traits measured in early adolescence with respect to multiple antisocial outcomes. *Journal of Abnormal Psychology*, *119*(4), 752. doi:10.1037/a0020796

Miron, C. D., Satlof-Bedrick, E., & Waller, R. (2020). Longitudinal association between callous-unemotional traits and friendship quality among adjudicated adolescents. *Journal of Adolescence*, *81*, 19-26.

Moffitt, T. E. (2018). Male antisocial behaviour in adolescence and beyond. *Nature Human Behaviour*, 1. doi:10.1038/s41562-018-0309-4

Monahan, K. C., & Piquero, A. R. (2009). Investigating the longitudinal relation between offending frequency and offending variety. *Criminal Justice and Behavior*, *36*(7), 653-673. doi:10.1177/0093854809335527

Muñoz, L. C., Frick, P. J., Kimonis, E. R., & Aucoin, K. J. (2008). Types of aggression, responsiveness to provocation, and callous-unemotional traits in detained adolescents. *Journal of Abnormal Child Psychology*, *36*(1), 15-28. doi:10.1007/s10802-007-9137-0

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. doi:10.1177/0093854807310159

Newcomb, A. F., Bukowski, W. M., & Pattee, L. (1993). Children's peer relations: A meta-analytic review of popular, rejected, neglected, controversial, and average sociometric status. *Psychological Bulletin*, *113*(1), 99-128. doi:10.1037/0033-2909.113.1.99

Odgers, C. L., Moffitt, T. E., Broadbent, J. M., Dickson, N., Hancox, R. J., Harrington, H., . . . Caspi, A. (2008). Female and male antisocial trajectories: From childhood origins to

adult outcomes. *Developmental Psychopathology*, *20*, 673-716. doi:10.1017/s0954579408000333

Pardini, D. A., & Fite, P. J. (2010). Symptoms of conduct disorder, oppositional defiant disorder, attention-deficit/hyperactivity disorder, and callous-unemotional traits as unique predictors of psychosocial maladjustment in boys: Advancing an evidence base for DSM-V. *Journal of the American Academy of Child & Adolescent Psychiatry*, *49*(11), 1134-1144. doi:10.1016/j.jaac.2010.07.010

Pardini, D. A., Frick, P. J., & Moffitt, T. E. (2010). Building an evidence base for DSM–5 conceptualizations of oppositional defiant disorder and conduct disorder: Introduction to the special section. *Journal of Abnormal Psychology*, *119*(4), 683. doi:10.1037/a0021441

Patrick, C. J., & Drislane, L. E. (2015). Triarchic model of psychopathy: Origins, operationalizations, and observed linkages with personality and general psychopathology. *Journal of Personality*, *83*(6), 627-643. doi:10.1111/jopy.12119

Pelham Jr., W. E., 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. doi: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. doi:10.1007/bf00927116

Piatigorsky, 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. doi:10.1023/b:jacp.0000037782.28482.6b

Prasad, A. H., & Kimonis, E. R. (2018). Effects of the "limited prosocial emotions" specifier for conduct disorder on juror perceptions of juvenile offenders. *Criminal Justice and Behavior*, *45*(10), 1547-1564.

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.* doi:10.1080/15374416.2018.1504297

Ray, J.V., Frick, P.J., Thornton, L.C., Wall Myers, T.D., Steinberg, L., Cauffman, E. (2017). Callous-unemotional traits predict self-reported offending in adolescent boys: The mediating role of delinquent peers and the moderating role of parenting practices. *Developmental Psychology, 53*(2), 319-328. doi:10.1037/dev0000210

Rivenbark, J. G., Odgers, C. L., Caspi, A., Harrington, H., Hogan, S., Houts, R. M., . . . Moffitt, T. E. (2018). The high societal costs of childhood conduct problems: evidence from administrative records up to age 38 in a longitudinal birth cohort. *Journal of Child Psychology and Psychiatry, 59*(6), 703-710. doi:10.1111/jcpp.12850

Robertson, E.L., Frick, P.J., Walker, T.M., Kemp, E.C., Ray, J.V., Thornton, L.C., Wall Myers, T.D., Steinberg, L., & Cauffman, E. (in press). Callous-Unemotional Traits Increases Risk of Gun Carrying and Gun Use during Crime.

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. doi:10.1177/1073191109344153

Rowe, R., Maughan, B., Moran, P., Ford, T., Briskman, J., & Goodman, R. (2010). The role of callous and unemotional traits in the diagnosis of conduct disorder. *Journal of Child Psychology and Psychiatry, 51*(6), 688-695. doi:10.1111/j.1469-7610.2009.02199.x

Saukkonen, S., Laajasalo, T., Jokela, M., Kivivuori, J., Salmi, V., & Aronen, E.T. (2016). Weapon carrying and psychopathic-like features in a population-based sample of Finnish adolescents. *European Child & Adolescent Psychiatry, 25*(2), 183-191. doi:10.1007/s00787-015-0724-2

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. doi:10.1111/j.1475-3588.2008.00483.x

Thornberry, T., & Krohn, M. (2000). The self-report method for measuring delinquency and crime. *Criminal Justice, 4*(1), 33-83.

Tyrer, P. (2013). The classification of personality disorders in ICD-11: Implications for forensic psychiatry.

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*, doi:1073191119847766.

Waller, R., Baskin-Sommers, A. R., & Hyde, L. W. (2016). Examining predictors of callous unemotional traits trajectories across adolescence among high-risk males. *Journal of Clinical Child & Adolescent Psychology*, 1-14. doi:10.1080/15374416.2015.1102070

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. doi:10.1007/s10578-006-0033-6

White, S. F., Frick, P. J., Lawing, K., & Bauer, D. (2013). Callous-unemotional traits and response to Functional Family Therapy to adolescent offenders. *Behavioral Sciences & the Law, 31*, 271-285. doi:10.1002/bsl.2041

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. doi:10.1111/jcpp.12494

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

Viding, E. (2004). Annotation: Understanding the development of psychopathy. *Journal of Child Psychology and Psychiatry, 45*(8), 1329-1337. doi:10.1111/j.1469-7610.2004.00323.x

Viding, E., Fontaine, N. M., & McCrory, E. J. (2012). Antisocial behaviour in children with and without callous-unemotional traits. *Journal of the Royal Society of Medicine*, *105*(5), 195-200. doi:10.1258/jrsm.2011.110223

Yoshida, S., Adachi, M., Takahashi, M., Takanyanagi, N., Yasuda, S., Osada, H., & Nakamura, K. (2019). The factor structure and construct validity of the parent-reported Inventory of Callous-Unemotional Traits among school-aged children and adolescents. *PloS one*, *14*(8). doi:10.1371/journal.pone.0221046

Youngstrom, E. A. (2013). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *Journal of Pediatric Psychology, 39*(2), 204-221. doi:10.1093/jpepsy/jst062

Vita

Emily was born in Richardson, Texas and was raised in both Dallas and Houston, Texas. She graduated one semester early and with highest (summa cum laude) and departmental honors, with a Bachelor of Science (B.S.) Degree in psychology and minor in sociology, from the University of Texas at Austin in December 2014. Following graduation, she moved to New Haven, Connecticut to receive full-time research training as a Clinical Research Assistant in the Yale School of Medicine's Psychiatry Department. In 2018, she joined the Clinical Psychology Ph.D. Program at Louisiana State University under Dr. Paul Frick in the Developmental Psychopathology Lab. Emily plans to receive her Master of Arts (M.A.) Degree in August 2020 and her Ph.D. in clinical psychology in 2023.