Article

Assessment
2021, Vol. 28(2) 567–584

DOI: 10.1177/1073191119847766
journals.sagepub.com/home/asm
SAGE

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

Kathrin Ueno[1], Katharina Ackermann[1], Christine M. Freitag[1], and Christina Schwenck[1,2]

Abstract

Background: This article reports reliability, validity, and norms for the German version of the multi-informant questionnaire Inventory of Callous–Unemotional Traits (ICU). **Method:** The ICU was filled in by nonreferred children aged 13 to 18 years old (n = 645), parents of children aged 6 to 18 years old (n = 1,005), and their teachers (n = 955). **Results:** Confirmatory factor analysis resulted in a two-factor solution giving the best fit. Still none of the models showed an adequate model-fit applying the chi-square exact fit test. The internal consistency of the parent's, teacher's, and self-report version were α = .830, α = .877 and α = .769, respectively. Interrater reliability was moderate. Convergent validity with the Youth Psychopathic Traits Inventory, the externalizing scores of the Youth Self-Report/Child Behavior Checklist, and with the German oppositional Defiant Disorder/Conduct Disorder Rating Scale "FBB-SSV" were good. German norms were calculated. **Conclusions:** The ICU is a reliable and valid dimensional measure to describe callous–unemotional traits.

Keywords

Inventory of Callous–Unemotional Traits, ICU, multi-informant questionnaire, factor structure, reliability, validity, norms.

The fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*; American Psychiatric Association, 2013) introduced the specifier "With Limited Prosocial Emotions" (LPE; 312.8) for conduct disorder (CD), which specialists refer to as "callous–unemotional" (CU) traits. To meet criteria for the specifier, at least two of the following behaviors have to be persisting over at least 12 months and affecting multiple relationships and settings: (a) lack of remorse or guilt, (b) callous lack of empathy, (c) lack of concern about performance, and (d) shallow or deficient affect (American Psychiatric Association, 2013). CU traits are found in up to 50% of children and adolescents with CD (Kahn, Frick, Youngstrom, Findling, & Youngstrom, 2012), but also can be observed in individuals without any mental disorder (Kahn et al., 2012; Kimonis, Fanti, & Singh, 2014; Rowe et al., 2010). CU traits are related to severity and chronicity of CD as well as a reduced response to therapy (Edens, Campbell, & Weir, 2006; Frick, Cornell, Barry, Bodin, & Dane, 2003; Högström, Enebrink, & Ghaderi, 2013; Kimonis, Frick, Munoz, & Aucoin, 2007; Pardini & Fite, 2010; Ray, Frick, Thornton, Steinberg, & Cauffman, 2016; Viding, Simmonds, Petrides, & Frederickson, 2009). Children and adolescents with high CU traits but without the diagnosis of CD regularly display subclinical CD problems, lower academic achievement, less prosocial behavior, and more social problems than their age- and sex-matched peers (Rowe et al., 2010). Therefore, a valid and reliable assessment of CU traits in children with and without CD is needed. The Inventory of Callous–Unemotional Traits (ICU; Frick, 2004) is a comprehensive and commonly used instrument to assess these traits in children and adolescents.

Factor Structure of the Inventory of Callous–Unemotional Traits

The ICU contains 24 items (of which 12 are positively and 12 are negatively worded) and is available in various

[1]Goethe University, Frankfurt, Germany
[2]Justus-Liebig-University, Giessen, Germany

Corresponding Author:
Kathrin Ueno, Department of Child and Adolescent Psychiatry, Psychosomatics, and Psychotherapy, University Hospital Frankfurt, Goethe University, Frankfurt, Germany.
Email: kathrin.ueno@kgu.de

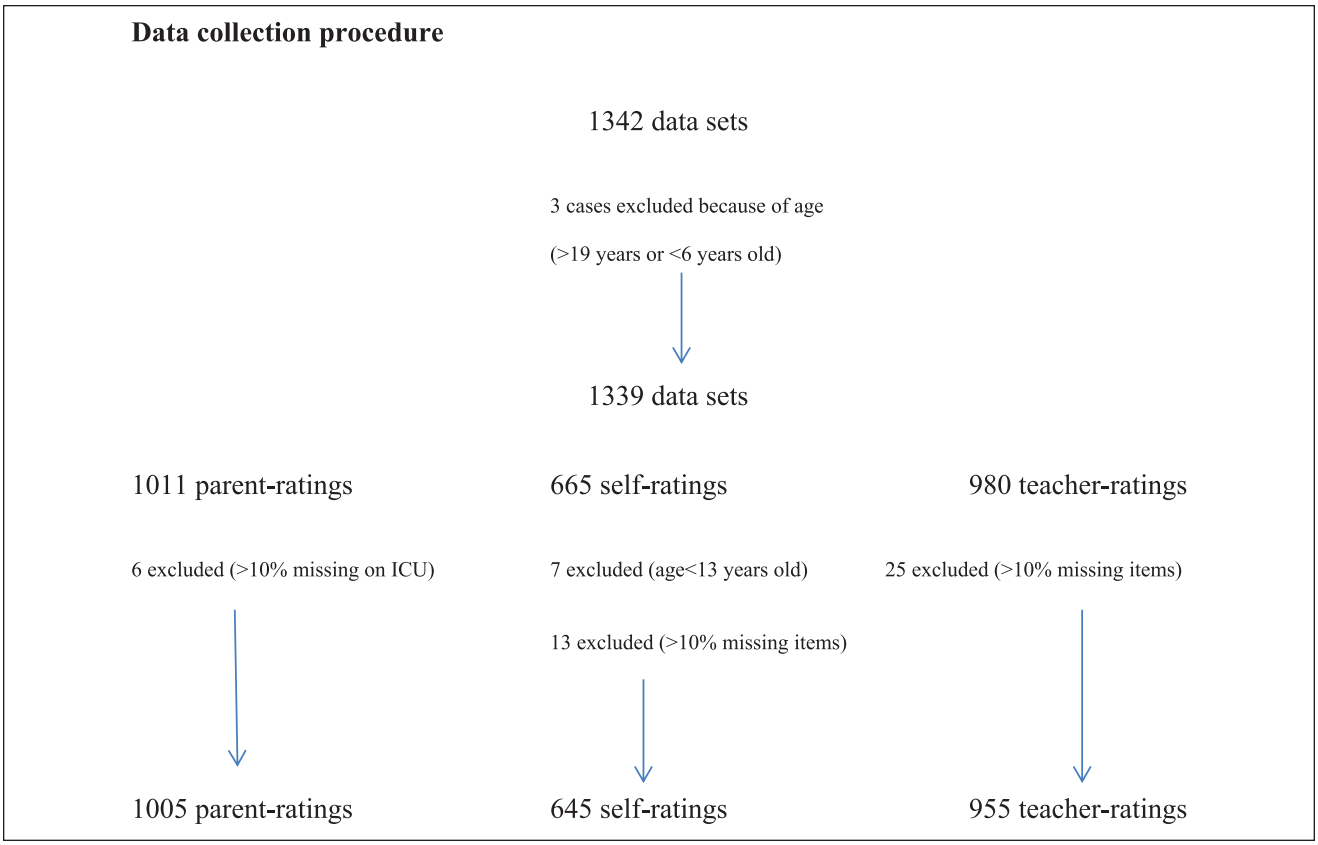

Figure 1. Data collection procedure.

languages and versions (self, parent, and teacher). Factor structure has mainly been studied for the self-report version (results presented first) and resulted in three models. One-factor structure yielded was a model with the three factors callousness (i.e., "lack of empathy, guilt, and remorse of misdeeds"), uncaring (i.e., "lack of caring about one's performance in tasks and for the feelings of other people") and unemotional (i.e., "absence of emotional expression"; Essau, Sasagawa, & Frick, 2006; Model 1: Supplement Figure 1 [Supplementary materials are available online]). In a Portuguese sample of detained juvenile offenders, this factor structure was found (Pechorro, Ray, Barroso, Maroco, & Goncalves, 2016). Other studies detected a bifactorial model with the three intercorrelated second-order factors (callous, uncaring, and unemotional) loading on a first-order general factor (CU; Model 2: Supplement Figure 2). This model was found in nonreferred (Essau et al., 2006; Fanti, Frick, & Georgiou, 2009) and detained (Kimonis et al., 2008) adolescents (12-20 years) from Germany, Greece, and the United States. However, the callousness subscale consists of negatively worded items and the uncaring subscale of positively worded items, suggesting that the bifactorial structure may result from different item-endorsement rates and item difficulties (Ray, Frick, et al., 2016). Finally, a two-factor structure with the dimensions uncaring and callousness was detected in an Australian study of nonreferred boys, based on a short 12-item ICU (Model 3: Supplement Figure 3; Houghton, Hunter, & Crow, 2013). The structure was replicated in a Belgian sample of 12 to 17 years old detained girls (Colins, Andershed, Hawes, Bijttebier, & Pardini, 2016) and in a Portuguese sample of 7 to 17 years old nonreferred children (Carvalho, Faria, Conceicao, Gaspar de Matos, & Essau, 2018).

Studies on the three models applying the parent-report version are less common and involve samples based in the United States. Here, the bifactorial model (Model 2) was found in nonreferred adolescents (Horan, Brown, Jones, & Aber, 2015) and high-risk 9-year-olds (Waller et al., 2015). The two-factor structure (Model 3) was confirmed in samples of referred boys aged 6 to 12 years (Hawes et al., 2014) and of at-risk 7- to 12-year-olds (McDonald et al., 2018).

In a Dutch sample of nonreferred 14- to 20-year-olds applying the self-, parent-, and teacher-report version, the bifactorial model (Model 2) was found to fit best the data of the self-report version. However, Model 2 was discussed to be insufficient for the teacher- and parent-report version (Roose, Bijttebier, Decoene, Claes, & Frick, 2010). In an American sample of nonreferred 8- to

10-year-olds, the two-factor solution (Model 3) had the best fit for the self-report version and the three-factor solution (Model 1) for the parent-report version (Gao & Zhang, 2016). This illustrates the need to study the different ICU versions (self, parent, and teacher) to confirm the best-factor structure for each version.

Reliability and Validity of the Inventory of Callous–Unemotional Traits

Most reliability studies of the ICU (total scale) were done on the self-report version and, to a lesser degree, the parent-report version. In culturally diverse samples, the internal consistency of the ICU was found to be acceptable to good with $\alpha = .71$ to .90 for the self-report version (Berg et al., 2013; Colins et al., 2016; Decuyper, De Bolle, De Fruyt, & De Clercq, 2011; Decuyper, De Caluwe, De Clercq, & De Fruyt, 2014; Eisenbarth, Demetriou, Kyranides, & Fanti, 2016; Essau et al., 2006; Fanti et al., 2009; Feilhauer, Cima, & Arntz, 2012; Kimonis et al., 2015; Kimonis, Kennealy, & Goulter, 2016; Kongerslev, Bo, Forth, & Simonsen, 2015; Latzman, Lilienfeld, Latzman, & Clark, 2013; Levy et al., 2017; Lopez-Romero, Gomez-Fraguela, & Romero, 2015; Marsee et al., 2011; Pechorro et al., 2016; Ray, Pechorro, & Goncalves, 2016; Roose et al., 2010; White, Frick, Lawing, & Bauer, 2013) and $\alpha = .70$ to .89 for the parent-report version (Decuyper et al., 2011; Decuyper et al., 2014; Henry, Pingault, Boivin, Rijsdijk, & Viding, 2016; Herpers et al., 2017; Latzman et al., 2013; McDonald et al., 2018). Only one study from Israel tested the teacher-report version in an at-risk sample of adolescent males and reported a good internal consistency ($\alpha = .86$; Levy et al., 2017).

The test–retest reliabilities of the self- and parent-report version were acceptable to good in different samples: Self-report version: mixed Dutch sample, 43 weeks apart, $r = .72$ (Feilhauer et al., 2012); at-risk American sample, 6 to 10 weeks apart, $r = .459$ to .671, $p < .01$ (Berg et al., 2013); Danish male forensic sample, 9 days apart, intraclass correlation [ICC] = .88 (Kongerslev et al., 2015); Parent-report version: sample of nonreferred American twins, mean of 23 days apart, $r = .84$ (Moore et al., 2017); nonreferred Greek sample, 6/12 months apart, $r = .56$ to .62 (Kimonis et al., 2014). No studies have been done on the teacher-report version.

Varying by rater, sample and ICU form (full vs. shortened) interrater reliability was low to moderate. Low interrater reliabilities were found in nonreferred American samples for the full form, self-guardian: $r = .19$, self-teacher: $r = .24$, guardian–teacher: $r = .17$ (Docherty, Boxer, Huesmann, O`Brien, & Bushman, 2016), the shortened form across self-, teacher-, and guardian-report versions ($r = .15$ to .23; Gao & Zhang, 2016) and between the self- and teacher-report version in an at-risk Israeli male adolescent sample for the full form (Levy et al., 2017). Moderate interrater reliabilities existed for the full forms in an at-risk sample of 13- to 17-year-old Americans (self-caregiver: $r = .40$, $p < .01$; Berg et al., 2013) and in nonreferred Dutch-samples of adolescents aged 12 to 20 years (Decuyper et al., 2011; Decuyper et al., 2014).

All in all, reliability of the ICU seems to be acceptable to good. However, studies were mainly conducted in the United States and focused on the self- and parent-report version, missing results applying the teacher-report version.

CU traits are defined as a specifier of CD, and also show shared additive genetic effects with CD (American Psychiatric Association, 2013; Saunders, 2016). Accordingly, convergent validity of the ICU is assumed by positive correlations with other instruments covering the same construct or aspects of CD. Studies applying the self-report version in different countries showed moderate positive correlations with self-ratings of psychopathic traits ($r = .49$ to .63; Colins et al., 2016; Fink, 2010; Pechorro et al., 2016; Roose et al., 2010) and the Youth Self-Report (YSR; $r = .30$; Essau et al., 2006). Effects were most pronounced for callousness ($r = .38$) and uncaring ($r = .52$) compared with the unemotional scale ($r = .14$; Roose et al., 2010). Correlations with the Psychopathy Checklist: Youth Version (PCL:YV; Forth, Kosson, & Hare, 2003), the gold standard for psychopathy traits in delinquent adolescents based on a semistructured interview and review of legal reports, were positive and mostly significant ($r = .13$ to .76; Feilhauer et al., 2012; Fink, Tant, Tremba, & Kiehl, 2012; Harenski, Harenski, & Kiehl, 2014; Johnson et al., 2013; Kongerslev et al., 2015) as were with measures of aggression and delinquency ($r = .28$ to .57; Colins et al., 2016; Fanti et al., 2009; Feilhauer et al., 2012). Applying the parent-/caregiver-report version in divergent countries, moderate to high positive correlations were found with the PCL:YV ($r = .27$; Feilhauer et al., 2012; Fink et al., 2012; Harenski et al., 2014; Johnson et al., 2013; Kongerslev et al., 2015), the Child Behavior Checklist (CBCL) aggressive behavior/rule-breaking scale ($r = .51/.55$; Berg et al., 2013) and with measures of aggression and delinquency ($\beta = .52$ and .24; Colins et al., 2016; Fanti et al., 2009; Feilhauer et al., 2012; Horan et al., 2015). No validity study has been performed for the ICU teacher-report version.

Frick, Ray, Thornton, and Kahn (2014) proposed that CU traits may be associated with low levels of anxiety. Accordingly, elevated CU traits are expected to be associated with a reduced responsiveness to anxiety and grief (Anastassiou-Hadjicharalambous & Warden, 2008; Frick & Dickens, 2006; Frick et al., 2014; Leist & Dadds, 2009). Divergent validity of the ICU is assumed by negative correlations with anxiety/internalizing disorders. The ICU self-report version showed a significant negative association with internalizing dimensions (Carvalho et al., 2018) and anxiety disorders ($\beta = -.23$; Colins et al., 2016) in Portuguese and Belgian samples. However, other studies

conducted in divergent countries did not detect associations between the ICU self-report version and the internalizing scale of the YSR ($r = .03$; Essau et al., 2006), anxiety ($r = -.01$; Fanti, Demetriou, & Kimonis, 2013), and affective disorders ($\beta = .08$; Colins et al., 2016), or weak to strong associations with the CBCL anxious/depression ($r = .15$) and trait anxiety/depression ($r = .51/36$; Berg et al., 2013). Furthermore, Fink (2010) posits the affective presentation of CU to be similar to that of depression in adolescence. Accordingly, positive correlations were found with depressive symptoms (Fink, 2010), depression/anxiety ($r = .36/.51$; Berg et al., 2013), the anxious/depressed subscales of the CBCL and YSR ($r = .16/.24$), and the withdrawal/depressed subscales of the CBCL and the YSR ($r = .17/.27$; Gao & Zhang, 2016) in American studies. The ICU parent-/caretaker-report version showed mostly weaker associations with anxiety ($\beta = -.01$; Horan et al., 2015), the CBCL anxious/depression ($r = .19$), trait anxiety/depression ($r = .18/.16$; Berg et al., 2013), the CBCL/YSR anxious/depressed ($r = .30/.07$) and withdrawal/depressed subscales ($r = .42/.17$; Gao & Zhang, 2016) in American studies. Results regarding validity are altogether inconsistent, and again, studies on the teacher-report version have not been done.

Distinctive correlations with external factors emerged when applying the ICU subscales in studies conducted in divergent countries (Charles, Acheson, Mathias, Furr, & Dougherty, 2012; Essau et al., 2006; Kimonis et al., 2008; Lopez-Romero et al., 2015; Ray, Pechorro, et al., 2016; Roose et al., 2010; Waller et al., 2015). The unemotional scale was low on convergent (Charles et al., 2012; Essau et al., 2006; Kimonis et al., 2008; Lopez-Romero et al., 2015; Ray, Pechorro, et al., 2016; Roose et al., 2010; Waller et al., 2015) and divergent validity across different versions of the ICU (Waller et al., 2015), showed poor internal reliability (Gao & Zhang, 2016) and integrated some low-loading items (Kimonis et al., 2008). Hence, it is questionable whether the items marked as "unemotional" on the ICU indeed capture this quality as it is defined by the nomological network of the LPE specifier (Waller et al., 2015). In one study, the callousness scale was significantly related with anxious-depressed and withdrawn-depressed ratings (Waller et al., 2015). However, other studies that applied different versions of the ICU had callousness and uncaring scales in line with the total ICU score (Lopez-Romero et al., 2015; Ray, Pechorro, et al., 2016; Roose et al., 2010).

Effects of Sex and Age

Males generally scored higher on the ICU and were twice as likely as females to meet the criteria for the CU specifier in community, clinical, and at-risk samples from divergent countries (Carvalho et al., 2018; Charles et al., 2012; Decuyper et al., 2011; Decuyper et al., 2014; Docherty et al., 2016; Essau et al., 2006; Fanti et al., 2013; Henry et al., 2016; Kimonis et al., 2015; Marini & Stickle, 2010; Moore et al., 2017; Stickle, Kirkpatrick, & Brush, 2009; Viding et al., 2009). However, in American studies of nonreferred 8- to 10-year-olds (Gao & Zhang, 2016) and at-risk 13- to 17-year-olds (Berg et al., 2013), Gao and Berg did not identify sex differences for the ICU self-report version. Among American males, strong correlations were found for CU traits and the CBCL-derived scales rule-breaking/aggressive behavior and withdrawn/depressed. Among American girls, CU traits correlated with the CBCL-derived scales attention problems, aggressive behavior, and withdrawn/depressed (Charles et al., 2012).

Studies on nonreferred samples have given conflicting results regarding the effect of age. These studies have used different versions of the ICU, and no study was done on the ICU teacher rating version. Two studies did not observe age effects, one American study of 9- to 14-year-old children, applying the parent-report version ($\beta = .25$; $p = .21$; Moore et al., 2017) and one Dutch study of 12- to 18-year-old adolescents, applying the self-/mother-report version (Decuyper et al., 2011). A Portuguese study based on a shortened ICU version found in contrast that 7- to 10-year-old children showed higher self-reported ICU scores than 11- to 14-year-old and 15- to 17-year-old adolescents (Carvalho et al., 2018). Essau et al. (2006) found in a German sample a curvilinear effect, whereby 15- to 16-year-old adolescents scored higher on the ICU self-report version than their 13- to 14-year-old and 17- to 18-year-old counterparts.

Open Issues

The inclusion of the specifier LPE in *DSM-5* and the developmental impact of CU traits on referred and nonreferred children and adolescents necessitates a valid and reliable assessment of CU traits. Despite a growing body of evidence supporting the ICU as a promising assessment tool, previous studies show several limitations: First, most studies focused on the self-report version, some included the parent-report version, but hardly any study does exist on the teacher-report version. Second, studies have generally focused on either children or adolescents without reporting data on the entire age range (6- to 18-year-olds). Third, most studies have been conducted in the United States, and some in Western Europe, restricting the generalizability of the results. To date, only one study (Essau et al., 2006) has applied the German version of the ICU; this study investigated a restricted age range (13- to 18-year-olds) and only implemented the self-report version. Fourth, despite numerous studies on the ICU, norms are missing to identify nonnormative CU traits. Last, the factorial structure of the different ICU version needs to be clarified, given heterogeneous results of previous studies.

Aims of This Study

This study examines the factor structure, reliability, validity of the ICU self-, parent-, and teacher-report versions and aims to derive norms for German children and adolescents aged 6 to 18 years. More specifically, the first aim is to explore the ICU factor structure through confirmatory factor analysis (CFA) by comparing different factor models. Considering the questionable reliability of the unemotional scale and the results stated above, we expect the two-factor model based on a shortened ICU to best fit the data of the three ICU versions. The second aim is to assess the reliability of the ICU regarding internal consistency, test–retest reliability, and interrater reliability. We expect good internal consistencies, good test–retest reliability, but only moderate interrater reliability of the three versions. The third aim is to assess the convergent and divergent validity of the ICU by relating it to measures designed to tap similar personality traits (Youth Psychopathic Traits Inventory [YPI]) or external behavior problems (i.e., externalizing symptom scores of the CBCL and the YSR, Observer Rating Scale for Oppositional Defiant and Conduct Disorders [FBB-SSV]) versus internal behavior problems (i.e., Beck Depression Inventory–Second edition [BDI-II], internalizing symptom scores of the CBCL and the YSR, Screen for Child Anxiety Related Disorders [SCARED]). For the three ICU versions, we expect positive correlations for the YPI and scales targeting external behavior problems and negative correlations for scales targeting internal behavior problems. We assume the uncaring and callousness subscales will correlate with the ICU total score and the unemotional subscale to show a significant positive correlation with internal behavior problems. Finally, the study was designed to derive norms (percentile rank and Stanine) to estimate individual specifications of CU traits. Since ICU scores are gender-specific, we also expect to find gender-specific norms; for example, boys will need higher ICU scores to reach the same norm as girls. Furthermore, based on the inconsistent results regarding the relationship of age and ICU scores, this study aims to explore the appropriateness of age-specific ICU norms.

Method

Participants

Participants were recruited from 49 schools in three German federal states (Bayern, Hessen, and Rheinland-Pfalz) with approval by the local federal government as well as the school director. Schools were randomly selected to cover all curricula (23 primary schools, 22 secondary schools, 1 vocational school, and 1 high school) as well as different areas (urban and rural). No incentives were provided for participation. Data were collected on 1,342 children and adolescents; 1,011 parents (75.33% of the initially invited group), and 980 teachers (73.03% of the initially invited group) returned the corresponding assessment. Children younger than 13 years old did not complete the self-report version. Figure 1 provides an overview on the number of included questionnaires.

Incomplete ICUs (i.e., one or more items unanswered) were returned by 10.54% of participating adolescents, 7.56% of parents, and 11.62% of teachers. If 10% or more of the data were missing, data were not included into analysis. For scales with less than 10% missing data, the missing item value was replaced by the sex- and age-specific group mean of the respective item. ICU data were provided by all three raters (self, parent, and teacher) for 324 children and adolescents, parent- and self-report versions were available for 392, teacher- and self-report versions for 454, and parent- and teacher-report versions for 755 participants. The sample characteristics are displayed in Table 1.

Socioeconomic status of the participants was representative of the average net monthly family income in Germany (Statistisches Bundesamt, 2017). Children from rural areas and those without an immigration background were overrepresented in the study (Statistisches Bundesamt, 2017).

Measures

Inventory of Callous–Unemotional Traits (Frick, 2004). This study assesses the German self-report version (for children ≥13 years old), the parent- and teacher-report versions of the ICU (German translation of the self- and parent-report version by Essau et al. [2006], teacher-report version adapted accordingly). Answers are coded on a 4-point Likert-type scale ranging from 0 (*not at all true*) to 3 (*definitely true*).

Observer Rating Scale for Oppositional Defiant and Conduct Disorders (Döpfner, Görtz-Dorten, & Lehmkuhl, 2008). Parents and teachers filled in the FBB-SSV to give information on oppositional defiant disorder and CD. Of the original 25 items, 2 of 4 items are pooled, resulting in 23 final items. Eight items measure oppositional defiant behavior and 15 items dissocial aggressive behavior. Items are rated on a 4-point Likert-type scale (0-3), which is dichotomized (0, 1 vs. 2, 3) to describe the presence and absence of single symptoms. The cutoff score for oppositional defiant behavior and for dissocial aggressive behavior are four and three dichotomized items related to the respective *DSM-IV* defined disorders. The questionnaire is internally consistent and has a satisfactory discriminant validity (Görtz-Dorten, Ise, Hautmann, Walter, & Döpfner, 2014). The internal consistency of the teacher- and parent-report versions in this study were $\alpha = .860$ and $\alpha = .787$, respectively.

Youth Psychopathic Traits Inventory (Andershed, Kerr, Stattin, & Levander, 2002). This self-rating questionnaire taps psychopathic traits in adolescents (≥12 years old). Each of the 50 items is scored on a 4-point Likert-type scale from 1 (*does*

Table 1. Descriptive Data of the Sample.

	Years	*M*	*SD*	Percentage
Age	6-18	11.58	3.30[a]	
Gender				
Female				51.9
Male				48.1
Nationality				
German				72.4
Other nationality[b]				1.4
Family income				
Net/month under 3,000 Euros				38.7
3,000 Euros and more[c]				61.2
Living				
Village				57.7
Town				38.2
City[d]				4.1

[a]Seven cases without age reference. [b]26.2% Missing information on nationality: In 2015, 22.5% of the population were with a migration background. [c]Based on parents who gave information on the income with 29.2% missing information on income; the average net family income (monthly) for Germany in 2015 was 3,218 Euros (Statistisches Bundesamt, 2017). [d]Based on parents who gave information on the current living situation with 27.9% missing information on the current living situation: In 2015, 22.8% of Germans lived in sparsely populated areas, 41.5% in medium populated, and 35.7% in high-populated areas (Statistisches Bundesamt, 2017).

not apply at all) to 4 (*applies very well*). It comprises 10 scales consisting of five items, respectively (dishonest charm, grandiosity, lying, manipulation, remorselessness, callousness, unemotionality, impulsiveness, thrill seeking, and irresponsibility) that map onto the three domains grandiose/manipulative, callous/unemotional, and impulsive/irresponsible. The items describe psychopathic features as abilities (Kimonis et al., 2008; Skeem & Cauffman, 2003). The YPI has an excellent internal consistency, an acceptable test–retest reliability, and its validity is supported by positive correlations with various forms of conduct problems (Andershed et al., 2002; Colins, Bijttebier, Broekaert, & Andershed, 2014; Pechorro et al., 2016; Poythress, Dembo, Wareham, & Greenbaum, 2006; Skeem & Cauffman, 2003; Stadlin, Perez, Schmeck, Di Gallo, & Schmid, 2015). In the present study, the internal consistency of the total scale was $\alpha = .879$.

Child Behavior Checklist (Achenbach & Rescorla, 2001) and Youth Self-Report (Döpfner et al.,2008). The CBCL (parent-report) and the YSR (self-report for adolescents aged 13- to 18-year-olds) were used to describe internalizing and externalizing behavior. These standardized behavior rating scales quantify emotional and behavioral difficulties in children and adolescents over the past 6 months. Both questionnaires comprise 118 items rated on a 3-point Likert-type scale ranging from 0 (*never or not true*) to 2 (*often or very true*). Items on the anxious/depressed, withdrawn-depressed, and somatic complaints scales are summed to an aggregated internalizing subscale. Items describing rule-breaking and aggressive behavior are aggregated into the externalizing subscale. Both, the CBCL and YSR showed high test–retest reliability, and criterion and construct validity (Achenbach & Rescorla, 2001). In the present study, the internal consistency of the total scale was $\alpha = .955$ for the CBCL and $\alpha = .974$ for the YSR.

Beck Depression Inventory–Second Edition (Hautzinger, Keller, & Kühner, 2009). The BDI-II is a self-report questionnaire with 21 items capturing depressive symptoms over the previous 2 weeks in adolescents/adults aged 13 years or older. Each item is scored on a Likert-type scale of 0 to 3, resulting in a summary score ranging from 0 to 63. The measure is internally reliable and has a good convergent validity (Dozois, Dobson, & Ahnberg, 1998). In the present study, internal consistency for the total scale was $\alpha = .826$.

Screen for Child Anxiety Related Disorders (Birmaher, Kheterpal, Cully, Brent, & McKenzie, 2005). This 41-item self-report questionnaire captures anxiety-related behaviors in children aged 9 to 18 years. Items are scored on a 3-point Likert-type scale from 0 (*not true or hardly ever true*) to 2 (*true or often true*), which are summarized into a total and five subscales. Studies have shown good internal consistency of the total and subscales as well as convergent and discriminant validity (Mittenzwei, 2013; Weitkamp, Romer, Rosenthal, Wiegand-Grefe, & Daniels, 2010). The internal consistency of the total scale in the present study was $\alpha = .863$.

Procedure

This study was approved by the Ethical Committee of the Medical Faculty at Goethe University, Frankfurt, Germany. All participants and their parents gave written informed consent prior to or on the date of their assessment. Adolescents

aged 13 years and older individually completed questionnaires in a classroom setting. Research assistants were available to provide assistance if needed and to ensure independent responding. An individual ID number based on child-specific information was derived for every version to ensure anonymity. Participants were given questionnaires for their parents with a return envelope. Additionally, one teacher of every participating child or adolescent received the teacher questionnaire to individually report on the respective child or adolescent.

Results

Data were analyzed using SPSS (2016; version 23.0). In individuals with missing parent-report versions, adolescents self-reported higher ICU scores, $F(1, 643) = 7.639$, $p < .05$, similarly teacher-reported increased ICU scores, $F(1, 953) = 19.058$, $p < .001$. For individuals with or without self-report or teacher-reports, no differences in the filled in ICU versions were found. No sex differences were observed regarding any of the three ICU versions. The ICU scores of different raters of the parent-report version (mother, father, and both) did not differ.

Item Difficulties

Reliability indices, means, standard deviations, corrected item-scale correlations, and item difficulties of all three ICU versions are shown in supplement Tables S1 to S3 (Supplementary tables are available online). Items 2 and 10 demonstrated the lowest corrected item total correlation in the three versions, ranging from .147 to .178 (Item 2) and −.219 to .103 (Item 10). In the teacher-report version, Item 19 showed a low corrected item total correlation (.037). As the purpose of the study was to validate and extract norms of the three ICU versions, the full item spectrum was retained for further analyses.

Confirmatory Factor Analysis

To examine the factor structure of the ICU, AMOS 23 (Arbuckle, 2014) was used to perform a CFA employing maximum likelihood estimation. Three a priori models were tested for each report version: A model with three correlated factors (Model 1), a bifactorial, three-factor model (Model 2), and a two-factor solution based on a short ICU-version (Model 3; see Introduction for descriptions). Contrary to previous studies (e.g., Essau et al., 2006; Fanti et al., 2009; Houghton et al., 2013; Waller et al., 2015), we did not correlate error terms nor drop any items, in the first two models of the factor analysis, as our study aimed to compare the results of the models based on different raters. Model 3 was calculated only for the 12-item short ICU version proposed by Houghton et al. (2013) and Hawes et al. (2014) as we aimed to compare the fit of the model based on a nonreferred sample with a wide age range (6-18 years). To be able to compare our results with the results of previous studies and to consider multiple fit indices (Lei & Wu, 2007), we calculated all fit indices previously reported (e.g., Essau et al., 2006; Fanti et al., 2009; Houghton et al., 2013; Roose et al., 2010). The following goodness-of-fit indices were calculated: χ^2 exact fit test (Barrett, 2007), χ^2/degrees of freedom [*df*] ratio, the goodness-of-fit index (GFI), the adjusted goodness-of-fit index (AGFI), the comparative fit index (CFI), the root mean square error of approximation (RMSEA), the Akaike information criterion (AIC), and the consistent Akaike information criterion (CAIC). An adequate model fit is indicated by $2 < \chi^2/df \leq 3$, GFI > .90, AGFI > .85, CFI > .95, RMSEA ≤ .10, and small values of AIC and CAIC (Barrett, 2007; Lei & Wu, 2007; Schermelleh-Engel, Moosbrugger, & Müller, 2003). Fit indices for every model and version are presented in Table 2.

Model 1 showed an acceptable fit of all three ICU versions only according to RMSEA. No other fit index indicated an acceptable fit of Model 1. Model 2 showed acceptable values for the parent- and self-report versions according to the GFI and for all three versions according to AGFI and RMSEA. Model 3 showed acceptable fit indices for all three versions based on GFI, AGFI, RMSEA, AIC, and CAIC. For all ICU versions and models, the CFI and the χ^2/df ratio did not reach a good fit, and the χ^2 goodness-of-fit test failed to fit in any of the three versions of the three models.

Reliability

Reliability analysis was performed for the ICU total version. The internal consistency was examined via Cronbach's alpha with reliability coefficients >.8 considered good and >.7 as acceptable (Blanz, 2015). The internal consistency of the parent- ($\alpha = .830$) and teacher-report versions ($\alpha = .877$) was good, and acceptable for the self-report version ($\alpha = .769$).

To examine the test–retest reliability, a subsample of raters (individuals, parents, and teachers) received the same ICU version after a period of up to 15 months after the first data collection (Range = 6.10-15.52 months; $M = 12.67$; $SD = 2.67$). High correlations (Spearman's rho) for the corresponding measures of the summary score were found for the self-report ($r_s = .584$, $p < .05$, $n = 90$), the parent-report ($r_s = .710$, $p < .05$, $n = 162$), and the teacher-report versions ($r_s = .523$, $p < .05$, $n = 87$).

To examine the interrater reliability, we applied a two-way random effects ICC model with absolute agreement. The average ICC for all scales was .614, $F(323, 646) = 2.632$, 95% confidence interval [CI: .535, .682], $p < .001$. The ICC of the self-report and parent-report version was .525, $F(391, 391) = 2.167$, 95% CI [.416, .614], $p < .001$, the parent- and teacher-report version showed an ICC of

Table 2. Model-Fits for the Three ICU Versions (Parent, Teacher, and Self).

Model	χ^2/df	GFI	AGFI	CFI	RMSEA	AIC	CAIC	χ^2
Parent-report version								
1	6.576	.867	.836	.795	.075	1448.556	1726.455	1354.556**
2	4.009	.927	.903	.880	.055	1057.995	1483.712	913.995**
3	4.447	.961	.943	.919	.059	285.667	433.486	235.667**
Teacher-report version								
1	10.041	.757	.708	.797	.097	2602.238	2901.185	2500.238**
2	5.972	.890	.855	.898	.072	1505.729	1927.773	1361.729**
3	6.503	.939	.910	.930	.076	394.675	541.217	344.675**
Self-report version								
1	4.651	.847	.815	.676	.075	1259.994	1538.926	1157.994**
2	3.060	.912	.884	.833	.057	841.670	1235.456	697.670**
3	4.708	.937	.907	.840	.076	299.539	436.270	249.539**

Note. df = degrees of freedom; GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index; RMSEA = root mean square error of approximation; AIC = Akaike information criterion; CAIC = consistent Akaike information criterion. Model 1: model with three correlated factors, Model 2: bifactorial three-factor model, and Model 3: two-factor model.
**$p<.001$.

.522, $F(756, 756) = 2.103$, 95% CI [.449, .586], $p < .001$. The agreement between self- and teacher-report version was poor, with an ICC of .488, $F(453, 453) = 1.968$, 95% CI [.385, .574], $p < .001$.

Convergent and Divergent Validity

To test the convergent validity, the three ICU total versions and the three subscales (i.e., callousness, unemotional, and uncaring) were correlated with the YPI total score, the YPI subscales (callousness, remorseless, and unemotionality), the externalizing scores of the YSR and the CBCL and with the FBB-SSV (parent- and teacher-report version). Assuming that the three subscales are intercorrelated, correlations were computed showing correlations and partial correlations, controlling for the effect of the other two ICU subscales. Within each set of correlations, we controlled for the family-wise error rate and set an $\alpha_{pc} = .017/.003$. For interpretation of effect sizes, we applied three different ρ values: $\rho = .1$ as a small, $\rho = .3$ as a medium, and $\rho = .5$ as a large effect size. Results are presented in Table 3.

Regarding the total score of the self-report version, positive correlations with a medium to large effect size were found with the YPI total score, with the callousness and the remorseless subscales of the YPI and with the externalizing scale of the YSR. The ICU total score parent-report version correlated positively with medium effect sizes with the YPI total score, with the externalizing scale of the CBCL and with the parental and teacher assessments of the FBB-SSV oppositional aggressive. The callousness subscale showed correlations in the same direction and in similar size as the ICU total score. However, the callousness subscale of the ICU teacher-report version failed to have significant correlations with the YPI and its subscales callousness and remorselessness, the externalizing scale of the YSR and the parent-report version of the FBB-SSV dissocial. The uncaring subscale also showed many correlations in the same direction and of similar size as the ICU total score, especially for the ICU teacher-report version. Compared with the ICU total score, the unemotional subscale often correlated differently, sometimes even negatively, with the external scales, however.

To examine the divergent validity of the ICU, the three ratings of the total scale with the subscales (i.e., callousness, unemotional, and uncaring) were correlated with the BDI-II, the SCARED and the internalizing scores of the YSR, and the CBCL. Similar to the convergent validity calculation, Spearman`s ρ was applied, correlations and partial associations were calculated, the family-wise error rate was controlled via an $\alpha_{pc} = .017/.003$ and effect sizes were calculated. Results are shown in Table 4.

The total scale of the ICU parent-report version correlated positively with medium effect-size with the internalizing scale of the CBCL. The unemotional subscale mostly showed correlations in the same direction and in similar size as the ICU total score. However, a significant correlation of small effect-size was present between the self-report version of the unemotional subscale and the SCARED. No significant correlation existed between the teacher-report version of the unemotional subscale and the internalizing scale of the CBCL. Unlike the total score, the correlations between the self-report versions of the uncaring and callousness subscales and the BDI-II were insignificant, between the self-report version of the uncaring subscale and the internalizing scale of the YSR was insignificant, between the self-report version of the callousness subscale and the internalizing scale of the YSR was negative and between the self-report version of the uncaring subscale and the SCARED was negative.

Table 3. Correlations and Partial-Correlations[a] Between the ICU Versions (With Individual Factors) and the YPI, the Externalizing Scores of the YSR and CBCL and the Oppositional Aggressive Scale of the FBB-SSV.

	ICU self-report version				ICU parent-report version				ICU teacher-report version			
	Cal	Une	Unc	Total	Cal	Une	Unc	Total	Cal	Une	Unc	Total
YPI	.547** / .498**	.144** / −.025	.380** / .262**	.500**	.351** / .292**	.051 / −.069	.273** / .126	.302**	.254** / .098	−.069 / −.150*	.298** / .223**	.238**
YPI Callousness	.279** / .242**	.132** / .055	.284** / .218**	.333**	.105 / .138*	.053 / −.041	.079 / .018	.093	.137* / .014	−.028 / −.082	.201** / .180**	.151**
YPI Remorselessness	.473** / .437**	.115* / .002	.347** / .255**	.445**	.224** / .246**	.085 / −.041	.177* / .048	.212**	.195** / .068	−.074 / −.144	.255** / .216**	.195**
YPI Unemotionality	.299** / .290**	.145** / .104*	.057 / −.079	.229**	.133* / .154*	.072 / .057	.007 / −.079	.087	.069 / .007	−.007 / −.047	.105* / .096	.075
YSR ext.	.368** / .342**	.144** / .016	.332** / .218**	.386**	.201** / .128*	.014 / −.097	.247** / .165**	.212**	.197** / .093	−.184** / −.246**	.210** / .177**	.140*
CBCL ext.	.173** / .177**	−.017 / −.080	.244** / .167**	.207**	.431** / .330**	.080* / −.069	.438** / .269**	.443**	.213** / .114**	−.065 / −.143**	.283** / .191**	.229**
FBB-SSV opp/agg (parent)	.211** / .119	.027 / −.064	.201** / .142*	.211**	.384** / .277**	.117** / −.059	.402** / .276**	.421**	.209** / .106*	.037 / −.058	.261** / .157**	.253**
FBB-SSV diss. (parent)	.131* / .183**	.049 / −.001	.038 / −.076	.116	.311** / .273**	.049 / −.012	.262** / .049	.286**	.124** / .024	−.041 / −.036	.158** / .117**	.125**
FBB-SSV opp/agg (teacher)	.169** / .124*	.060 / −.031	.189** / .114	.194**	.212** / .181**	−.044 / −.087	.173 / .110*	.166**	.423** / .321**	−.140** / −.277**	.451** / .204**	.383**
FBB-SSV diss. (teacher)	.164** / .107	.063 / −.024	.154* / .102	.174**	.155** / .152**	−.028 / −.081	.145** / .068	.155**	.305** / .263**	−.090* / −.150**	.321** / .114**	.285**

Note. ICU = Inventory of Callous–Unemotional Traits; Cal = Callousness; Unc = Uncaring; Une = Unemotional; YPI = Youth Psychopathy Inventory; YSR ext. = externalizing score of the Youth Self-Report; CBCL ext. = externalizing score of the Child Behavior Checklist; FFB-SSV opp/agg (parent) = parent-version of the oppositional aggressive scale of the Observer Rating Scale for Conduct Disorders; FFB-SSV diss. (parent) = parent-version of the dissocial scale of the Observer Rating Scale for Conduct Disorders; FFB-SSV opp/agg (teacher) = teacher-version of the oppositional aggressive scale of the Observer Rating Scale for Conduct Disorders; FFB-SSV diss. (teacher) = teacher-version of the dissocial scale of the Observer Rating Scale for Conduct Disorders.

[a]Partial correlations, controlling for other ICU scales, if applicable, are presented after the slash.

*Correlation significant at the .017 level. **Correlation significant at the .003 level.

Table 4. Correlations and Partial Correlations[a] Between the ICU Versions (With Individual Factors) and the BDI-II, the SCARED, and the Internalizing Scales of the YSR and the CBCL.

	ICU self-report version				ICU parent-report version				ICU teacher-report version			
	Cal	Une	Unc	Total	Cal	Une	Unc	Total	Cal	Une	Unc	Total
BDI-II	.157** / .088	.333** / .324**	.117* / −.024	.268**	.054 / .009	.107 / .083	.097 / .051	.094	.046 / −.003	.023 / .002	.093 / .056	.081
SCARED	−.042 / −.069	.196 / .236**	−.084 / −.115*	.011	−.011 / .065	.077 / .102	−.037 / −.108	.003	−.054 / −.058	.036 / .075	−.013 / .003	−.013
YSR int.	.093 / .047	.302** / .313**	.028 / −.101*	.175**	.001 / −.040	.100 / .105	.037 / .034	.046	.011 / .037	.077 / .108	.007 / −.045	.035
CBCL int.	.045 / −.002	.092 / .108	.069 / .004	.111	.248** / .196**	.288** / .250**	.219 / .021	.319**	.116** / .038	.118** / .087	.134 / .080	.169**

Note. ICU = Inventory of Callous–Unemotional Traits; Cal = Callousness; Unc = Uncaring; Une = Unemotional; BDI-II = Beck Depression Inventory–Second edition; SCARED = Screen for Child Anxiety Related Disorders; YSR int. = internalizing score of the Youth Self-Report; CBCL int. = internalizing score of the Child Behavior Checklist.

[a]Partial correlations, controlling for other ICU scales, if applicable, are presented after the slash.

*Correlation significant at the .017 level. **Correlation significant at the .003 level.

Table 5. Partial Correlations Between the ICU Versions (With Individual Factors) and the BDI-II, SCARED, Internalizing Scales of the YSR, CBCL Controlling for the Parent-Version of the Oppositional Aggressive scale and the Dissocial Scale of the FBB-SSV, and other ICU Scales[a].

	ICU self-report version				ICU parent-report version				ICU teacher-report version			
	Cal	Une	Unc	Total	Cal	Une	Unc	Total	Cal	Une	Unc	Total
BDI-II	.110	.357**	.036	.359**	−.017	.064	.036	.065	.058	−.055	.016	.054
SCARED	−.056	.253**	−.091	.055	.013	.089	−.006	.009	−.027	.091	.018	.033
YSR int.	.033	.350**	−.129	.172**	−.037	.083	.036	.060	.062	.089	−.083	.013
CBCL int.	−.057	.130*	−.040	.010	.067	.284**	−.043	.211**	−.004	.126**	.033	.100*

Note. ICU = Inventory of Callous–Unemotional Traits; Cal = callousness; Unc = uncaring; Une = unemotional; BDI-II = Beck Depression Inventory–Second edition; SCARED = Screen for Child Anxiety Related Disorders; YSR int. = internalizing score of the Youth Self-Report; CBCL int. = internalizing score of the Child Behavior Checklist; FBB-SV = Observer Rating Scale for Conduct Disorders.
[a]If applicable.
*Correlation significant at the .017 level. **Correlation significant at the .003 level.

Table 6. Age- and Gender-Specific Means and Standard Deviations of the Three ICU Versions (Parent, Self, and Teacher).

	ICU parent-report version				ICU self-report version				ICU teacher-report version			
	Male		Female		Male		Female		Male		Female	
Age	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
6-10 Years	18.13	7.32	15.82	7.16					21.54	9.38	16.14	9.39
11-14 Years[a]	21.58	8.15	17.65	7.43	23.64	7.51	20.58	7.28	24.14	10.70	18.68	8.50
15-18 Years	21.68	10.42	18.25	8.03	24.64	7.50	21.27	8.27	23.16	12.14	17.95	8.04

[a]Applies to the self-rating of children aged 13 to 14 years.

Furthermore, the correlations between the teacher-report versions of the uncaring and callousness subscales and the CBCL internalizing differed (compared with the ICU total scale) and between the parent-report version of the uncaring subscale and the CBCL internalizing.

To rule out that CD symptoms explained the correlations, we additionally calculated partial correlations with the parent rated, FBB-SSV-based oppositional defiant and dissocial aggressive behavior as possible confounding variables. As shown in Table 5, results did not change for the ICU total score, but for the subscales. The parent-rated callousness subscale as well as the teacher-rated unemotional subscale did not show the previously observed negative correlation with the CBCL-derived internalizing scale.

Norms

To calculate norms, we first controlled for age and sex effects on the ICU ratings. This was accomplished by performing analyses of variance with age and sex as the independent and ICU score as the dependent variables. To guarantee approximately equivalent age spectrums, categories were defined: 6 to 10 years, 11 to 14 years, and 15 to 18 years. Table 6 displays the age- and sex-specific means and standard deviations of the three ICU versions. For the ICU parent-report version, a main-effect of sex, $F(1, 1001) = 34.989$, $p < .001$, was found, with boys having higher ICU scores than girls. Also, children had lower scores compared with the other two groups, categorized; $F(2, 1001) = 15.899$, $p < .001$. Hence, we developed age- and sex-specific norms for the total scale of the ICU parent-report version. For the ICU self-report version, again a main-effect of sex, $F(1, 640) = 28.115$, $p < .001$, but not for age, with higher ICU summary scores in boys compared with girls was found. Therefore, we developed sex-specific norms for the ICU self-report version. For the ICU teacher-report version, a main-effect of sex, $F(1, 943) = 66.649$; $p < .001$ was found, with boys having higher ICU scores than girls. Also, children aged 6 to 10 years had lower ICU scores than children aged 11 to 14 years, categorized; $F(2, 943) = 6.710$, $p < .05$. Post hoc tests for the age groups revealed that ICU data of all versions (self, parent, and teacher) were not normally distributed, therefore, we report percentile rank and Stanine-norms (for an overview, see supplementary Tables S4-S8).

Discussion

This study aimed at examining the factor structure, reliability, and validity of the German ICU self-, parent-, and teacher-report versions. Furthermore, we developed norms based on the data of a large community sample of children and adolescents aged 6 to 18 years. The article adds to previous research

in several respects: First, this study covers a wider age spectrum (6-18 years) than previous studies; second, it jointly studied the three ICU versions (self, parent, and teacher) with reports on the same individuals; and third, it provides German norms for the ICU. This allows a German population-based comparison of ICU derived behaviors as basis to specify the new *DSM-5*-based LPE CD classifier.

Item analysis replicated previous findings with Items 2 and 10 showing the lowest item-total correlations (Henry et al., 2016; Kimonis et al., 2007; Munoz, Frick, Kimonis, & Aucoin, 2008; Ray, Frick, et al., 2016). As the purpose of the study was to validate and extract norms of the three ICU ratings, all items were included in the following analyses.

Regarding the results of the factor analysis, the best fit for all three ICU versions was found for the recently proposed two-factor (uncaring and callousness) model based on the short ICU version with 12 items (Carvalho et al., 2018; Colins et al., 2016; Hawes et al., 2014; Horan et al., 2015; Houghton et al., 2013; McDonald et al., 2018; Roose et al., 2010; Waller et al., 2015). However, none of the versions fit well when applying the χ^2 goodness-of-fit test, which has been discussed as "the only substantive test of fit" (Barrett, 2007). It might be argued that this is due to the large sample size, increasing the probability that a model will fail to fit when applying the χ^2 goodness-of-fit test. However, alternative explanations have to be taken into account as we relied on a moderate sample size (Barrett, 2007). The data were not multivariate normally distributed, and boot strapping did not improve the distribution; therefore, one possible option would have been to choose different CFA models. We refrained from adjusting the models, as our aim was to compare our results with models found in the literature. Furthermore, there are many obstacles to reaching a good model fit. Often personality trait inventories have shown a rather poor model fit when evaluated via CFA because of possible artefacts resulting from item wording, items tapping minor sources of variation, items having secondary factor loadings, and because of correlated residuals (Hopwood & Donnellan, 2010). Accordingly, many international studies focusing the ICU's factor structure had difficulties adequately fitting the model (Essau et al., 2006; Feilhauer et al., 2012; Gao & Zhang, 2016; Houghton et al., 2013; Kimonis et al., 2008; Lopez-Romero et al., 2015; Roose et al., 2010; Waller et al., 2015) and required post hoc modifications (e.g., dropping of items, correlated error terms, and explorative factor analysis to detect new factor structure models). Similar to our results, the best model fit was observed by studies on the short ICU version with 12 items, which resulted in the two factors uncaring and callousness, which may be used in further studies on the LPE CD specifier.

The internal consistency of the total score of the self- and parent-report version was acceptable, similar to that found in earlier international studies and other ICU versions (Berg et al., 2013; Eisenbarth et al., 2016; Essau et al., 2006; Fanti et al., 2009; Feilhauer et al., 2012; Henry et al., 2016; Herpers et al., 2017; Kimonis et al., 2015; Kimonis et al., 2016; Levy et al., 2017; Lopez-Romero et al., 2015; Marsee et al., 2011; McDonald et al., 2018; Pechorro et al., 2016; Ray, Pechorro, et al., 2016; Roose et al., 2010; White et al., 2013), indicating accurate measurement of underlying constructs. Strong internal consistency was proven for the total score of the teacher-report version, expanding on Levy et al.'s (2017) results to nonreferred children and adolescents of both genders outside of Israel. In line with the results from Cardinale and Marsh (2020), the other-report versions of the total score of the ICU had a higher internal consistency than the self-report version. Thus, our results underscore that regardless of which ICU version is applied, the ICU covers one dimension (CU).

Our data supports the high test–retest reliability of the total version found in earlier studies (Berg et al., 2013; Feilhauer et al., 2012; Kimonis et al., 2014; Moore et al., 2017). In contrast to past studies, which were much shorter in duration, our test–retest period was on average 12.7 months, supporting the high stability of CU traits. Furthermore, we could confirm a high retest reliability for the teacher-report version. Interestingly, the parent-report version had a higher test–retest reliability than the teacher- and self-report version. One reason could be that the teacher-report version might have been realized by different teachers or by teachers who have only known the child for a short time; however, resulting in an impressive teacher test–retest reliability. Another reason could be differences in setting (home vs. school), which could result in different displays of CU traits. For example, peer interactions and pressure at school may have resulted in a more obvious manifestation of CU traits in the school setting. Furthermore, the criteria "unconcern about performance" (American Psychiatric Association, 2013) will be more apparent in the school context, favoring the high test–retest reliability of the teacher-report version.

Our study was characterized by a moderate interrater reliability. Along with the results of Docherty et al. (2016), we found a higher interrater reliability of the self- and parent-report than of the self- and teacher-report. This is probably due to a closer relationship of parents and their children as compared with teachers. The high parent–teacher interrater-reliability is likely due to the fact that adults are more aware of certain CU aspects than children.

In terms of convergent validity, we found a positive correlation of all ICU versions with the YPI and with the externalizing scales of the YSR and the CBCL, as expected (Colins et al., 2016; Essau et al., 2006; Pechorro et al., 2016; Waller et al., 2015). The stronger correlations between the ICU self-report version (compared with the parent- and teacher-report version) and the YPI and the externalizing scale of the YSR likely is attributable to the self-reports implemented. The same holds true for the stronger correlation of the ICU parent-report version

(compared with the self-report version) and the externalizing scale of the CBCL (compared with the YSR). The parent-report versions of the FBB-SSV oppositional/defiant and dissocial/aggressive positively correlated with the ICU parent- and teacher-report versions, but not the self-report version. This may be due to the fact that adults seem to better judge the *DSM-5*-based ODD and CD behaviors and CU symptomatology than the children or adolescents themselves (e.g., Gao & Zhang, 2016). Self-reporting provides insights that are especially important for assessing adolescents and thus should be complemented by parent- and teacher-assessments resulting in a multi-informant perspective as recommended by the *DSM-5* CD specifier "with limited prosocial emotions" (American Psychiatric Association, 2013).

Regarding the subscales of the ICU, the callousness and uncaring subscales primarily correlated in the same direction and in comparable size as the total score. Correlations of the unemotional subscale substantially differed, underlining its questionable reliability (Gao & Zhang, 2016; Latzman, Malikina, Hecht, Lilienfeld, & Chan, 2016) and low convergent validity (Charles et al., 2012; Essau et al., 2006; Kimonis et al., 2008; Lopez-Romero et al., 2015; Ray, Pechorro, et al., 2016; Roose et al., 2010; Waller et al., 2015) and whether it even captures what "unemotional" is, as it is defined by the nomological network of the LPE specifier (Waller et al., 2015). Accordingly, one study suggested the unemotional subscale to rather represent the low end of positive affectivity (Latzman et al., 2013). The unique associations of the subscales with external criteria go in line with different associations of the ICU factors with personality traits (Decuyper et al., 2011; Latzman et al., 2013). This may indicate that CU is a multidimensional construct, what might be obscured when just combining the subscales into a total ICU score.

According to the theory of Frick, CU traits are characterized by low levels of arousal and therefore associated with lower levels of fear and anxiety (Frick et al., 2014). However, in line with other studies (Essau et al., 2006; Fanti et al., 2013; Horan et al., 2015; Waller et al., 2015), we failed to find negative correlations between any ICU version and the SCARED. Instead, in line with previous studies (Berg et al., 2013; Fink et al., 2012; Gao & Zhang, 2016) and Fink et al.'s (2012), assumption that the affective presentation of CU is similar to that of depression in adolescence, we found positive correlations between the ICU self-report version and depression (BDI-II), as well as self-reported internalizing symptoms (YSR). Furthermore, we found positive correlations between the ICU parent- and teacher-report versions with parent-reported internalizing symptoms (CBCL). This finding again is in contrast to Frick's theory (Frick, Lilienfeld, Ellis, Loney, & Silverthorn, 1999) and replicates previous results observed for the self-report version and extends them to the parent- and teacher-report version of the ICU in children aged 6 to 18 years. Some authors (e.g., Berg et al., 2013) interpreted the positive associations of CU traits and internalizing symptoms to reflect negative emotionality. Consequently, children and adolescents with high ICU scores would experience negative emotions, and be characterized by an emotional and interpersonal distance from others (Latzman et al., 2013), not a lack of emotionality (i.e., "unemotional"; Latzman et al., 2013). Other authors proposed to distinguish a primary versus a secondary variant of CU traits in adolescents (Kahn et al., 2013). Adolescents with the primary variant are suggested to be characterized by a low level of anxiety and depression. Individuals with the secondary variant are suggested to show a high level of anxiety, depression, impulsive aggression, an increased risk for self-harming behavior (such as drug use, unprotected sex), a history of trauma, and attachment disorganization (Cecil, McCrory, Barker, Guiney, & Viding, 2018; Kahn et al., 2013).

Regarding divergent validity of the subscales, the unemotional subscale of the ICU showed comparable correlations as the ICU total score. It seems likely that the unemotional subscale also drives the respective correlation of the ICU summary score with the different measures on internalizing behavior (such as the BDI-II, internalizing scales of the CBCL, and the YSR). In contrast, the uncaring and callousness subscales did show a differential correlation with self- and parent-reported internalizing behavior compared with the ICU summary score. Given that the callousness and uncaring scales showed higher correlations with other measures of aggression or other psychopathy and the unemotional scales primarily correlated with measures of emotional functioning (Feilhauer et al., 2012; Kimonis et al., 2008; Lopez-Romero et al., 2015; Roose et al., 2010), the subscales seem to capture rather differential behaviors. However, limiting the interpretations of the divergent validity is the fact that effect sizes of correlations were mostly small.

In accordance with previous studies, we found sex differences for all three ICU versions, with girls scoring lower than boys (Docherty et al., 2016; Essau et al., 2006; Fanti et al., 2013; Henry et al., 2016; Kimonis et al., 2015; Marini & Stickle, 2010; Stickle et al., 2009; Viding et al., 2009). This underlines the importance of conducting sex-specific research on CU traits. Studies have reported sex-differences in CD, for example, girls are more likely to endorse nonconfrontational behavior such as stealing, lying, and running away (overview: Berkout, Young, & Gross, 2011; Freitag, Boomsma, Glennon, Franke, & Holtel, 2018). In addition, Charles et al. (2012) reported that increased CU traits among girls are related to more adjustment problems compared with boys. Girls may also express CU traits somewhat differently to boys in ways not adequately covered by the questionnaire. Taken together, our findings clearly indicate that sex-specific norms should be applied to ensure that girls will be diagnosed correctly as basis for specific interventions taking the *DSM-5*-based LPE specifier into account.

In this study, we found increased CU traits in older children for the ICU parent- and teacher-report version, but not for the ICU self-report version. The main reason for the missing age-effect of the self-report is the age of the sample. In the parent-report version, children aged 6 to 10 years had lower ICU scores than children aged 11 to 14 and 15 to 18 years. In the teacher-report version, children aged 6 to 10 years had lower ICU scores than children aged 11 to 14 years. The results imply that children may exhibit fewer CU symptoms and be at a lower risk of ICU captured CU traits at age 6 to 10 years. Former studies showed divergent results, such as no correlation of age and ICU score (Moore et al., 2017), higher ICU scores in 7- to 10-year-olds compared with 11- to 17-year-olds (Carvalho et al., 2018) or highest scores in 15- to 16-year-olds compared with 13- to 14-year-olds and 17- to 18-year-olds (Essau et al., 2006). However, our study differs from earlier studies by covering the three ICU versions. Essau et al. (2006) only studied the self-report version of the ICU and did not cover children aged 6 to 12 years. Carvalho et al. (2018) only examined the ICU self-report version in children aged 7 to 17 years. Self-reporting in younger children (aged 6-12 years) was not studied here, so we cannot fully exclude possible age effects for the self-report version.

Strengths, Limitations, and Future Directions

Strengths of this investigation include the large sample size, the inclusion of multiple age groups and the corroboration of findings across informants. Nevertheless, the findings must be interpreted in light of the following limitations.

First, external validation was based solely on questionnaire data. Other correlates that are important to the construct of CU traits, such as cognitive, emotional, and biological variables as well as peer-nomination procedures and expert ratings of CU traits (e.g., PCL:YV; Forth et al., 2003), which have been shown to be positively correlated with CU traits (Feilhauer et al., 2012; Fink et al., 2012; Graziano et al., 2016; Harenski et al., 2014; Johnson et al., 2013; Kongerslev et al., 2015), should be considered when validating measures of this construct. Second, we were able to derive norms based on the nonreferred sample, but it should be considered that the suitability of the ICU for individual diagnostics should be proven with an ROC-analysis. This can only be achieved via clear cutoff scores based on clinical and forensic samples. Third, the study was based on a school sample. The working group of Docherty et al. (2016) found differing response styles when comparing a high school with a detained sample (e.g., higher agreement between adolescents and parents in the high school compared with the incarcerated sample). Furthermore, only adolescents attending school were assessed. Therefore, selection effects cannot be ruled out, especially as school absenteeism is associated with CD and CU traits. Future studies are needed to compare ICU scores in various settings (nonreferred, clinical, institutionalized individuals). Fourth, participation in the study relied on the approval of the director of the school, no incentives were provided and participation was time consuming. Therefore, selection effects (e.g., of schools with more active directors and teachers) cannot be ruled out. Participation rates per school varied accordingly dependent on the willingness of the director, how many classes she or he would allow to participate, and how many teachers were willing to participate. Specific school selection effects, hence, cannot be ruled out. Fifth, the study was based on a German sample with the German translation of the ICU, restricting the generalizability of the results to other countries. Further studies should examine the three versions in non-Western countries. Sixth, the overrepresentation of Germans without a migration background and living in rural areas also somewhat limits the representativeness of the study and the generalizability of the results. Seventh, as parents, who did not fill in the questionnaire, had children with higher ICU scores in the teacher- and self-report version, a selection bias for the parent data is evident. It can be speculated that the results based on the current parent data are rather positive. The interrater reliability would likely have been lower if the parent-report version of children with higher CU scores were available. Eight, the handling of missing data may have resulted in a reduced standard error (overestimating the precision of the results) and thus in potentially biased results. It also may have weakened the correlation estimates due to lower variability. Omitting data on individuals with many missing values may also have led to discarce important information especially on the more strongly affected individuals (Pederson et al., 2017). Finally, due to the covered age-range of the applied measures (e.g., YPI), the ICU self-report was only obtained from children aged 13 years and older, not allowing to generalize the results to younger children.

Conclusion

This study supports previous findings on the ICU as a promising assessment instrument to measure CU traits, which are part of the *DSM-5*-based LPE specifier of CD. This is the first study proposing norms for the German version of the ICU. These age and sex-specific norms now need to be applied in the scientific and clinical context to specify individual results based on these norms.

Future research should focus on establishing valid cutoff scores for the description of clinically relevant CU traits. Similar to Kimonis et al.'s (2016) proposal, we suggest that the ICU can be used to screen adolescents for increased CU traits, who may need a more thorough assessment of CD and its specifiers. It also may be used as a screener to identify children and adolescents at greater risk of future CD-related behavior, and to apply intense and individualized prevention

of the disorder. As the best model fit was found for the short 12 items ICU version, this may be used in further studies on the LPE CD specifier.

Acknowledgments

We sincerely thank the staff of the participating schools and all participants who made this research possible. Mrs. Feike and Mrs. Eisenbein supported data-collection, and Mr. Zerlaut provided the study database.

Declaration of Conflicting Interests

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: CMF receives royalties for books on ADHD, ASD, and MDD, and has served as consultant to Desitin and Roche regarding ASD. CS receives royalties for an intervention manual. The other authors declare no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The research was partly funded by the European Community, project FemNAT-CD (coordinator CM Freitag), FP7 Grant no. 602407.

Supplemental Material

Supplemental material for this article is available online.

ORCID iD

Kathrin Ueno https://orcid.org/0000-0002-7468-4279

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles: An integrated system of multi-informant assessment*. Burlington, VT: ASEBA.

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

Anastassiou-Hadjicharalambous, X., & Warden, D. (2008). Physiologically-indexed and self-perceived affective empathy in Conduct-Disordered children high and low on Callous-Unemotional traits. *Child Psychiatry and Human Development*, *39*, 503. doi:10.1007/s10578-008-0104-y

Andershed, H., Kerr, M., Stattin, H., & Levander, S. (2002). Psychopathic traits in non-referred youths: A new assessment tool. In E. Blauuw & L. Sheridan (Eds.), *Psychopaths: Current international perspectives* (pp. 131-158). The Hague, Netherlands: Elsevier.

Arbuckle, J. L. (2014). Amos (Version 23) [Computer software]. Chicago, IL: IBM.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, *42*, 815-824.

Berg, J. M., Lilienfeld, S. O., Reedy, S. D., Latzman, R. D., Roose, A., Craighead, L. W., . . . Raison, C. S. (2013). The Inventory of Callous and Unemotional Traits: A construct-validational analysis in an at-risk-sample. *Assessment*, *20*, 532-544.

Berkout, O. V., Young, J. N., & Gross, A. M. (2011). Mean girls and bad boys: Recent research on gender differences in conduct disorder. *Aggression and Violent Behavior*, *16*, 503-511.

Birmaher, B., Kheterpal, S., Cully, M., Brent, D., & McKenzie, S. (2005). Screen for Child Anxiety Related Disorders (SCARED): In L. VandeCreek & T. L. Jackson (Eds.), *Innovations in clinical practice: Focus on children and adolescents* (pp. 99-104). Sarasota, FL: Professional Resources Exchange.

Blanz, M. (2015). *Research methods and statistics for the social work: basics and applications*. Stuttgart, Germany: Kohlhammer.

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

Carvalho, M., Faria, M., Conceicao, A., Gaspar de Matos, M., & Essau, C. A. (2018). Callous-Unemotional Traits in children and adolescents. *European Journal of Psychological Assessment*, *34*, 101-110.

Cecil, C. A. M., McCrory, E. J., Barker, E. D., Guiney, J., & Viding, E. (2018). Characterising youth with callous–unemotional traits and concurrent anxiety: Evidence for a high-risk clinical group. *European Child & Adolescent Psychiatry*, *27*, 885-898.

Charles, N. E., Acheson, A., Mathias, C. W., Furr, R. M., & Dougherty, D. M. (2012). Psychopathic traits and their association with adjustment problems in girls. *Behavioral Sciences & the Law*, *30*, 631-642.

Colins, O. F., Andershed, H., Hawes, S. W., Bijttebier, P., & Pardini, D. A. (2016). Psychometric properties of the original and short form of the Inventory of Callous-Unemotional Traits in detained female adolescents. *Child Psychiatry and Human Development*, *47*, 679-690.

Colins, O. F., Bijttebier, P., Broekaert, E., & Andershed, H. (2014). Psychopathic-like traits among detained female adolescents: Reliability and validity of the Antisocial Screening Device and the Youth Psychopathic Traits Inventory. *Assessment*, *21*, 195-209.

Decuyper, M., De Bolle, M., De Fruyt, F., & De Clercq, B. (2011). General and maladaptive personality dimensions and the assessment of callous-unemotional traits in adolescence. *Journal of Personality Disorders*, *25*, 681-701. doi:10.1521/pedi.2011.25.5.681

Decuyper, M., De Caluwe, E., De Clercq, B., & De Fruyt, F. (2014). Callous-Unemotional traits in youth from a *DSM-5* trait perspective. *Journal of Personality Disorders*, *28*, 334-357. doi:10.1521/pedi_2013_27_120

Docherty, M., Boxer, P., Huesmann, L. R., O`Brien, M., & Bushman, B. (2016). Assessing callous-unemotional traits in adolescents: Determining cutoff scores for the Inventory of Callous and Unemotional Traits. *Journal of Clinical Psychology*, *73*, 257-278.

Döpfner, M., Görtz-Dorten, A., & Lehmkuhl, G. (2008). *DISYPS-II: Diagnostic System for Mental Disorders in Childhood and Adolescence II*. Bern, Lower Saxony, Germany: Huber.

Dozois, D. J. A., Dobson, K. S., & Ahnberg, J. L. (1998). A psychometric evaluation of the Beck Depression Inventory-II. *Psychological Assessment*, *10*, 83-89.

Edens, J. F., Campbell, J. S., & Weir, J. M. (2006). Youth psychopathy and criminal recidivism: A meta-analysis of the psychopathy checklist measures. *Law and Human Behavior*, *31*, 53-75.

Eisenbarth, H., Demetriou, C. A., Kyranides, M. N., & Fanti, K. A. (2016). Stability subtypes of callous–unemotional traits and conduct disorder symptoms and their correlates. *Journal of Youth and Adolescence*, *45*, 1889-1901.

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

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*, 964-979.

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*, 285-298.

Feilhauer, J., Cima, M., & Arntz, A. (2012). Assessing callous-unemotional traits across different groups of youths: Further cross-cultural validation of the Inventory of Callous-Unemotional Traits. *International Journal of Law and Psychiatry*, *35*, 251-262.

Fink, B. C. (2010). *Depressive symptoms predict callous-unemotional behavior in adjudicated adolescent males with conduct disorder* (Doctoral dissertation). Retrieved from https://digitalrepository.unm.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1041&context=psy_etds

Fink, B. C., Tant, A. S., Tremba, K., & Kiehl, K. A. (2012). Assessment of psychopathic traits in an incarcerated adolescent sample: A methodological comparison. *Journal of Abnormal Child Psychology*, *40*, 971-986.

Forth, A., Kosson, D. S., & Hare, R. D. (2003). *Hare PCL: Youth version (PCL: YV) rating booklet*. Toronto, Ontario, Canada: MHS.

Freitag, C. M., Boomsma, D., Glennon, J. C., Franke, B., & Holtel, A. (2018). Focused issue on conduct disorder and aggressive behaviour. *European Child & Adolescent Psychiatry*, *27*, 1231-1234.

Frick, P. J. (2004). *The Inventory of Callous-Unemotional Traits* (Unpublished rating scale). University of New Orleans, New Orleans, LA.

Frick, P. J., Cornell, A. H., Barry, C. T., Bodin, S. D., & Dane, H. E. (2003). 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*, 457-470.

Frick, P. J., & Dickens, C. (2006). Current perspectives on conduct disorder. *Current Psychiatry Reports*, *8*, 59-72.

Frick, P. J., Lilienfeld, S. O., Ellis, M., Loney, B., & Silverthorn, P. (1999). The association between anxiety and psychopathy dimensions in children. *Journal of Abnormal Child Psychology*, *27*, 383-392.

Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Annual research review: 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.

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*, 331-340.

Görtz-Dorten, A., Ise, E., Hautmann, C., Walter, D., & Döpfner, M. (2014). Psychometric properties of a German parent rating scale for oppositional defiant and conduct disorder (FBB-SSV) in clinical and community samples. *Child Psychiatry & Human Development*, *45*, 388-397.

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*, 201-214.

Harenski, C., Harenski, K., & Kiehl, K. A. (2014). Neural processing of moral violations among incarcerated adolescents with psychopathic traits. *Developmental Cognitive Neuroscience*, *10*, 181-189.

Hautzinger, M., Keller, F., & Kühner, C. (2009). *BDI-II. Beck-depressions-inventar: Revision*. Frankfurt, Germany: Pearson Assessment.

Hawes, S. W., Byrd, A. L., Henderson, C. E., Gazda, R. L., Burke, J. D., Loeber, R., & Pardini, D. A. (2014). Refining the parent-reported inventory of callous-unemotional traits in boys with conduct problems. *Psychological Assessment*, *26*, 256-266.

Henry, J., Pingault, J.-B., Boivin, M., Rijsdijk, F., & Viding, E. (2016). Genetic and environmental aetiology of the dimensions of Callous-Unemotional traits. *Psychological Medicine*, *46*, 405-414.

Herpers, P. C. M., Klip, H., Rommelse, N. N. J., Taylor, M. J., Greven, C. U., & Buitelaar, J. K. (2017). Taxometric analysis and predictive accuracy of callous-unemotional traits regarding quality of life and behavior problems in non-conduct disorder diagnoses. *Psychiatry Research*, *253*, 351-359.

Högström, J., Enebrink, P., & Ghaderi, A. (2013). The moderating role of child callous-unemotional traits in an Internet-based parent-management training program. *Journal of Family Psychology*, *27*, 314-323.

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, *14*, 332-346.

Horan, J. M., Brown, J. L., Jones, S. M., & Aber, J. L. (2015). Assessing invariance across sex and race/ethnicity in measures of youth psychopathic characteristics. *Psychological Assessment*, *27*, 657-668.

Houghton, S., Hunter, S. C., & Crow, J. (2013). Assessing callous unemotional traits in children aged 7- to 12-years: A confirmatory factor analysis of the Inventory of Callous Unemotional Traits. *Journal of Psychopathology and Behavioral Assessment*, *35*, 215-222.

Johnson, M. M., Dismukes, A. R., Vitacco, M. J., Breiman, C., Fleury, D., & Shirtcliff, E. A. (2013). Psychopathy`s influence on the coupling between hypothalamic-pituitary-adrenal and -gonadal axes among incarcerated adolescents. *Developmental Psychobiology*, *56*, 448-458.

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*, 271-282.

Kahn, R. E., Frick, P. J., Youngstrom, E. A., Youngstrom, J., Feeny, N. C., & Findling, R. L. (2013). Distinguishing primary and secondary variants of callous-unemotional traits among adolescents in a clinic-referred sample. *Psychological Assessment*, *25*, 966-978.

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*, 1249-1261.

Kimonis, E. R., Fanti, K. A., & Singh, J. P. (2014). Establishing cut-off scores for the parental-reported Inventory of Callous-Unemotional Traits. *Archives of Forensic Psychology*, *1*, 27-48.

Kimonis, E. R., Frick, P. J., Munoz, L. C., & Aucoin, K. J. (2007). Can a laboratory measure of emotional processing enhance the statistical prediction of aggression and delinquency in detained adolescents with callous-unemotional traits? *Journal of Abnormal Child Psychology*, *35*, 773-785.

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, 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 and Psychiatry*, *31*, 241-252.

Kimonis, E. R., Kennealy, P. J., & Goulter, N. (2016). Does the self-report inventory of callous-unemotional traits predict recidivism? *Psychological Assessment*, *28*, 1616-1624.

Kongerslev, M. T., Bo, S., Forth, A., & Simonsen, E. (2015). Assessment of the affective dimensions of psychopathy with the Danish version of the Inventory of Callous-Unemotional Traits among incarcerated boys: A study of reliability, criterion validity, and construct validity. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology*, *3*, 80-96.

Latzman, R. D., Lilienfeld, S. O., Latzman, N., & Clark, L. A. (2013). Exploring callous and unemotional traits in youth via general personality traits: An eye toward DSM-5. *Personality Disorders: Theory, Research and Treatment*, *4*, 191-202.

Latzman, R. D., Malikina, M. V., Hecht, L. K., Lilienfeld, S. O., & Chan, W. Y. (2016). The contribution of personality and refugee camp experience to callous and unemotional traits among immigrant adolescents in the United States: Implications for the DSM-5 "Limited Prosocial Emotions" specifier. *Child Psychiatry and Human Development*, *47*, 215-225.

Lei, P.-W., & Wu, Q. (2007). Introduction to structural equation modeling: Issues and practical considerations. *Educational Measurement: Issues and Practice*, *26*(3), 33-43.

Leist, T., & Dadds, M. R. (2009). Adolescents' ability to read different emotional faces relates to their history of maltreatment and type of psychopathology. *Clinical Child Psychology and Psychiatry*, *14*, 237-250.

Levy, T., Apter, A., Djalovski, A., Peskin, M., Fennig, S., Gat-Yablonski, G., . . . Bloch, Y. (2017). The reliability, concurrent validity and association with salivary oxytocin of the self-report version of the Inventory of Callous-Unemotional Traits in adolescents with conduct disorder. *Psychiatry Research*, *256*, 124-129.

Lopez-Romero, L., Gomez-Fraguela, A., & Romero, E. (2015). Assessing callous-unemotional traits in a Spanish sample of institutionalized youths: The Inventory of Callous-Unemotional traits. *Journal of Psychopathology and Behavioral Assessment*, *37*, 392-406.

Marini, V., & Stickle, T. R. (2010). Evidence for deficits in reward responsivity in antisocial youth with callous-unemotional traits. *Personality Disorders: Theory, Research, and Treatment*, *1*, 218-229.

Marsee, M. A., Childs, K. K., Frick, P. J., Munoz, L. C., Barry, C. T., Kimonis, E. R., . . . Lau, K. S. L. (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*, 792-804.

McDonald, S. E., Ma, L., Green, K. E., Hitti, S. A., Cody, A. M., Donovan, C., . . . Ascione, F. R. (2018). Evaluation of the parent-report Inventory of Callous-Unemotional Traits in a sample of children recruited from intimate partner violence services: A multidimensional Rasch analysis. *Journal of Clinical Psychology*, *74*, 418-441.

Mittenzwei, K. (2013). *The Screen for Child Anxiety Related Emotional Disorders (SCARED): Eine deutsche Übersetzung und erste empirische Anwendung des Fragebogens für Angststörungen im Kindes - und Jugendalter (SCARED-D)* (Doctoral dissertation). Universität Hamburg, Hamburg, Germany.

Moore, A. A., Carney, D., Moroney, E., Machlin, L., Towbin, K. E., Brotman, M. A., . . . Hettema, J. M. (2017). The Inventory of Callous-Unemotional Traits (ICU) in children: Reliability and heritability. *Behavior Genetics*, *47*, 141-151.

Munoz, 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*, 15-28.

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*, 1134-1144.

Pechorro, P., Ray, J. V., Barroso, R., Maroco, J., & Goncalves, R. A. (2016). Validation of the Inventory of Callous-Unemotional Traits among a Portuguese sample of detained juvenile offenders. *International Journal of Offender Therapy and Comparative Criminology*, *60*, 349-365.

Pederson, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., & Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology*, *9*, 157-166.

Poythress, N. G., Dembo, R., Wareham, J., & Greenbaum, P. E. (2006). Construct validity of the Youth Psychopathic Traits Inventory (YPI) and the Antisocial Process Screening Device (APSD) with justice-involved adolescents. *Criminal Justice and Behavior*, *33*, 26-55.

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

Ray, J. V., Pechorro, P., & Goncalves, R. A. (2016). A comparison of self-report measures of callous-unemotional traits among incarcerated youth. *Criminal Justice and Behavior*, *43*, 1293-1309.

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*, 44-57.

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*, 688-695.

Saunders, M. C. (2016). *The associations between callous-unemotional traits and symptoms of conduct problems, hyperactivity and anxiety: A twin study* (Master thesis). Universite de Montreal, Montreal, Quebec, Canada.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(8), 23-74.

Skeem, J. L., & Cauffman, E. (2003). Views of the downward extension: Comparing the Youth Version of the Psychopathy Checklist with the Youth Psychopathic Traits Inventory. *Behavioral Sciences & the Law*, *21*, 737-770.

SPSS. (2016). SPSS statistics [Computer software]. Chicago, IL: IBM.

Stadlin, C., Perez, T., Schmeck, K., Di Gallo, A., & Schmid, M. (2015). Konstruktvalidität und Faktorenstruktur des deutschsprachigen Youth Psychopathic Traits Inventory (YPI) in einer repräsentativen Schulstichprobe [Construct validity and factor structure of the German-speaking Youth Psychopathic Trait Inventory (YPI) in a representative school sample]. *Diagnostica*, *62*, 85-96.

Statistisches Bundesamt. (2017). *Annual book of statistics: Germany and International*. Wiesbaden, Germany: Author.

Stickle, T. R., Kirkpatrick, N. M., & Brush, L. N. (2009). Callous-unemotional traits and social information processing: Multiple risk-factor models for understanding aggressive behavior in antisocial youth. *Law and Human Behavior*, *33*, 515-529.

Viding, E., Simmonds, E., Petrides, K. V., & Frederickson, N. (2009). The contribution of callous-unemotional traits and conduct problems to bullying in early adolescence. *Journal of Child Psychology and Psychiatry*, *50*, 471-481.

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

Weitkamp, K., Romer, G., Rosenthal, S., Wiegand-Grefe, S., & Daniels, J. (2010). German Screen for Child Anxiety Related Emotional Disorders (SCARED): Reliability, validity, and cross-informant agreement in a clinical sample. *Child and Adolescent Psychiatry and Mental Health*, *4*, 19. doi:10.1186/1753-2000-4-19

White, S. F., Frick, P. J., Lawing, K., & Bauer, D. (2013). Callous-unemotional traits and response to functional family therapy in adolescent offenders. *Behavioral Sciences & the Law*, *31*, 271-285.