

# Further Validation of the Inventory of Callous–Unemotional Traits in Chinese Children: Cross-Informants Invariance and Longitudinal Invariance

Assessment  
1–13  
© The Author(s) 2019  
Article reuse guidelines:  
sagepub.com/journals-permissions  
DOI: 10.1177/1073191119845052  
journals.sagepub.com/home/asm



Meng-Cheng Wang<sup>1</sup> , Yiyun Shou<sup>2</sup>, Jinghui Liang<sup>1</sup>, Hongyu Lai<sup>1</sup>,  
Hong Zeng<sup>1</sup>, Lina Chen<sup>1</sup>, and Yu Gao<sup>3,4</sup>

## Abstract

The present study examined the factor structure and measurement invariance of the shortened versions of the Inventory of Callous–Unemotional Traits (ICU) with data from multiple informants. Five short versions of the ICU proposed in previous studies were tested and compared through confirmatory factor analysis. The measurement invariance across different informants (i.e., self-report, parent-report, and teacher-report) and longitudinal measurement invariance for the resulting best-fitting model were tested thoroughly. Results indicated that a shortened form that consists of 11 items (ICU-11) to assess callousness and uncaring factors had excellent overall fit. Moreover, the ICU-11 was invariant across informant and occasions. However, the ICU-11 was not without limitations; the internal consistency  $\alpha$  for the uncaring factor with self-report scores was marginal. In conclusion, our findings suggest that the ICU-11 was an excellent fit for our data and displayed measurement invariance across informants and over time. The ICU-11 may be a promising assessment tool that could be used in research to assess callous–uncaring traits in Chinese children.

## Keywords

callous–unemotional traits, measurement invariance, longitudinal invariance, multiple informant, Chinese children

The callous and unemotional (CU) traits are characterized by a lack of guilt and empathy, a lack of concern about performance, and shallow and/or deficient affects (Frick, 2009). Numerous studies have demonstrated that CU traits are positively associated with severe, stable, and instrumental patterns of aggressive behavior (Byrd, Loeber, & Pardini, 2012; Kahn, Byrd, & Pardini, 2013; see also Frick, Ray, Thornton, & Kahn, 2014, for a review). For example, Byrd et al. (2012) found that CU traits in childhood (age 7) were associated with persistent criminal behaviors in adulthood (age 25) even when controlling for attention-deficit/hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), and conduct disorder (CD). Given the importance of CU traits in diagnosing CD, CU traits have been added as a specifier to the diagnosis of CD in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*; American Psychiatric Association, 2013).

The Inventory of Callous–Unemotional Traits (ICU) is a comprehensive and focused measure of CU traits (Frick, 2004). Its 24 items are worded to be appropriate for use with youth from preschool-age to older adolescence (e.g., “I do not feel remorseful when I do something wrong.”). There are several formats of the ICU, including the Youth

Self-Report, Parent-Report, Teacher-Report, Parent-Report (Preschool), and Teacher-Report (Preschool).

While extensive research has been done on the self-report version of the ICU (Ciucci & Baroncelli, 2014; Essau, Sasagawa, & Frick, 2006; Pihet, Etter, Schmid, & Kimonis, 2015), the factor structure of the parent- and teacher-reported ICUs has not been investigated extensively (Gao & Zhang, 2016; Hawes et al., 2014; Roose, Bijttebier, Decoene, Claes, & Frick, 2010; Waller et al., 2015; Willoughby, Mills-Koonce, Waschbusch, Gottfredson, & Family Life Project Investigators, 2015). As suggested by the *DSM-5*, to identify the CU traits (i.e., limited prosocial

<sup>1</sup>Guangzhou University, Guangzhou, China

<sup>2</sup>The Australian National University, Canberra, Australian Capital Territory, Australia

<sup>3</sup>Brooklyn College, Brooklyn, NY, USA

<sup>4</sup>The Graduate Center of the City University of New York, New York, NY, USA

## Corresponding Author:

Meng-Cheng Wang, Department of Psychology, Guangzhou University, 230 Wai Huan Xi Road, Guangzhou Higher Education Mega Center, Guangzhou 510006, China.  
Email: wmcheng2006@126.com

emotions) multiple information sources and occasions are required (American Psychiatric Association, 2013). Teacher-report and parent-report formats are commonly used information sources for psychopathological study of children (Kraemer et al., 2003). Furthermore, measurement invariances (MIs) of the different reporting formats of the ICU are requirement to ensure that the ICU is assessing the same construct across different sources. However, there has been no previous attempt to examine the MI (i.e., whether the construct is measured equivalently) of the ICU across different informants (i.e., self-report, parent-report, and teacher-report), nor over time (i.e., longitudinal invariance). The present study aims to address this gap using a large sample of Chinese children using the following ICU formats: self-report, mother-report, and teacher-report data.

### *Factor Structure of ICU and Its Shortened Forms*

With terms of the 24-item self-report version of the ICU, the original planned study proposed a three-factor bifactor model in which all items loaded onto a general factor as well as onto three specific factors: uncaring, callousness, and unemotional (Essau et al., 2006). Despite the poor model fit, according to traditional fitting standards, several independent investigations supported this original bifactor model based on the results that the bifactor model outperformed both the unidimensional model and the intercorrelated three-factor model (Byrd et al., 2012; Fanti, Frick, & Georgiou, 2009; Kimonis et al., 2008; see Wang et al., 2017, for a review). For instance, Fanti et al. (2009) compared three different models in a sample of 347 Greek Cypriot adolescents. The bifactor model fit their data better than the single-factor model and intercorrelated three-factor model, achieving an acceptable fit after including 17 pairs of correlated error variance.<sup>1</sup>

Furthermore, there is only limited evidence supporting the original bifactor model for the parent-report version of the ICU (Roose et al., 2010; cf. Waller et al., 2015). Only one study (Roose et al., 2010) reported acceptable model fit indices for the original three-factor bifactor model. Roose et al. (2010) examined the factor structure of the ICU in a sample of 455 community adolescents (56% male; mean age = 16.67 years old) using three informants' data: parent-report, self-report, and teacher-report). Their findings supported the original three-factor bifactor model when using any of the three informants' versions (self-, parent-, and teacher-report). More recently, Waller et al. (2015) achieved a reasonable model fit with parent-report data for a sample of 450 high-risk 9-year-old children only when a substantial number of modifications had been made to the original model.

The most problematic issue in previous studies has been regarding the unemotional factor. Most studies have reported that the unemotional factor has had much lower internal consistency reliability estimates than the other two

factors (Deng, Wang, Zhang, Shou, & Gao, 2019; Lee-Rowland, Barry, Gillen, & Hansen, 2017). The low reliability of the unemotional factor could suggest that either the scale's items are inconsistent in assessing a latent factor, or that they are assessing more than one single latent trait. Previous studies also found that the unemotional factor had either a low or nonexistent relationship with both the other two ICU factors, as well as the external criteria variables (e.g., Cardinale & Marsh, 2017). Both the reliability and the validity of the unemotional factor measurements, therefore, become questionable.

To address the issues relating to the construct validity of the ICU, several short versions of the ICU have been proposed which select a subset of items from the original ICU. For example, Houghton, Hunter, and Crow (2013) used 16 items of the original scale and tested it using a sample of 268 Australian children. The researchers found that a two-factor model with eight pairs of correlated errors had a marginally acceptable model fit.

Gao and Zhang (2016) proposed another shortened version of self-report ICU using 13 items (ICU-13). The ICU-13 consists of two factors: callousness (seven items) and uncaring (six items). The ICU-13 was validated in a sample of community children for its internal consistency reliability and external validity. The ICU-13 total score and its two-factor scores had acceptable Cronbach's  $\alpha$  values and displayed a reasonable correlation with external measures (Gao & Zhang, 2016). However, the factor structure of the ICU-13 has not been examined. A recent study by Wang et al. (2017) also reported a low model fit for the ICU-13 when used in a sample of Chinese adults.

More recently, Ray, Frick, Thornton, Steinberg, and Cauffman (2016) applied an item response theory approach to develop a unidimensional scale to assess the overall CU traits using self-report data in a sample of delinquent adolescents (Ray et al., 2016). Their shortened version of the ICU consists of 10 items (ICU-10), including 7 items from the original uncaring subscale and 3 items from the original callousness subscale. The ICU-10 was reported to have an acceptable  $\alpha$  coefficient ( $\alpha = .78$ ), as well as test-retest reliability over 6 months ( $r = .59$ ). The ICU-10 was further replicated using a sample of incarcerated Portuguese male juvenile offenders (incremental fit index = .90, comparative fit index [CFI] = .89, root-mean-square error of approximation [RMSEA] = .14; Pechorro, Gonçalves, Hawes, & Ray, 2017). However, the signal-factor model of the ICU-10 did not achieve a good model fit in Chinese adult sample (CFI = .86, Tucker-Lewis index [TLI] = .82, RMSEA = .11; Wang et al., 2017). With regard to the parent- and teacher-report versions of the ICU, Hawes et al. (2014) developed a short version of the test with consideration of item response theory using mother-reported data. This form of the ICU has 12 items (ICU-12) and consists of two factors: callousness (seven items) and uncaring (five items; Hawes et al., 2014).

The factor structure of the ICU-12 was replicated in several studies using self-report data (Colins, Andershed, Hawes, Bijttebier, & Pardini, 2016; Paiva-Salisbury, Gill, & Stickle, 2017) as well as parent-report data (Kimonis, Fanti, Mertan, Goulter, & Katsimicha, 2016; Waller et al., 2015). For example, using both parent- and teacher-report data of a mixed sample (i.e., high-risk and community children) of 214 children, Kimonis et al. (2016) compared several models using both the full and shortened versions of the ICU. They found that the ICU-12 fit the data best. Colins et al. (2016) used the parent-report data of detained female adolescents and demonstrated that the ICU-12 without Item 6 also achieved satisfactory model fit. This finding (ICU-12 without Item 6) was also replicated using Chinese adults and detained Chinese adolescents (Wang et al., 2017; Zhang, Wang, Shou, Luo, & Deng, 2019).

### *Cross-Informant and Longitudinal Measurement Invariance of ICU*

Meta-analyses by De Los Reyes et al. (2015) found that psychopathological assessments usually had low-to-moderate levels of agreement across informants (De Los Reyes et al., 2015; Kraemer et al., 2003). This is known as informant discrepancies (De Los Reyes & Kazdin, 2005). This discrepancy has also been observed when measuring CU traits. For example, Frick et al. (2014) summarized 13 studies that reported 24 correlations between measures of CU traits using different informants or methods; the average correlation was  $r = .24$ , and ranged from  $-.09$  to  $.54$ . However, the cause of these low agreement levels is not yet well understood.

It is unclear whether these discrepancies reflect true differences on the same construct (e.g., the parent has rated CU traits to a different degree than the child has in self-reporting), or if these differences are due to the instrument measuring different constructs across informants. Previous multi-informant assessments appear to have neglected the importance of MI, that is, the extent to which items have equal meaning across groups and/or time, across different informants. Ignoring MI would hamper the validity of tests regarding score comparison, cross-informant agreement, and incremental validity (De Los Reyes et al., 2015).

Another type of MI is longitudinal measurement invariance (LMI; Liu et al., 2017; Widaman, Ferrer, & Conger, 2010). LMI refers to how the relationships between the latent variable and their manifest indicators are invariant over time. If the LMI is violated, the differences observed in the scores over time do not necessarily reflect changes in the latent variable. This may lead to limited or even misleading conclusions, however, and can hinder the validity of score comparison, especially in interventional studies (Wu, 2017). To date, there are no studies that have examined and compared the factor structures of the

shortened versions of the ICU across different informant formats. Additionally, there has been no investigation as of yet that tests the LMI of the ICU. This current study was designed to fill those gaps.

### *The Present Study*

The first aim of this study was to examine and compare the factor models of the ICU, including the original ICU-24 (bifactor model) and the five short versions—ICU-10, ICU-11, ICU-12, ICU-13, and ICU-16—across different informants (i.e., self-, mother-, and teacher-reporting).

Our second aim was to examine the MI of the resulting best-fit model across three informants using multiple-group confirmatory factor analyses (CFAs). The correlations of the latent factors' scores across the informants would also be tested for agreement among the same informants. We also examined the LMI of the best-fitting model in a sample of children (self-report method) after a 6-month period. Based on previous research (Colins et al., 2016; Kimonis et al., 2016; Paiva-Salisbury et al., 2017; Waller et al., 2015; Wang et al., 2017), we predicted that the ICU-12 or/and the ICU-11 would fit the data best.

The third aim was to examine the criterion validity of the different versions of the ICU measuring the aforementioned three informants using the Child Problematic Traits Inventory (CPTI). The ICU measured at Time Point 1 will be used to correlate with the CPTI scores at Time Point 2.

## **Method**

### *Participants*

Data were collected from a primary school in Guangdong, China. The self-report ICU data were gathered from a sample total of 512 children (Grades 4-6; girls = 47.6%). A total of 362 of the 512 children were tested a second time 6 months after Test Point 1. The responses of these 362 children were used to test the longitudinal invariance of the self-report ICU.

For the parent- and teacher-report data, a total of 977 children in Grades 1 to 6 (6- to 12-year-olds) were included in the data collection. Mothers of 967 children (99%) responded to the invitation to participate in the study. The mean age of the mothers was 37.85 years old ( $SD = 3.79$ , range: 28-50 years). The teacher-report data involved 24 teachers connected to one of the 977 students. The mean time of acquaintance between the teacher and the student they were measuring was 2.27 years ( $SD = 1.51$ , range: 1-6 years). The children in the self-report data (Grades 4 to 6) were a subset of the teacher-rated child sample. Information regarding the child's grade, number of siblings, family structure, and parents' education and monthly income are presented in Table 1.

**Table 1.** Demographic Information for the Current Sample ( $N = 967$ ).

Variables	Parent-report ( $N = 967$ )	Self-report ( $n = 512$ )	Self-report T2 ( $n = 362$ )
Sex, $n$ (%)			
Boys	478 (49.43)	267 (52.15)	181 (50)
Girls	448 (46.33)	245 (47.85)	181 (50)
Missing information	41 (4.24)	0	0
Grades, $n$ (%)			
First grade	167 (17.3)		
Second grade	178 (18.4)		
Third grade	143 (14.8)		
Fourth grade	188 (19.4)	169 (33.01)	38 (10.5)
Fifth grade	127 (13.1)	175 (34.18)	163 (45)
Sixth grade	164 (17)	168 (32.81)	161 (44.5)
Number of siblings, $n$ (%)			
None	693 (71.7)	382 (74.61)	260 (71.8)
One or more	221 (22.9)	130 (25.39)	102 (28.2)
Missing information	53 (5.4)	0	0
Family composition, $n$ (%)			
Father and mother	893 (92.3)	486 (94.92)	350 (96.7)
Single parent	27 (2.8)	4 (0.78)	7 (1.9)
Missing information	47 (4.9)	22 (4.3)	5 (1.4)
Educational level, $n$ (%)			
Elementary school	20 (2.1)	10 (1.95)	12 (3.3)
Junior middle school	92 (9.5)	50 (9.77)	36 (9.9)
Senior middle school	129 (13.3)	76 (14.84)	51 (14.1)
Technical secondary	112 (11.6)	58 (11.33)	40 (11.0)
Junior college	322 (33.3)	183 (35.74)	131 (36.2)
Bachelor degree or above	257 (26.6)	125 (24.41)	92 (25.4)
Missing information	35 (3.6)	10 (1.95)	0
Monthly income, <sup>a</sup> $n$ (%)			
<4,000	171 (17.7)	90 (17.58)	63 (17.4)
4,000-8,000	373 (38.6)	187 (36.52)	132 (36.5)
8,000-12,000	202 (20.9)	115 (22.46)	84 (23.2)
12,000-16,000	78 (8.1)	48 (9.34)	34 (9.4)
16,000-20,000	48 (5)	26 (5.08)	19 (5.2)
>20,000	53 (5.5)	39 (7.62)	25 (6.9)
Missing information	42 (4.3)	7 (1.37)	5 (1.4)

<sup>a</sup>The monetary unit of income is Yuan; US\$1 equals about 6.9 Yuan.

## Procedure

The head of the children's school and their teachers were informed about the purposes of the study. Information and parental consent forms were sealed in envelopes with the questionnaire and were taken home by the school children. The parents were instructed to return the questionnaires in sealed envelopes to their child's teacher within 2 days. Children in Grades 4 to 6 completed a self-report version of the ICU as well as other questionnaires (not used in this present study) during a 45-minute class session on a regular school day. Teachers completed the teacher-report version of the ICU and were instructed to return the questionnaires along with the questionnaires completed by the students' parents within 2 weeks to the third study author. Teachers,

parents, and children who completed the questionnaires were each paid by gift vouchers worth approximately 15 Yuan or US\$2 for their participation. The study was reviewed and approved by the Human Subjects Review Committee at Guangzhou University (Review No. 20141008).

## Measures

*Inventory of Callous-Unemotional Traits* (Frick, 2004). Parents, teachers, and children completed the 24-item parent-, teacher-, and self-report versions of the ICU (Frick, 2004) separately. Items were scored on a 4-point scale (*not at all true* = 0, *somewhat true* = 1, *very true* = 2, and *definitely true* = 3). Selected items were analyzed, depending on the testing model (see Table 4 for model specifications).

*The Child Problematic Traits Inventory.* The CPTI (Colins et al., 2014; Wang et al., 2018) is a parent-report measure of child problematic traits. It has 28 items and each item is rated on a 4-point Likert-type scale (*does not apply at all* = 1, *does not apply well* = 2, *applies fairly well* = 3, and *applies very well* = 4). These 28 items assess three dimensions: interpersonal (Grandiose–Deceitful [GD], 8 items), affective (Callous–Unemotional [CU], 10 items), and behavioral (Impulsive–Need for Stimulation [INS], 10 items). In addition to the three factors' scores, the CPTI yields a total score to measure a general level of psychopathy. A higher CPTI score indicates a higher psychopathic level. The data of the CPTI in the present study was taken from the parent-report CPTI at Time Point 2. The alpha coefficients for the GD, CU, and INS subscale and total scores were .84, .89, .86, and .94, respectively.

### Data Analysis

**Step 1: CFA.** Several short forms of the ICU were tested and compared using CFA via *Mplus* 7.4 (Muthén & Muthén, 1998-2010).<sup>2</sup> Maximum likelihood estimation was inappropriate when response categories were less than five (DiStefano, 2002). Therefore, the robust weighted least-squares with mean and variance adjustment (WLSMV) estimator was used (Flora & Curran, 2004). Multiple fit indices were evaluated: chi-squares, RMSEA, TLI, and CFI. RMSEA values  $\leq .08$  indicate an acceptable model fit, and  $\leq .05$  indicate a good model fit, and a CFI or TLI  $\geq .90$  indicates an adequate model fit (Kline, 2010). Given that cutoff standards may be specific to particular models and data sets (West, Taylor, & Wu, 2012), multiple indices were applied to compare models. If all the fit indices used did not give the same conclusion on the best model, a decision was made by the investigative team based on majority rule.

**Step 2: Measurement Invariance.** MI tests were performed using the sequential strategy described by Meredith and Teresi (2006). Five levels of MI were tested through multiple-group CFA. The first level was configural invariance, in which no parameters were constrained. Good model fit indices indicated satisfied configural invariance, which meant that the item-factor structure was similar across different samples. The second level was metric invariance or weak invariance, in which all item factor loadings were constrained to be equal. The third level is the scalar invariance or strong invariance, in which the item intercepts or thresholds were equally constrained. The fourth level was the unique variance invariance or strict invariance, in which the item unique variance was constrained equally. The last level was the latent mean invariance, in which the latent mean of a factor was constrained to be equal. The latent mean invariance testing does not require the unique variance invariance (Marsh et al., 2009), therefore, the last level models were compared with the scalar invariance model.

A nonsignificant chi-square difference test indicates that the more constrained model fit the data as well as the less constrained model. However, the chi-square difference test is highly sensitive to sample size, and thus the change in CFI ( $\Delta$ CFI) was reported. According to Cheung and Rensvold (2002), a  $\Delta$ CFIs  $< .01$  indicates that the invariance hypothesis should not be rejected, as mean differences exist when  $\Delta$ CFIs are from  $.01$  to  $.02$ , and definite differences exist when  $\Delta$ CFIs are  $> .02$  (Cheung & Rensvold, 2002). Nevertheless, the DIFFTEST function of *Mplus* that was designed to compare chi-square difference with WLSMV estimator was employed.

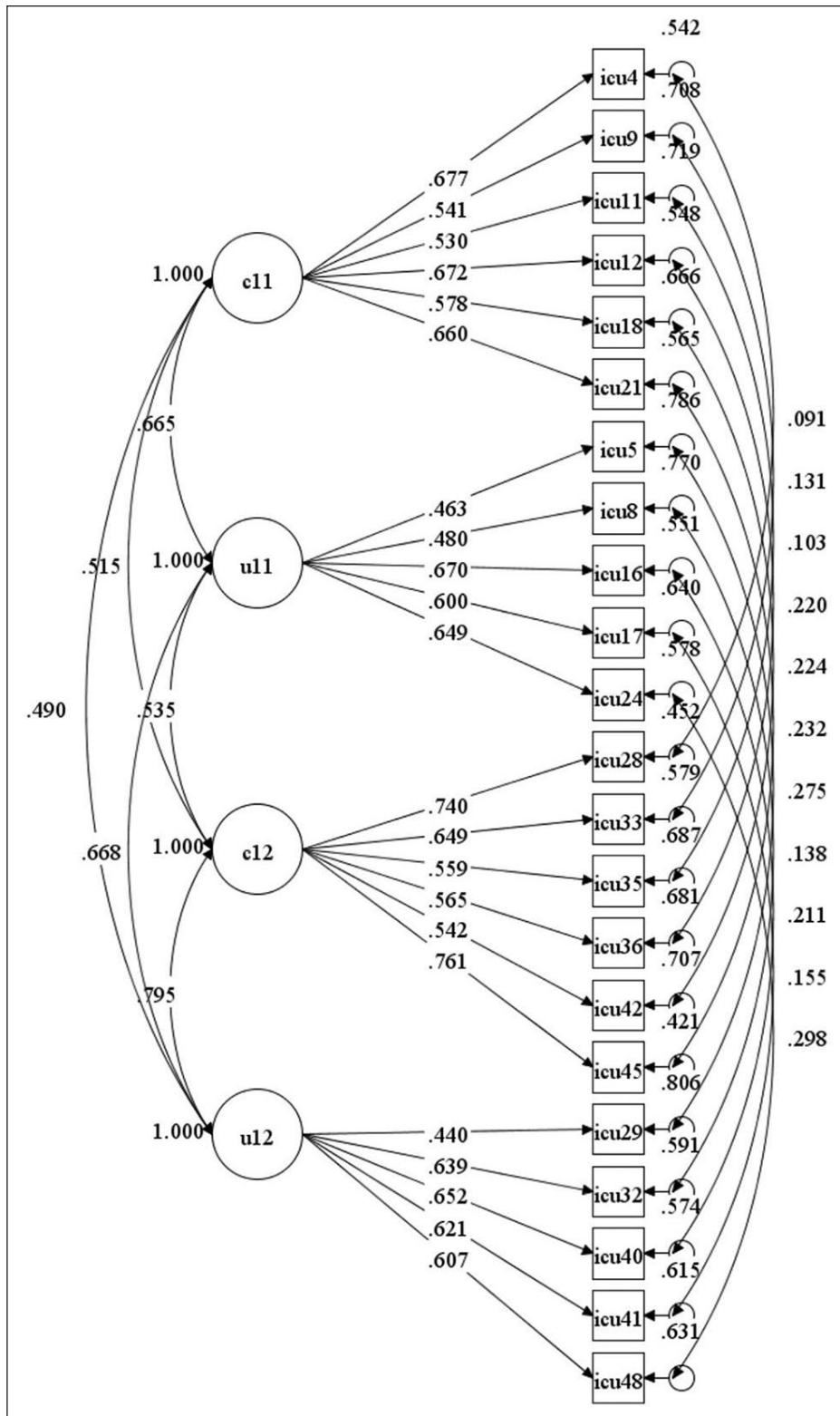
**Step 3: Longitudinal Measurement Invariance.** The LMI testing procedure with ordered-categorical indicator proposed by Liu et al. (2017) was adopted in the current study. The specification of the longitudinal invariance differs from MI through a multiple-group CFA. Testing LMI of the factor structure requires analyzing data by fitting the two or more waves of data with two separate models simultaneously (configural invariance model specified, see Figure 1). Therefore, two separate factor models were constructed for Time Point 1 and Time Point 2 data, with factor correlations estimated between Time Point 1 and Time Point 2 factors. In addition, items' residual error variance for Time Point 1 was allowed to covary with that at Time Point 2. In other words, correlations were estimated among all possible pairs of error variance between Time Point 1 and Time Point 2 as the same items were used across the two different time points (Pitts, West, & Tein, 1996). Additionally, the latent mean of the factors over the two time points were tested.

**Step 4: Reliability Analysis.** Cronbach's alpha coefficients were interpreted as follows:  $\alpha < .60$  = insufficient;  $.60$  to  $.69$  = marginal;  $.70$  to  $.79$  = acceptable;  $.80$  to  $.89$  = good, and  $.90$  or higher = excellent (Barker, Pistran, & Elliot, 1994). Given that  $\alpha$  depends on the number of items, we also reported mean interitem correlations (MICs), which are independent of scale lengths and should be in the range of  $.15$  to  $.50$  to be considered acceptable (Clark & Watson, 1995).

## Results

### Confirmatory Factor Analysis

The fit indices of various short forms of the ICU are displayed in Table 2. As expected, the fit of the original ICU bifactor model for the data were unacceptable in terms of all fit indices across all three rated formats. With regard to the shorter ICU versions, all models showed an acceptable fit (CFIs  $> .90$  and TLIs  $> .90$ ) for all three informant-report versions (self-, parent-, and teacher-report). Generally, within each informant version, the ICU-11 and ICU-12 fit the data better than the other models. The factor loadings



**Figure 1.** The diagram for longitudinal configural invariance.  
 Note. ICU = Inventory of Callous–Unemotional Traits.

**Table 2.** Goodness-of-Fit Indices for Tested Models in the Confirmatory Factor Analysis.

Model	$\chi^2$	df	CFI	TLI	RMSEA [90% CI]
<b>Teacher-report</b>					
ICU-24	2595.94*	228	.88	.85	.10 [.10, .11]
ICU-10	479.77*	35	.95	.93	.11 [.11, .12]
<b>ICU-11</b>	<b>476.55*</b>	<b>43</b>	<b>.95</b>	<b>.93</b>	<b>.10 [.09, .11]</b>
ICU-12	541.91*	53	.95	.93	.10 [.09, .10]
ICU-13	792.33*	64	.94	.93	.11 [.10, .11]
ICU-16	1148.05*	103	.93	.92	.10 [.10, .11]
<b>Self-report</b>					
ICU-24	746.33*	228	.88	.86	.07 [.06, .07]
ICU-10	394.93*	35	.95	.93	.14 [.13, .15]
<b>ICU-11</b>	<b>148.81*</b>	<b>43</b>	<b>.98</b>	<b>.97</b>	<b>.07 [.06, .08]</b>
ICU-12	223.37*	53	.97	.96	.08 [.07, .09]
ICU-13	301.09*	64	.96	.96	.08 [.07, .09]
ICU-16	326.04*	103	.93	.92	.06 [.06, .07]
<b>Parent-report</b>					
ICU-24	1710.97*	228	.91	.89	.08 [.08, .09]
ICU-10	478.39*	35	.93	.91	.12 [.11, .13]
ICU-11	260.12*	43	.97	.96	.07 [.07, .08]
<b>ICU-12</b>	<b>269.40*</b>	<b>53</b>	<b>.97</b>	<b>.96</b>	<b>.07 [.06, .07]</b>
ICU-13	615.87*	64	.95	.93	.10 [.09, .10]
ICU-16	864.03*	103	.94	.93	.09 [.08, .09]

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval. Best-fitting models are in bold. \* $p < .001$ .

for both the ICU-11 and ICU-12 are exhibited in Table 3.<sup>3</sup> The factor loadings for ICU-11 and ICU-12 fall in the range of .36 and .82.

Across all informants, the ICU-11 and ICU-12 fit the data similarly well among all models. The ICU-11 fit the data best for in the self- and teacher-report versions, while the ICU-12 fit the parent-report data slightly better than the ICU-11 did. Close inspection of the factor loading revealed that although the ICU-12 had the best model fit for the parent-report data, Item 6 had a low factor loading (.36). The low factor loading of Item 6 was also reported in several other studies that used Chinese-speaking samples. It has been argued that Item 6 (“I do not show my emotions to others”) may not be suitable for assessing CU traits in Chinese-speaking subjects, as it can be affected by the Chinese cultural norm of suppressing emotions to others (see Discussion for more details). Considering the model fit and factor loading, we decided to select the ICU-11 as the best model for the subsequent analyses.

**Internal Consistency of ICU Scores.** As presented in Table 4, the  $\alpha$  for the scores of the two factors of all tested models of the parent- and teacher-report ICU formats ranged from acceptable to good, with one exception for the uncaring

factor in the ICU-11 and ICU-12, due in part to the short scale length (five items). The  $\alpha$  for the self-report ICU scores were relatively lower than those of the parent- and teacher-report ICU scales. The MICs that account for the scale length were acceptable for all factors in all informant versions.

Finally, the cross-informant correlations among three informants are presented in Table 5. The correlations between the self- and parent-report ICUs were .38 and .35, and the correlations between the self- and teacher-report ICUs were .21 and .20. Meanwhile, the correlations between the parent- and teacher-report ICUs were .11 and .15 for the callousness and uncaring subscale scores, respectively ( $p < .01$  for each).

### Measurement Invariance Across Informants

Given that the ICU-11 fit the data best across all informant formats (see the above section), we examined the baseline model on the basis of the ICU-11 (i.e., configural invariance model) across all three different informant report groups. The results indicated that the ICU-11 fit all the three groups well (see Table 6). The metric invariance model was then tested by constraining the factor loadings equally across the three groups. In terms of the  $\Delta$ CFI and  $\Delta$ TLI (both less than .01), the results indicated that there were no significant differences in item loading across groups. Likewise, the result for the scalar invariance test also showed that intercepts were invariant across all three groups. By contrast, the unique variance invariance or strict invariance were not supported in terms of  $\Delta$ CFI (.028) and  $\Delta$ TLI (.018).

Finally, the latent means for two factors across the three informant formats were constrained to be equal, and results showed a significant decrease in model fit for the last model, suggesting that the latent factor means were significantly different across the three groups. With the parent-report data set as the reference group, the self-report group reported latent mean scores of  $-0.75$  ( $p < .001$ ) and  $-0.52$  ( $p < .001$ ) for callousness and uncaring factors, respectively, while the teacher-report latent mean scores were  $0.54$  ( $p < .001$ ) and  $0.51$  ( $p < .001$ ) for these two factors, respectively. These indicate that the callousness and uncaring factors had lower latent means for the self- and parent-report versions of the ICU than it did for the teacher-report versions, while teacher-report versions had highest latent means for the callousness and uncaring factors than it did for the self- and parent-report version. Complete model fit indices are presented in Table 6.

### Longitudinal Measurement Invariance

The LMI test was conducted by self-report after a 6-month interval. The configural invariance model provided a great fit to the data in terms of all fit indices (CFI = .96,

**Table 3.** Standardized Factor Loadings of the ICU-11 and ICU-12 for Three Versions.

Item	Parent-report	Self-report	Teacher-report
<i>Callousness</i>			
4. I do not care who I hurt to get what I want.	.65/.66	.79/.79	.64/.64
6. I do not show my emotions to others.	—/.36	—/.58	—/.48
9. I do not care if I get into trouble.	.67/.67	.73/.72	.73/.73
11. I do not care about doing things well.	.57/.57	.69/.69	.70/.70
12. I seem very cold and uncaring to others.	.73/.73	.78/.81	.73/.74
18. I do not feel remorseful when I do something wrong.	.77/.77	.73/.73	.62/.61
21. The feelings of others are unimportant to me.	.79/.79	.78/.77	.77/.77
<i>Uncaring</i>			
5. I feel bad or guilty when I do something wrong. (R)	.66/.66	.69/.69	.60/.60
8. I am concerned about the feelings of others. (R)	.59/.59	.65/.65	.67/.67
16. I apologize ('say I am sorry') to persons I hurt. (R)	.67/.66	.82/.82	.76/.75
17. I try not to hurt others' feelings. (R)	.70/.70	.80/.80	.74/.74
24. I do things to make others feel good. (R)	.64/.64	.80/.80	.59/.59

Note. ICU = Inventory of Callous–Unemotional Traits. (R) indicates reverse code. All factor loadings were significant at the .001 level. The values in front of the slash were factor loading for the ICU-11; values behind the slash were factor loading for ICU-12.

**Table 4.** Model Specification and Internal Consistency for Five Tested Models.

Model number	Model specification and items	Cronbach's $\alpha$ (self/ parent/teacher)	MIC (self/ parent/teacher)	Number of items
Ray et al. (2016) ICU-10	ICU: 3, 5, 7, 8, 11, 15, 16, 17, 23, 24	.72/.79/.86	.21/.27/.39	10
Colins et al. (2016) ICU-11	Callousness: 4, 9, 11, 12, 18, 21	.69/.79/.81	.27/.38/.41	6
	Uncaring: 5, 8, 16, 17, 24	.63/.69/.75	.25/.31/.37	5
Hawes et al. (2014) ICU-12	Callousness: 4, 6, 9, 11, 12, 18, 21	.68/.77/.81	.24/.33/.38	7
	Uncaring: 5, 8, 16, 17, 24	.63/.69/.75	.25/.31/.37	5
Gao and Zhang (2016) ICU-13	Callousness: 7, 9, 11, 12, 18, 20, 21	.69/.80/.83	.24/.37/.42	7
	Uncaring: 3, 13, 15, 16, 17, 24	.69/.72/.82	.27/.30/.43	6
Houghton et al. (2013) ICU-16	Callousness: 4, 7, 8, 9, 11, 12, 18, 21	.70/.80/.84	.23/.33/.39	8
	Uncaring: 3, 5, 13, 15, 16, 17, 23, 24	.75/.79/.86	.27/.32/.43	8

Note. CU = callousness–unemotional; ICU = Inventory of Callous–Unemotional Traits; MIC = mean interitem correlations.

**Table 5.** Correlation Among Factors of ICU-11 Across Three Informants.

	SC	SU	PC	PU	TC
Self-report uncaring (SU)	.72**				
Parent-report callous (PC)	.38**	.30**			
Parent-report uncaring (PU)	.35**	.35**	.90**		
Teacher-report callous (TC)	.21**	.26**	.11**	.09*	
Teacher-report uncaring (TU)	.19**	.20**	.17**	.15*	.83*

\* $p < .05$ . \*\* $p < .01$ .

TLI = .95, RMSEA = .04). The correlations within and across factors for the longitudinal invariance (configural) model were present in Figure 1. The magnitudes of correlation were mostly moderate.

Next, a more constrained model was tested sequentially. The results also indicated that no significant differences in

factor loadings and item thresholds were found across the two time points ( $\Delta$ CFIs and  $\Delta$ TLLs  $< .01$ ). Overall, the results suggested that the ICU-11 displayed LMI across the two different occasions. In addition, when the latent mean differences were tested with two waves, the results indicated that the latent means for the two factors in the second wave were lower than that of the first wave ( $-.18$  for callousness,  $p = .031$ ;  $-.25$  for uncaring,  $p < .001$ ). The complete model fit indices are presented in Table 6.

### Criterion validity

Table 7 shows the correlations between the different versions of the ICU measured at Time Point 1 and the parent-reported CPTI at Time Point 2. Significant correlations were observed between the ICU scores and both the factors and total scores of the CPTI. In general, the parent-report ICU had the strongest correlations with the parent-report



**Table 6.** Results of Measurement Invariance.

	$\chi^2$	df	CFI	TLI	RMSEA [90% CI]	$\Delta\chi^2$	$\Delta df$	p	$\Delta CFI$	$\Delta TLI$
Across informant										
Configural invariance	802.07**	129	.96	.95	.08 [.08, .09]					
Metric invariance	935.50**	147	.96	.95	.08 [.08, .09]	150.35	18	<.001	.007	.001
Scalar invariance	1017.28**	191	.95	.96	.07 [.07, .08]	245.31	44	<.001	.002	-.009
Strict invariance	1553.24**	213	.93	.94	.09 [.08, .09]	550.19	22	<.001	.028	.018
Latent mean invariance	3181.16**	195	.84	.86	.14 [.13, .14]	812.11	4	<.001	.091	.082
Longitudinal invariance										
Configural invariance	284.10**	192	.96	.95	.04 [.03, .05]					
Metric invariance	280.34**	201	.97	.96	.03 [.02, .04]	3.51	9	.941	-.005	-.008
Scalar invariance	297.16**	221	.97	.97	.03 [.02, .04]	20.62	20	.420	-.002	-.005
Strict invariance	334.26**	232	.96	.96	.04 [.03, .04]	37.08	11	<.001	.011	.009
Latent mean invariance	316.27**	223	.96	.96	.03 [.03, .04]	14.82	2	<.001	.008	.007

Note. df = degrees of freedom; CFI = comparative fit index; TLI = Tucker–Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval.

\*\*p < .001.

**Table 7.** Zero-Order Correlations Between External Criteria and the Various Versions of ICU Scores.

	GD	CU	INS	TOT
Self-report <sup>a</sup>				
ICU-24	.22**	.30**	.26**	.30**
ICU-10	.20**	.24**	.21**	.25**
ICU-11	.22**	.29**	.27**	.30**
ICU-12	.22**	.29**	.27**	.29**
ICU-13	.20**	.29**	.22**	.27**
ICU-16	.22**	.29**	.24**	.29**
Parent-report <sup>b</sup>				
ICU-24	.42**	.54**	.50**	.57**
ICU-10	.38**	.49**	.47**	.52**
ICU-11	.42**	.55**	.50**	.57**
ICU-12	.42**	.55**	.50**	.57**
ICU-13	.43**	.54**	.52**	.57**
ICU-16	.44**	.55**	.52**	.58**
Teacher-report <sup>c</sup>				
ICU-24	.21**	.21**	.16**	.22**
ICU-10	.22**	.20**	.15**	.22**
ICU-11	.21**	.19**	.15**	.21**
ICU-12	.20**	.19**	.14**	.20**
ICU-13	.22**	.21**	.15**	.22**
ICU-16	.22**	.21**	.16**	.22**

Note. GD = grandiose-deceitful; CU = callous–unemotional; ICU = Inventory of Callous–Unemotional Traits; INS = impulsive–need for stimulation; CTPI = Child Problematic Traits Inventory; TOT = total scores.

<sup>a</sup>The correlations were calculated basing on all available self-report ICU and mother-report CPTI data, the sample size at the ranged from 302 to 487 due to missing data. <sup>b</sup>The correlations were calculated basing on all available mother-report ICU and mother-report CPTI data, the sample size at the ranged from 531 to 878 due to missing data. <sup>c</sup>The correlations were calculated basing on all available teacher-report ICU and mother-report CPTI data, the sample size at the ranged from 545 to 871 due to missing data.

\*\*p < .001.

CPTI, followed by the self-report and the teacher-report versions. The self- and parent-report ICUs also had clear discriminability validity in terms of having stronger correlations with the CPTI CU factor than with the other two CPTI factors. The ICU-11 had similarly strong correlations in comparison to the original version of the ICU, or with other longer versions.

## Discussion

The primary purpose of this study was to examine five shortened forms of the ICU as proposed in previous studies, and to test the MI of the resulting best-fitting model across informant-type and time point in a large sample of Chinese children. Our results indicated that the ICU-11 fit the data best. The ICU-11 also achieved MI across informant format type and time points.

### Confirmatory Factor Analysis

Five different short versions, as identified in previous studies, and three different informant versions of those models were tested and compared in the current study. Both the ICU-11 and ICU-12 contained similar items (the only item differing between the two short versions was Item 6) as well as having similar factor structures, and thus had similar model fit results. Both versions outperformed the other three short versions of the ICU, indicating that both the items and factors structures may best suit the current Chinese sample.

The ICU-11 achieved the best fit for the self- and teacher-report ICU versions. For the parent-report version, although the ICU-12 had slightly higher TLI than the ICU-11, Item 6 (“I do not show my emotions to others”) had a much lower factor loading than the other items in the questionnaire.

Similar results were also reported in several previous studies, using numerous sample types. These have included samples of Chinese adults and detained adolescent populations (Wang et al., 2017; Zhang et al., 2019), two Portuguese samples—one mixed (i.e., detained female juvenile offenders and community youths; Pechorro, Hawes, Gonçalves, & Ray, 2017) and another of incarcerated male juvenile offenders (Pechorro, Gonçalves, et al., 2017), and one of detained Belgian female adolescents (Colins et al., 2016). One possible explanation is that not showing emotion to others could be due to a lack of emotion, or it could be due to simply being unwilling to show emotions. This item may have been interpreted by parents as shy behavior rather than “lacking emotion” in the CU sense of the statement (Hawes et al., 2014).

The current study also compared the different candidate models in terms of their correlations with the CPTI. The ICU-11 had similarly strong correlations to the CPTI when compared with the original version of the ICU. This suggests that the ICU-11 can achieve better efficiency (i.e., is shorter in length) without losing much useful information.

### *Measurement Invariance and Informant Agreement*

The ICU-11 was found to be invariant across all three informant types, indicating that the ICU-11 can assess all intended constructs, regardless of the source of the information. This provides evidence that a multi-informant approach can be applied to integrate information when assessing CU traits.

As for the latent factor mean differences among the three report versions of the ICU, the two ICU factors were lowest when using the self-report ICU version, and were highest when using the teacher-report version. These findings were partially consistent with previous studies which showed that teacher-report scores on the CU factor of the APDS were higher than parent- or youth-report scores (Docherty, Boxer, Huesmann, O'Brien, & Bushman, 2017; Frick, Kimonis, Dandreaux, & Farell, 2003). Docherty et al. (2017) provided some possible explanations for this. On one hand, children might underreport on the ICU due to a social desirability factor, while teachers are least biased by the need to avoid shame. On the other hand, assessing emotions and feelings requires understanding and insight into the self; teachers may be more knowledgeable about a child's emotions than the parent or the child itself. Docherty et al. (2017) suggested that the parent-report ICU may be optimal in terms of balancing knowledge and social desirability.

The current study also tested the MI of a shortened version of the ICU over time. Results indicate that the factor model of the ICU-11 is invariant across all measurement occasions during an interval of 6 months. The modest correlation between CU traits at two measured occasion over 6

months indicated that the CU trait is stable at least in the short-term, supporting previous findings (e.g., Frick et al., 2003; Lynam et al., 2009; Lynam, Caspi, Moffitt, Loeber, & Stouthamer-Loeber, 2007). Additionally, the latent means for the two factors in the second wave were lower than in the first wave, suggesting that the level of CU traits measured by the ICU decreased as the subject's age increases. This may reflect the development of a child's ability to identify and/or express emotion, both due to their increase in age as well as increased exposure to others in the social (school) environment. However, it was difficult to determine the growth tendency given that there were only two time points used, and there could be significant within-person heterogeneity across childhood and adolescence (Byrd, Hawes, Loeber, & Pardini, 2018; Hawes et al., 2017). A longer time interval is required to better determine the change of CU traits over time.

With regard to internal consistency, current study findings were consistent with those of most previous studies (Colins et al., 2016; Pechorro, Hawes, et al., 2017; Wang et al., 2017). As previously mentioned, the low values of Cronbach's  $\alpha$  values are partially due to the small number of items in the scale. The  $\alpha$  values are lower for the child self-report version than for the two adult self-report versions. This could be due to children younger than 12 years old having limited reading comprehension (Soto, John, Gosling, & Potter, 2008), which may influence their ability to provide consistent responses on the ICU items and lead to their tendency to follow items' framing. Future studies should investigate how items' valence could influence the scoring of child-report ICU.

Furthermore, the cross-informant correlations for the three ICU subscales ranged from .113 to .384, all of which were lower than correlations found in past studies of CU traits among Western samples ( $r = .29$  to  $.57$ ; Frick et al., 2003; Roose et al., 2010). These values, however, were similar to Gao and Zhang's (2016) findings ( $r = .15$  to  $.23$ ). The low correlations may be due to cultural factors, specifically that Chinese culture in general does not encourage free public emotional expression. This high expectation of emotional regulation in public could bias others' perceptions of emotional expression.

It was also noticed that the strength of the correlations between the self- and parent-report formats were higher than those between self- and teacher-report versions. This may suggest that mothers may have had a better understanding of their children than the teachers did. Another possible cause of this could be due to the fact that each teacher was responsible for at least 40 students in the Chinese primary schools the samples were from, thus limiting the attention they can give to each child.

On the positive side, as suggested by several meta-analyses of multi-informant reports of child psychopathology (Achenbach, McConaughy, & Howell, 1987; De Los Reyes

et al., 2015; De Los Reyes & Kazdin, 2005), the low correlation between the parent- and teacher-report formats might suggest that the information provided by both the parents and the teachers covered different aspects of a child's behavior, particularly in different settings (i.e., home vs. school). Further investigation is needed to test the unique predictive utility of the CU traits assessed by different informants in terms of predicting different outcomes. For instance, a teacher-report CU may better predict school-based outcome variables, such as bullying (Ciucci & Baroncelli, 2014), than a parent- or self-report on CU traits.

### Limitations and Future Direction

Two main limitations of the current study should be acknowledged. First, as the sample used in this study was recruited from a single community primary school, the results—especially the teacher-report data—were subject to a clustering effect (e.g., class effect, with each teacher providing more than 40 children's data). Future studies should include more diverse samples, particularly looking at clinical settings and/or juvenile offenders (Colins et al., 2016; Pechorro, Hawes, et al., 2017). Second, we examined the LMI of the ICU-11 in our sample with only a 6-month interval. Longer time intervals and more repeated intervals are needed to better test the stability of the factor structure of the ICU over time (Obradović, Pardini, Long, & Loeber, 2007).

In conclusion, our findings suggested that that a shortened form of the ICU with 11 items (the ICU-11) may be a promising assessment tool that can be used in longitudinal research to assess callous–uncaring traits in children. The MI and LMI of the ICU-11 were also verified with multiple informants, including parent-, teacher-, and self-report formats. We encourage the replication of the present findings to wider range of samples (i.e., different age groups, and clinical/forensic sample).

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by grants from the National Natural Science Foundation of China (Grant 31400904) and Guangzhou University's 2017 training program for top-notch young people (Grant BJ201715).

### Notes

1. Simulation studies demonstrated that models using correlated errors (i.e., correlated unique, CU) could yield sustainably biased estimates of factor loadings and factor correlations (Conway, Lievens, Scullen, & Lance, 2004; Lance, Noble, &

Scullen, 2002).

2. We also examined alternative models such as bifactor model of the ICU-24 and bifactor model of the ICU-11. The results and detailed factor loading are available in the online supplementary file.
3. The factor loadings for the other three models are presented in supplemental tables, which is available online.

### Supplemental Material

Supplemental material for this article is available online.

### ORCID iD

Meng-Cheng Wang  <https://orcid.org/0000-0002-0855-2219>

### References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin, 101*, 213-232. doi:10.1037/0033-2909.101.2.213
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Barker, C., Pistran, N., & Elliot, R. (1994). *Research methods in clinical and counselling psychology*. Chichester, England: Wiley.
- Byrd, A. L., Hawes, S. W., Loeber, R., & Pardini, D. A. (2018). Interpersonal callousness from childhood to adolescence: Developmental trajectories and early risk factors. *Journal of Clinical Child & Adolescent Psychology, 47*, 467-482. doi:10.1080/15374416.2016.1144190
- Byrd, A. L., Loeber, R., & Pardini, D. A. (2012). Understanding desisting and persisting forms of delinquency: The unique contributions of disruptive behavior disorders and interpersonal callousness. *Journal of Child Psychology and Psychiatry, 53*, 371-380. doi:10.1111/j.1469-7610.2011.02504.x
- Cardinale, E. M., & Marsh, A. A. (2017). The reliability and validity of the Inventory of Callous Unemotional Traits: A meta-analytic review. *Assessment*. Advance online publication. doi:10.1177/1073191117747392
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255. doi:10.1207/S15328007SEM0902\_5
- Ciucci, E., & Baroncelli, A. (2014). The emotional core of bullying: Further evidences of the role of callous–unemotional traits and empathy. *Personality and Individual Differences, 67*, 69-74. doi:10.1016/j.paid.2013.09.033
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309-319. doi:10.1037//1040-3590.7.3.309
- Colins, O. F., Andershed, H., Frogner, L., Lopez-Romero, L., Veen, V., & Andershed, A. K. (2014). A new measure to assess psychopathic personality in children: The Child Problematic Traits Inventory. *Journal of Psychopathology and Behavioral Assessment, 36*, 4-21. doi:10.1007/s10862-013-9385-y
- 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 & Human Development*, 47, 679-690. doi:10.1007/s10578-015-0601-8
- Conway, J. M., Lievens, F., Scullen, S. E., & Lance, C. E. (2004). Bias in the correlated uniqueness model for MTMM data. *Structural Equation Modeling*, 11, 535-559. doi:10.1207/s15328007sem1104\_3
- De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, 141, 858-900. doi:10.1037/a0038498
- De Los Reyes, A., & Kazdin, A. E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study. *Psychological Bulletin*, 131, 483-509. doi:10.1037/0033-2909.131.4.483
- Deng, J., Wang, M.-C., Zhang, X., Shou, Y., & Gao, Y. (2019). The Inventory of Callous Unemotional Traits: A reliability generalization meta-analysis. *Psychological Assessment*. Advance online publication. doi:10.1037/pas0000698
- 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, 257-278. doi:10.1002/jclp.22313
- DiStefano, C. (2002). The impact of categorization with confirmatory factor analysis. *Structural Equation Modeling*, 9, 327-346. [http://dx.doi.org/10.1207/S15328007SEM0903\\_](http://dx.doi.org/10.1207/S15328007SEM0903_)
- Essau, C. A., Sasagawa, S., & Frick, P. J. (2006). Callous-unemotional traits in community sample of adolescents. *Assessment*, 13, 454-469. doi:10.1177/1073191106287354
- Fanti, K., 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. doi:10.1007/s10862-008-9111-3
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9, 466-491. doi:10.1037/1082-989X.9.4.466
- Frick, P. J. (2004). *The Inventory of Callous-Unemotional Traits*. New Orleans, LA: The University of New Orleans.
- Frick, P. J. (2009). Extending the construct of psychopathy to youths: Implications for understanding, diagnosing, and treating antisocial children and adolescents. *Canadian Journal of Psychiatry*, 54, 803-812. doi:10.1177/070674370905401203
- Frick, P. J., Kimonis, E. R., Dandreaux, D. M., & Farell, J. M. (2003). The 4 year stability of psychopathic traits in non-referred youth. *Behavioral Science Law*, 21, 713-736. doi:10.1002/bsl.568
- Frick, P. J., Ray, J. V., Thornton, L. C., & Kahn, R. E. (2014). Can callous-unemotional traits enhance the understanding, diagnosis, and treatment of serious conduct problems in children and adolescents? A comprehensive review. *Psychological Bulletin*, 140, 1-57. doi:10.1037/a0033076
- 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. doi:10.1007/s10862-015-9527-5
- Hawes, S. W., Byrd, A. L., Gonzalez, R., Cavanagh, C., Bechtold, J., Lynam, D. R., & Pardini, D. A. (2017). The developmental course of psychopathic features: Investigating stability, change, and long-term outcomes. *Journal of Research in Personality*, 77, 83-89. doi:10.1016/j.jrp.2018.09.009
- 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. doi:10.1037/a0034718
- 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. doi:10.1007/s10862-012-9324-3
- 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, 87-97. doi:10.1037/b0000003
- Kimonis, E. R., Fanti, K. A., Mertan, B., Goulter, N., & Katsimicha, E. (2016). Can callous-unemotional traits be reliably measured in preschoolers? *Journal of Abnormal Child Psychology*, 44, 625-638. doi:10.1007/s10802-015-0075-y
- 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. doi:10.1016/j.ijlp.2008.04.002
- Kline, R. B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press.
- Kraemer, H. C., Measelle, J. R., Ablow, J. C., Essex, M. J., Boyce, W. T., & Kupfer, D. J. (2003). A new approach to integrating data from multiple informants in psychiatric assessment and research: Mixing and matching contexts and perspectives. *American Journal of Psychiatry*, 160, 1566-1577. doi:10.1176/appi.ajp.160.9.1566
- Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods*, 7, 228-244. doi:10.1037/1082-989X.7.2.228
- Lee-Rowland, L. M., Barry, C. T., Gillen, C. T., & Hansen, L. K. (2017). How do different dimensions of adolescent narcissism impact the relation between callous-unemotional traits and self-reported aggression? *Aggressive Behavior*, 43, 14-25. doi:10.1002/ab.21658
- Liu, Y., Millsap, R. E., West, S. G., Tein, J. Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22, 486-506. doi:10.1037/met0000075
- 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, 155-165. doi:10.1037/0021-843X.116.1.155

- Lynam, D. R., Charnigo, R., Moffitt, T. E., Raine, A., Loeber, R., & Stouthamer-Loeber, M. (2009). The stability of psychopathy across adolescence. *Development and Psychopathology, 21*, 1133-1153. doi:10.1017/S0954579409990083
- Marsh, H. W., Muthén, B., Asparouhov, T., Ludtke, O., Robitzsch, A., Morin, A. J. S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439-476. doi:10.1080/10705510903008220
- Obradović, J., Pardini, D. A., Long, J. D., & Loeber, R. (2007). Measuring interpersonal callousness in boys from childhood to adolescence: An examination of longitudinal invariance and temporal stability. *Journal of Clinical Child & Adolescent Psychology, 36*, 276-292. doi:10.1080/15374410701441633
- Paiva-Salisbury, M. L., Gill, A. D., & Stickle, T. R. (2017). Isolating trait and method variance in the measurement of callous and unemotional traits. *Assessment, 24*, 763-771. doi:10.1177/1073191115624546
- Pechorro, P., Gonçalves, R. A., Hawes, S. W., & Ray, J. V. (2017). Psychometric properties of two short versions of the Inventory of Callous-Unemotional Traits among incarcerated youth. *Journal of Emotional and Behavioral Disorders, 26*, 243-256. doi:10.1177/1063426617717940
- Pechorro, P., Hawes, S. W., Gonçalves, R. A., & Ray, J. V. (2017). Psychometric properties of the Inventory of Callous-Unemotional Traits short version (ICU-12) among detained female juvenile offenders and community youths. *Psychology, Crime & Law, 23*, 221-239. doi:10.1080/1068316X.2016.1239724
- Pihet, S., Etter, S., Schmid, M., & Kimonis, E. R. (2015). Assessing callous-unemotional traits in adolescents: Validity of the Inventory of Callous-Unemotional Traits across gender, age, and community/institutionalized status. *Journal of Psychopathology and Behavioral Assessment, 37*, 407-421. doi:10.1007/s10862-014-9472-8
- Pitts, S. C., West, S. G., & Tein, J. Y. (1996). Longitudinal measurement models in evaluation research: Examining stability and change. *Evaluation and Program Planning, 19*, 333-350. doi:10.1016/S0149-7189(96)00027-4
- Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*, S69-S77. doi:10.1097/01.mlr.0000245438.73837.89
- Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- 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. doi:10.1037/pas0000183
- Roose, A., Bijttebier, P., Decoene, S., Claes, L., & Frick, P. J. (2010). Assessing the affective features of psychopathy in adolescence: A further validation of the Inventory of Callous and Unemotional Traits. *Assessment, 17*, 44-57. doi:10.1177/1073191109344153
- Soto, C. J., John, O. P., Gosling, S. D., & Potter, J. (2008). The developmental psychometrics of Big Five self-reports: Acquiescence, factor structure, coherence, and differentiation from ages 10 to 20. *Journal of Personality and Social Psychology, 94*, 718-737. doi:10.1037/0022-3514.94.4.718
- Waller, R., Wright, A. G., Shaw, D. S., Gardner, F., Dishion, T. J., Wilson, M. N., & Hyde, L. W. (2015). Factor structure and construct validity of the parent-reported Inventory of Callous-Unemotional Traits among high-risk 9-year-olds. *Assessment, 22*, 561-580. doi:10.1177/1073191114556101
- Wang, M.-C., Colins, O. F., Deng, Q., Deng, J., Huang, Y., & Andershed, H. (2018). The Child Problematic Traits Inventory in China: A multiple informant-based validation study. *Psychological Assessment, 30*, 956-966. doi:10.1037/pas0000545
- Wang, M.-C., Gao, Y., Deng, J., Lai, H., Deng, Q., & Armour, C. (2017). The factor structure and construct validity of the Inventory of Callous-Unemotional Traits in Chinese undergraduate students. *PLoS ONE, 12*, e0189003. doi:10.1371/journal.pone.0189003
- West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209-246). New York, NY: Guilford Press.
- Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives, 4*, 10-18. doi:10.1111/j.1750-8606.2009.00110.x
- Willoughby, M. T., Mills-Koonce, W. R., Waschbusch, D. A., & Gottfredson, N. C., & Family Life Project Investigators. (2015). An examination of the parent report version of the Inventory of Callous-Unemotional Traits in a community sample of first-grade children. *Assessment, 22*, 76-85. doi:10.1177/1073191114534886
- Wu, P. C. (2017). Longitudinal measurement invariance of Beck Depression Inventory-II in early adolescents. *Assessment, 24*, 337-345. doi:10.1177/1073191115608941
- Zhang, X., Wang, M.-C., Shou, Y., Luo, J., & Deng, J. (2019). *Psychometric evaluation of the ICU and its shorten version in a sample of Chinese detained juvenile offenders*. Manuscript submitted for publication.