

## DOCUMENT SUMMARY

This research article details a study to validate the **Michigan Limited Prosocial Emotion Addendum (M-LPE)**, a semi-structured interview method for assessing **callous-unemotional (CU) traits**, also known as the **Limited Prosocial Emotions (LPE)** specifier in **DSM-5**. The study found the M-LPE to have acceptable inter-rater reliability and demonstrated convergent, construct, and incremental validity when compared to other measures in a sample of at-risk, low-income, and primarily African American adolescents. The results support the M-LPE as a promising assessment tool for the **LPE** specifier in clinical research settings.

## FILENAME

Walker\_2021\_clinical\_research\_article\_lpe\_assessment\_validation

## METADATA

Category: CLINICAL

Type: research\_article

Relevance: Reference

Update Frequency: Static

Tags: #limited-prosocial-emotions #lpe #callous-unemotional-traits #cu-traits #assessment #k-sads-pl #dsm-5 #conduct-disorder #psychometrics #m-lpe

Related Docs: N/A

Supersedes: N/A

## FORMATTED CONTENT

### **Walker\_2021\_Psychometric\_Properties\_of\_a\_Semi-Structured Interview to Assess Limited Prosocial Emotions**

Toni M. Walker<sup>1</sup>, Paul J. Frick<sup>1,2</sup>, Tatiana M. Matlasz<sup>1</sup>, Emily L. Robertson<sup>1</sup>, Amy J. Mikolajewski<sup>3</sup>, Colter Mitchell<sup>4</sup>, Nestor Lopez-Duran<sup>5</sup>, Christopher S. Monk<sup>4,5</sup>, Luke W. Hyde<sup>4,5</sup>

<sup>1</sup>Department of Psychology, Louisiana State University

<sup>2</sup>Institute for Learning Sciences and Teacher Education, Australian Catholic University

<sup>3</sup>Department of Psychiatry and Behavioral Sciences, Tulane University

<sup>4</sup>Survey Research Center, Institute for Social Research, University of Michigan

<sup>5</sup>Department of Psychology, University of Michigan

## Abstract

**Callous-unemotional (CU) traits** have recently been added to the diagnostic criteria of **Conduct Disorder** in the **DSM-5** and of **Conduct-dissocial and Oppositional Defiant Disorders** in the **ICD-11** as the **Limited Prosocial Emotions** specifier.

This change necessitates the assessment of these traits with validated measures in both research and clinical contexts.

The current study sought to validate a semi-structured diagnostic interview method, the **Michigan Limited Prosocial Emotion Addendum (M-LPE)** to the **K-SADS-PL**, of assessing **CU traits** based on a recently developed clinician rating system (**CAPE 1.1**) in a sample of at-risk youth. Results supported the inter-rater reliability of the **M-LPE** with moderate agreement and high reliability between raters. The **M-LPE** demonstrated convergent and incremental validity with **CU traits** and various measures of antisocial behavior. The results provide preliminary evidence for the use of a semi-structured interview assessment of **CU traits** in research contexts and build the foundation for further validation.

**Keywords:** callous-unemotional traits; limited prosocial emotions; assessment; structured diagnostic interview; psychometric properties

Research has consistently shown that conduct problems can vary greatly in their severity, stability, and causes in children and adolescents (Frick, 2012). Substantial research has further suggested that the presence of elevated levels of **callous-unemotional (CU) traits** may be beneficial in explaining some of this heterogeneity. **CU traits** have been conceptualized as a downward extension of the affective features of psychopathy or the affective components of conscience and are defined by four key components: absence of guilt or remorse, a callous-lack of empathy, a failure to put forth effort in important activities, and a constricted display of affect (Hare & Neumann, 2008; Kimonis et al., 2015). Research has shown that youth with serious conduct problems who are elevated on these traits show a more severe and chronic pattern of antisocial behavior, and they show different cognitive, biological, environmental, and temperamental risk factors compared to other youth with conduct problems (Frick, Ray, Thornton, & Kahn, 2014). Further, youth with elevated **CU traits** start with more severe conduct problems and, while their behavior improves with treatment, often still show more severe behavior problems after treatment relative to other youth with serious conduct problems (Frick et al., 2014; Hawes, Price, & Dadds, 2014; Hyde, Waller, & Burt, 2014). As a result of this research, **CU traits** have recently been added to the major classification systems for diagnosing children with serious conduct problems. Specifically, the **Diagnostic and Statistical Manual of Mental Disorders -5th Edition (DSM- 5)** now includes a specifier for **Conduct Disorder (CD)** for those high on **CU traits** called "**with Limited Prosocial Emotions**" (**LPE**) (American Psychiatric Association, 2013) and the **International Classification of Disease 11th edition (ICD-11)** includes this as a potential specifier for the diagnoses of **Conduct-dissocial Disorder** and **Oppositional Defiant Disorder (ODD)** (World Health Organization, 2018).

This recent inclusion in the major classification systems for mental health diagnoses likely will lead to increases in the frequency with which these traits are assessed in a variety of research and clinical settings. To date, research has typically relied on multi-informant rating scales to assess **CU traits** (Frick & Ray, 2015; Kotler & McMahon, 2005). This assessment method is particularly beneficial because the measures are time-efficient, cost-effective, require little to no training to administer, and result in highly reliable scores (Frick & Ray, 2015). One of the most commonly used measures to assess **CU traits** in research is the **Inventory of Callous-Unemotional Traits (ICU)**; (Kimonis et al., 2008). The **ICU** is a 24-item behavior rating scale that includes forms for self-report, as well as parent- and teacher-report. The **ICU** was developed to a) provide a focused and comprehensive assessment of **CU traits** only (and not other dimensions of psychopathy), b) include a rating format that allows for sufficient variability in responses without including a central tendency point (i.e., items are anchored on a four-point Likert scale from 0 (Not at all true) to 3 (Definitely true), and c) include equal numbers of items rated in the positive (i.e., higher rating indicating higher levels of **CU traits**) and negative (i.e., higher ratings indicating lower levels of **CU traits**) directions (Frick & Ray, 2015). To date, the **ICU** has been translated into over 25 different languages and has been used widely in research, with over 200 published studies in samples ranging in age from 3 years to young adulthood (Frick & Ray, 2015). Lastly, a meta-analysis from 115 samples (n=27,947) reported that the **ICU** total score generally showed adequate internal consistency (pooled  $\alpha=0.83$ ; positive correlations with aggression (pooled  $r=.41$ ), delinquency (pooled  $r=.34$ ), and externalizing behaviors (pooled  $r=.34$ ); and negative correlations with measures of empathy (pooled  $r=-.42$ ; Cardinale & Marsh, 2017).

Although questionnaires, like the **ICU**, have proven to be beneficial in the research context, there are still limitations in relying solely on rating scales for clinical decision making. First, rating scales do not allow the clinician to assess if the informant understood the questions and was answering them in the way that they were intended. Second, most of these measures do not include clinical cut-off scores to help determine if the levels of **CU traits** are impairing and/or non-normative (Kimonis, Fanti, & Singh, 2014). Finally, with the exception of the **ICU** (Kimonis et al., 2015), most scales do not directly assess the symptoms used to define **LPE** in the **DSM-5** diagnostic criteria (see Kotler & McMahon, 2005).

To overcome these limitations, the **Clinical Assessment of Prosocial Emotions, Version 1.1 (CAPE 1.1)**; Frick, 2013) was developed as a clinician rating system for use in a wide range of clinical settings. This measure uses semi-structured clinical interviews combined with professional judgment to assess the diagnostic criteria for the **LPE** specifier in youth ages 3 to 21 years old. The development of this assessment was guided by research and closely tied to the way **CU traits** are measured by both the **ICU** and the **DSM-5** criteria (Frick, 2013). The **CAPE 1.1** includes semi-structured interviews that are designed to be completed with both the child and another informant (e.g., primary caretaker) separately. These interviews consist of nine stem questions (e.g., "Does show his or her feelings openly to others?"), followed by requests for examples, and additional supplementary questions (e.g., "Is this how he/she is most of the time and with most people?") that assess each of the four aforementioned diagnostic criteria

of the **Limited Prosocial Emotions** specifier. Informants are encouraged to respond to the stem questions in “yes/no” fashion and then provide additional examples to aid making the final ratings. Based on the responses and examples provided, a highly trained clinician can then follow-up with any questions they feel are needed in order to rate the child on each symptom using a three-point scale from 0 (Not at all or mildly descriptive), 1 (moderately descriptive), or 2 (Highly descriptive). This final professional judgment on the presence of symptoms is based on information from multiple informants (at least the interviews with the child and another informant) and clinical information from other sources.

The **CAPE 1.1** has been subjected to recent psychometric evaluation and shown promising reliability and validity of symptom counts and diagnostic cutoffs in international samples of high-risk (Centifanti et al., 2019) and detained youth (Molinuevo et al., 2019). Hawes, Kimonis, Mendoza Diaz, Frick, and Dadds (2020) found evidence to support the reliability and validity of the **CAPE 1.1** in a clinic-referred sample of children and adolescents (3 to 15 years) with conduct problems. **CAPE 1.1** scores were associated with established correlates of **CU traits** (i.e., **ODD** and **CD** symptom severity, proactive aggression, and affective empathy). Also, youth with diagnoses of **ODD** and **CD** with or without the **LPE** specifier differed significantly on maternal ratings of affective empathy.

While the **CAPE 1.1** is a promising clinical tool to assess **CU traits** as defined by the **LPE** specifier (Hawes, Kimonis, Diaz, Frick, & Dadds, 2019; Molinuevo et al., 2019), it requires extensive training in the assessment of psychopathology, extensive training in the **CAPE** specifically, and a fair amount of time in asking follow-up questions specifically about **LPE**. Such a clinician rating system may not be feasible in many clinical or research settings where the child needs to be assessed comprehensively for psychopathology by a semi-structured interview and without a clinician with extensive training. Thus, to bridge the gap between rating scales used in research and a clinician rating like the **CAPE 1.1**, the current study tested the **Michigan Limited Prosocial Emotional Addendum (M-LPE)**, a semi-structured interview method for assessing the **LPE** based on the **CAPE 1.1**. The **M-LPE** was developed to be administered as part of the **Schedule for Affective Disorders and Schizophrenia in Children-Present and Lifetime Version (K-SADS-PL)**; Kaufman et al., 1997, 2016) embedded in the **CD** section with a similar structure. Thus the **M-LPE** uses the stem questions from the **CAPE 1.1** (see <https://sites01.lsu.edu/faculty/pfricklab/cape/items>) and then, as in the broader **K-SADS**, offers additional **CAPE** questions as follow-up questions as needed, but excludes the request for examples and does not require that the clinician use all of the follow-up questions used by the **CAPE 1.1**. That is, the **M-LPE** adds content from the **CAPE** to screen for and assess **LPE** within the **K-SADS-PL**. Thus, the **M-LPE** can be used by trained researchers without extensive clinical experience and can be administered in a much shorter period of time as part of an overall diagnostic interview (10-15 minutes for **M-LPE** vs. 40-50 minutes for the **CAPE 1.1**).

## Current Study

In this study, we assessed the inter-rater reliability and conducted initial tests of the validity of the **M-LPE**. First, we tested the inter-rater agreement between initial scores made from the original interviewer at a case conference (see below) and scores made by a second rater who watched only videotaped recordings of the interviews with parents and children and made ratings without any discussion. These second raters had been trained on the **CAPE 1.1**. Second, we tested the convergent validity between the **M-LPE** and self- and parent-report versions of the **ICU**. This aim is a critical test of the validity of the **M-LPE**, given that much of the research supporting the validity of the construct of **CU traits** used the **ICU** or other rating scales (Frick et al., 2014). Third, we tested the validity of the **M-LPE** scores with measures of externalizing and internalizing symptoms, to determine if **LPE** scores would be positively correlated with measures of various types of externalizing behavior. Fourth, we tested the incremental validity of **M-LPE** scores by testing whether the associations with measures of construct validity (i.e., externalizing and antisocial behaviors) would remain when controlling for ratings of **CU traits** on the **ICU**. This aim provided a critical test to determine if the use of a semi-structured interview assessment provided useful information in the prediction of important criteria, over and above more time-efficient rating scales in a sample of mostly African American participants, oversampled to include more adolescents at socioeconomic risk.

## Method

### Participants

Participants were drawn from a sample adolescents from Detroit, Toledo, and Chicago who participated in the Study of Adolescent Neural Development (SAND) at the University of Michigan (Goetschius et al., 2019; Hein et al., 2018). This SAND is a sub-study of the Fragile Families and Child Wellbeing Study (FFCWS; Reichman, Teitler, Garfinkel, & McLanahan, 2001), a representative, longitudinal cohort of 4,898 children (52.4% boys) born in 20 major U.S. cities between 1998 and 2000 that was recruited from urban hospitals and oversampled for non-marital births (~3:1). This sample contains substantial representation of African American youth, as well as adolescents from families living in low-income contexts. Members of the SAND research team attempted to contact all of the families from the original Detroit and Toledo subsamples of FFCWS to take part in additional data collection at the University of Michigan as part of the SAND study when focal child was 15 years old. The team also contacted a small subset of families from the Chicago subsample to increase the total number of participants. In total, 237 of the 513 families that the team attempted to contact participated in the SAND data collection. The University of Michigan Medical School Institutional Review Board approved this study (UM IRBMED: HUM00074392). All adolescent participants provided written informed assent, and their primary caregivers provided written consent for both themselves and their adolescent children, after the study was explained and questions were answered.

There were no significant differences between the SAND sample and the original FFCWS sample from those cities on measures of maternal education, family income, and maternal marital status. The sample for the current analyses consisted of 144 parent-teen dyads from the age 15 wave of the SAND study. Participants were included



in the current study if the **M-LPE** was administered and videotaped for both the primary caregiver and child, and the audio was sufficient to allow for independent coding of responses. Though 237 participated in the SAND study, a subset of families (N=52) did not have full high-quality videotapes of the interviews for both parent and child because the video was cut off prior to the administration of the **M-LPE** items. Additionally, 32 videos did not have codeable responses for both parent and child report of each symptom. Of the 144 adolescents who were included in the present analyses, 51.4% were female, 76.4% were Black/African American, 16% were White/European-American, and 42% of families reported annual income below \$25,000. The primary caregivers reporting on the teens' symptoms were biological mothers (89.6%), biological fathers (4.9%), adoptive mothers (2.1%), and other relatives (3.5%). The sample used in the current analyses were not significantly different from the full SAND sample on maternal education and marital status. The sample in the current analyses did have marginally significantly lower scores on CBCL Rule Breaking (M=1.14) at age 15 than those not included (M=1.69);  $t(232)=1.97$ ,  $p=.05$ . The samples also significantly differed on race ( $X^2(5, N=237)=16.39$ ,  $p<.01$ ), with the sample included in the current analyses having a smaller proportion of minorities (84%) and higher proportion of White/European-American (15.9%) than the proportion of minorities (92.4%) and White/European-Americans (7.5%) not included.

## Procedure

IRB approval for the study was obtained by all research sites where the data was collected and analyzed. The **M-LPE** was scored and interviews were recorded at time of data collection. Combined total lifetime symptom count (i.e., both past and present clinical threshold symptoms) of **DSM-5 Limited Prosocial Emotions** (American Psychiatric Association, 2013) based on clinician-ratings assessed via a modified version of the **Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS)**; Kaufman et al., 1997). A trained clinical interviewer (e.g., psychology doctoral student, post-baccalaureate staff) administered the semi-structured interview to the target child and primary caregiver each individually. Assessors were trained by two licensed clinical psychologists with 25+ years of combined experience with the **K-SADS** but with no experience with the **CAPE** (authors LWH, NLD). Training for the **K-SADS** broadly included practice interviews and live supervision of interviews with families. The interviewer arrived at initial symptom ratings, symptom counts, and **DSM-5** diagnoses for each informant, which were then reviewed in case conferences with two licensed clinical psychologists and the assessment team to determine the best report score for the aforementioned ratings. Training for the **M-LPE** included reading the **CAPE 1.1** manual, but nothing additional. Symptoms are rated on a 3-point scale (0= not present; 1= present at subclinical; 2= present at clinical threshold). The secondary coders (e.g. graduate students, post-doctoral scholar) were from a different university who did not participate in the initial interviews but had more extensive experience with the **CAPE 1.1**. These coders reviewed the recorded videos of the parent interview and child interview and coded symptoms on a similar 3-point scale based on the **DSM-5 LPE** criteria. As a result, these reliability codes did not involve a clinical case conference. The three point scores were used to create the dichotomous CU diagnosis variable, which was coded as "present" if at least two symptoms were coded as present, and "not

present" if only one or no symptoms were coded as present, consistent with the **DSM-5** criteria for the **LPE** specifier (American Psychological Association, 2013).

## Measures

**Callous-Unemotional Traits.**—**CU traits** were assessed with the **Inventory of Callous Unemotional Traits (ICU)**; Kimonis et al., 2008). As described above, the **ICU** is a 24-item rating scale that assesses a wide range of indicators of **CU traits** and contains equal numbers of items worded in the positive (meaning higher levels of **CU traits**; e.g., "I do not feel remorseful when I do something wrong") and negative (meaning lower levels of **CU traits**; e.g., "I am concerned about the feelings of others") directions. To create a total score, the negatively-worded items are recoded so that higher scores indicate higher levels of **CU traits**. The current analyses also utilized a resolved **ICU** total score, created by taking the higher score between the parent- and child-reports for each symptom and summing them. As noted above, the total **ICU** score has been consistently associated with antisocial behavior (positively) and empathy (negatively) across a range of adolescent samples (Cardinale & Marsh, 2017). Internal consistency in this sample was acceptable for both parent-report ( $\alpha = .78$ ), self-report ( $\alpha = .78$ ), and the resolved score ( $\alpha = .80$ ).

**Conduct Problems.**—Conduct problems were measured using behavior rating scales and a semi-structured clinical interview. Parent- and child-reported behavior ratings of conduct problems were assessed using the Youth Self Report (YSR) and Child Behavior Checklist (CBCL) scales of the Achenbach System of Empirically Based Assessment measures (Achenbach, 2009). A Rule Breaking (RB) syndrome scale was created by summing 12 items from each measure generally covert conduct problems (e.g., "Steals outside of the home", "Runs away from home"). The Aggressive Behavior (AGG) syndrome scale consists of 20 items that generally measure more overt conduct problems and aggression (i.e., "Destroys things belonging to his/her family or others", "Physically attacks people", "Threatens people"). For both scales, participants rated each item on a 3-point Likert scale from 0 (Not true) to 2 (Very true or often true), which were summed to create a total score. Internal consistency for the RB scale was acceptable for both the CBCL ( $\alpha = .69$ ) and the YSR ( $\alpha = .68$ ). Internal consistency was strong for both the CBCL AGG scale ( $\alpha = .86$ ) and the YSR AGG scale ( $\alpha = .85$ ). Previous studies have demonstrated external validity of the Rule Breaking and Aggressive Behavior scales through strong correlations with **ODD** and **CD** and significant predictive validity of overt and covert **CD** dimensions measured by semi-structured interviews (Gomez, Vance, & Gomez, 2014; Tackett, Kreuger, Sawyer, & Graetz, 2003). Moderate associations have also been found between both scales and **CU traits** in both boys (RB=.47; AGG=.42) and girls (RB=.43; AGG=.44; Charles, Acheson, Mathias, Furr, & Dougherty, 2012).

The **Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS-PL)** was used as a measure of **DSM-5** symptoms of conduct problems (Kaufman et al., 1997). The **K-SADS-PL** is a semi-structured interview that assesses the diagnostic criteria of the major DSM-IV disorders that are displayed by children and adolescents. Since there was no change in the actual symptoms of **CD** and **ODD** between the 4th and 5th

editions of the DSM, the method of scoring did not need to change. The youth and their parent independently reported on the youth's symptoms of **CD** and **ODD**. For each informant rating of the symptom, interviewers coded the symptoms as 0 (Not present), 1 (Sub-clinical), or 2 (Present). The parent and youth's ratings on each symptom were used to create a summary "best" rating by clinical judgement, which was the consensus decision from the research team based on all the information presented, as described above for the assessment of the **LPE** specifier. We created a symptom count variable for all above threshold present endorsements for each symptom of **CD** and **ODD**. Of the current sample, only 1.4% (N=2) and 4.2% (N=6) met criteria for past and present **CD** and **ODD**, respectively.

**Aggression.**—Aggression was assessed with the self-report **Reactive-Proactive Aggression Questionnaire (RPQ)**; Raine et al., 2006). Participants rated the frequency of both their reactive and proactive forms of verbal and physical aggression using a 3-point Likert scale from 0 (Never) to 2 (Often). The 26 items (13 reactive and 13 proactive) were summed to create a total aggression score. Internal consistency for this scale was strong ( $\alpha = .85$ ). Previous research shows that proactive and reactive aggression as measured by the **RPQ** are significantly associated with **CU traits**, psychopathy, delinquency, and impulsivity in detained and community samples of adolescents and young adults (Raine et al., 2006; Fanti, Frick, & Georgiou, 2009; Feilhauer, Cima, & Arntz, 2012).

**Delinquency.**—Participant's frequency of antisocial behaviors was measured with the **Self-Report of Delinquency (SRD)**; Elliot & Ageton, 1984). Participants self-reported how often they engaged in each of the 62 items in the past year from 0 (Never) to 2 (More often). Total sum scores were used in the current analyses such that higher scores indicated more engagement in delinquent behavior. Internal consistency for the **SRD** in this sample was strong ( $\alpha = .85$ ). Previous studies have shown significant correlations between self-reported delinquency and official records of delinquent involvement and arrests (Huizinga & Elliott, 1986). The positive relationship between self-reported **CU traits** and this measure of delinquency is well-established (Ansel, Barry, Gillen, & Herrington, 2014; Frick, Stickle, Dandreaux, Farrell, & Kimonis, 2005).

### **Analytic Plan**

The inter-rater reliability of the **M-LPE** was assessed in three ways. First, the absolute level of agreement between the interviewer and coders on the presence or absence of each symptom and a diagnosis of **LPE** was estimated. Second, the phi coefficient was used to determine the statistical significance and effect size of the level of association between the two raters' scores, with scores below 0.3 considered weak, 0.3 to 0.5 moderate, and 0.5 and above strong (McHugh, 2018). Third, the level of agreement between raters for the total symptom counts was assessed through the intraclass correlation, which not only considers the level of association (i.e., correspondence between the relative level between the two raters), but the correspondence in the absolute value of the ratings.

To test the convergent validity of the **M-LPE**, total symptom counts from the **M-LPE** were correlated with the **ICU**. Correlations were also used to determine construct



validity of the **M-LPE** symptom counts, by testing their correlations with measures of various types of externalizing behaviors: **CD** and **ODD** symptoms, delinquency, and aggression. Finally, simultaneous multiple regressions were run to test the incremental association of the **M-LPE** symptom count with externalizing behaviors, delinquency, and aggression after controlling for **ICU** scores. The **M-LPE** symptom counts and **ICU** scores were used as independent variables, with the measures of externalizing behaviors, delinquency, and aggression as the dependent measures.

## Results

### Reliability

The first aim of the current study was to examine the inter-rater agreement between the interviewers' initial ratings and the coders' ratings on symptoms of **CU traits** as measured by the **M-LPE**. Of note, the **LPE** diagnosis had a very low base rate in this sample and was rated as present 3.5% (N=5) of the time by interviewer and 1.4% (N=2) by coder regardless of whether they met criteria for a **CD** diagnosis. There was 98% agreement between interviewer and coders' ratings on the presence of a **Limited Prosocial Emotions** diagnosis and this high level of agreement is reflected by a strong phi coefficient ( $\phi = .63$ ,  $p < .001$ ). However, given the low base rate of the diagnosis, the reliability of the overall diagnosis should be interpreted with caution and more attention should be given to the reliability of the symptom counts, which did show more variability. Specifically, a high degree of reliability was also found between the interviewer and coder ratings of total symptom counts, with the average measure ICC at .75 ( $p < .001$ ) with a 95% confidence interval from .65 to .82. For the individual symptoms, the prevalence rates for the interviewer ranged from 0.7% for the parent reported "lack of empathy" symptom to 6.9% for the self-reported "shallow affect" symptom. Prevalence rates for the coder ranged from 0% for parent reported "shallow affect" to 3.5% for self-reported "lack of remorse". The level of agreement for individual symptoms ranged from 93% for the self-reported "shallow affect" symptom to 99% for both the parent reported "lack of remorse" and best report "lack of empathy" symptom, with phi coefficients ranging from .27 to .81 for these symptoms (all  $ps < .01$ ).

### Validity

The second aim was to test the convergent validity of the **M-LPE** with the **ICU**. The zero-order correlations between the **M-LPE** symptom counts and self-reported ( $r = .29$ ,  $p < .01$ ), parent-reported ( $r = .27$ ,  $p < .01$ ), and resolved ( $r = .36$ ,  $p < .01$ ) **ICU** scores are provided in Table 1. These correlations show a moderate level of convergence across the two methods for assessing **CU traits**.

Table 1 also displays the results addressing the third aim, which was to test the construct validity of the **M-LPE** scores by their associations with various measures of externalizing behaviors. As expected, **M-LPE** symptom count was significantly positively associated with self-reported delinquency ( $r = .18$ ,  $p < .05$ , **RPQ** aggression ( $r = .37$ ,  $p < .01$ ), parent-reported rule breaking ( $r = .31$ ,  $p < .01$ ), parent-reported aggression ( $r = .18$ ,  $p < .05$ ), **ODD** symptoms ( $r = .28$ ,  $p < .01$ ), and **CD** symptoms ( $r = .40$ ,  $p < .01$ ). Contrary to expectations, the **M-LPE** symptom count was not related to youth-reported rule

breaking or aggression. Table 1 also reveals that the **M-LPE** and **ICU** scores show different correlations with measures of antisocial behavior. Comparisons of the correlates using Fisher's r-to-z transformation for dependent correlations (Lee & Preacher, 2003) showed that **M-LPE** and **ICU** Resolved scores had significantly different correlations with YSR Rule Breaking ( $z = 2.45$ ;  $p < .05$ ), YSR Aggression ( $z = 2.40$ ;  $p < .05$ ), CBCL Aggression ( $z = 2.48$ ,  $p < .05$ ), with the **ICU** showing stronger associations. The **ICU** Child and **M-LPE** differed in their relationship with YSR Rule Breaking ( $z = 2.12$ ,  $p < .05$ ), YSR Aggression ( $z = 2.50$ ;  $p < .05$ ), **K-SADS-PL CD** symptoms ( $z = 2.05$ ,  $p < .05$ ), and delinquency ( $z = 2.36$   $p < .05$ ), again with the **ICU** showing stronger associations. Finally, the **M-LPE** showed stronger associations with the **K-SADS-PL CD** symptoms than the **ICU** Child ( $z = 2.05$ ,  $p < .05$  and Parent ( $z = 2.53$ ,  $p < .05$  report).

Finally, the fourth aim of the current study was to test the incremental validity of the **M-LPE** symptom count, relative to the more time efficient ratings from the **ICU**. The results of these analyses are provided in Table 2 and should be interpreted in light of the zero-order correlations presented in Table 1. In the first two rows of the table, the results of the multiple regression analyses testing the incremental association of the **M-LPE** symptom count controlling for parent-reported **ICU** are provided. The **M-LPE** symptom count accounted for a significant incremental portion of the variance in **K-SADS-PL ODD** symptoms ( $b^* = .22$ ,  $p < .01$ ) and **K-SADS-PL CD** symptoms ( $b^* = .38$ ,  $p < .001$ ). The **M-LPE** symptom count also accounted for a significant incremental variance in youth-reported aggression ( $b^* = .33$ ,  $p < .01$  over and above parent-reported **CU traits**). In the second section of Table 2, the results of the multiple regression analyses testing the incremental contribution of the **M-LPE** symptom count controlling for self-reported **CU traits** are provided. When entered in the model with self-reported **CU traits**, **M-LPE** symptoms again accounted for incremental variance in parent-reported rule breaking behavior ( $b^* = .30$ ,  $p < .001$ ), **K-SADS-PL ODD** ( $b^* = .25$ ,  $p < .01$ ), and **CD** ( $b^* = .39$ ,  $p < .001$ ) symptoms, and self-reported aggression ( $b^* = .30$ ,  $p < .01$ ). When entered into the model with a resolved **ICU** total score, **M-LPE** symptoms accounted for incremental variance in parent-reported rule breaking ( $b^* = .23$ ,  $p < .01$ ), **K-SADS-PL CD** symptoms ( $b^* = .36$ ,  $p < .001$ ), and aggression ( $b^* = .27$ ,  $p < .05$ ).

## Discussion

In the current study, we provided initial data on the reliability and validity (i.e., convergent, construct, and incremental) of a structured interview method, the **M-LPE**, for assessing the **LPE** specifier, which was recently added to the diagnosis of **Conduct Disorder** in the **DSM-5** and **Conduct-dissocial Disorder** or **Oppositional Defiant Disorder** in the **ICD-11**. The **M-LPE** was developed to provide a method for assessing the symptoms of **LPE** embedded within the **K-SADS** in a way that provides more information than self or informant-based rating scales, but that does not require the same level of training, nor administration time, as the full **CAPE**. Further, the reliability coder ratings were purely based on the answers provided by the participants without further discussion of symptom ratings by coders expert in the **CAPE** (as opposed to the original codes which were made after discussion in a clinical case conference but with non-**CAPE** experts). The high reliability between these codes suggests that one would obtain fairly consistent information without the more intensive clinical case conference.

Thus, the **M-LPE** format is well suited for use in many clinical and research settings, particularly as embedded within the **K-SADS** and to be used by clinicians or researchers trained and experienced in the use of the **K-SADS-PL**.

Overall, our findings demonstrated an acceptable degree of inter-rater reliability between the initial ratings made by the interviewer and ratings from re-coded videos of the **M-LPE** in this sample of mostly African American, and low-income adolescents. These results suggest the information gained from the **M-LPE** questions lead to responses that can be interpreted consistently across raters, even without the more extended request for examples and clinician determined follow-up questions used by the **CAPE 1.1**. This level of reliability is consistent with the level of reliability found for the assessment of other forms of psychopathology using the **K-SADS-PL** in various samples of adolescents (Ambrosini, 2000; Kaufman et al., 1997; de la Peña et al., 2018; Lauth, Magnusson, Ferrari, & Petursson, 2008). Thus, the symptoms of the **LPE** specifier, which focuses on the child's emotional and interpersonal style, can be assessed as reliably as overt behaviors, when assessed through a structured interview format (Frick & Nigg, 2012).

Our results also showed the symptom count from the **M-LPE** is modestly correlated with parent- and self-reported ratings on the **ICU**, one of the most common ways that **CU traits** have been assessed in research to date (Frick & Ray, 2015). Of note, these validity coefficients were statistically significant, but modest in size (e.g.,  $r=.20$  and  $.27$ , for child- and parent-report, respectively). However, these are comparable to the low to moderate associations between the NIMH Diagnostic Interview Schedule for Children (DISC; Costello, Edelbrock, & Dulcan, 1984) and parent- ( $r=.29-.31$ ) and teacher- ( $r=.14-.28$ ) ratings of conduct problems (Hodges, 1993). Thus, this low level of agreement may reflect typical levels of association when assessing constructs using different informants and methods (Achenbach, McConaughy, & Howell, 1987; De Los Reyes & Kazdin, 2005). Alternatively, this may reflect differences between information gathered from interviews, during which the interviewer can ensure that the informant is understanding the question appropriately and are able to weigh the report of various informants when making decisions on whether or not a symptom is present. This is supported by the fact that the **M-LPE** scores were more highly correlated with conduct problems assessed by the semi-structured interview (**K-SADS**), whereas **ICU** was more highly correlated with rating scale measures (YSR) of antisocial behavior.

Convergent validity was established through positive associations between **M-LPE** symptom counts and measures of delinquency, **CD** symptoms, and aggression. Specifically, the **M-LPE** was consistently associated with these measures of more severe types of externalizing behaviors at a level consistent with what has been reported in past research. Specifically, Frick et al. (2014) reviewed 118 studies (70 cross-sectional and 48 longitudinal) and reported that the average correlation between **CU traits** and measures of externalizing behaviors was  $.33$ .

Of note, the **M-LPE** explained incremental variance in parent-reported **CD** and **ODD** symptoms, parent-reported rule-breaking, and child-reported

aggression (but not child-reported delinquency), even after controlling for **ICU** scores.

This finding of the incremental utility of the **M-LPE** is particularly important for suggesting that the information gathered from semi-structured interviews adds to the variance explained in certain clinically important outcomes, relative to the more time-efficient behavior ratings. Notably, our findings revealed that the resolved **ICU** score sometimes performed worse than when using scores from a single rater. This may suggest that if one informant is a better reporter than the other, combining their reports mathematically (rather than via clinician-led interview) may decrease the validity of the better reporter.

These results need to be considered within the context of several limitations. Of most importance, the use of a non-referred sample led to a very small number of participants meeting the diagnostic criteria for **ODD** (past diagnosis: n=8,7.1%, current diagnosis: n=7,4.9%), **CD** (past diagnosis: n=9,6.3%, current diagnosis: n=2,1.4%), and the **LPE** specifier (past diagnosis: n=5,3.5%, current diagnosis: n=5,3.5%). This finding is consistent with past studies finding that the rate of children and adolescents meeting the threshold for the **LPE** specifier is below 5% in community samples (Kahn, Frick, Youngstrom, Findling, & Youngstrom, 2012; McMahon et al., 2010; Pardini et al., 2012; Seijas et al., 2018). However, it meant that we could not restrict our analyses to only those who met the **LPE** specifier and who also met criteria for **CD**, as specified in the **DSM-5** criteria. Further, it meant that our tests had to largely focus on the validity of the symptom counts for the **LPE** specifier, rather than on the very low base rate diagnosis. Thus, the reliability and validity of the **M-LPE** will need to be tested in much larger samples with higher rates of **CD** (e.g., clinic-referred and forensic samples), which would lead to higher base rates of the specifier with and without **CD**. In addition, the method for assessing the inter-rater reliability of the **M-LPE** was to have coders make ratings from videotapes of the original interview, rather than conducting independent interviews. This may have led to inflated reliability estimates, although the structured format used by the interviewers led to little variability in how the questions were asked on the **M-LPE**. Further, the tests of the construct validity of the **M-LPE** was limited to its correlations with externalizing behaviors. As noted by Frick et al. (2014), the construct of **CU traits** has shown theoretically and clinically important correlations with other measures, such as being associated with reduced sensitivity to fear and distress in others, with less sensitivity to punishment cues under certain conditions, and with lower levels of fear, as well as moderating the effectiveness of certain treatments for serious conduct problems. Thus, more comprehensive tests of the **M-LPE's** construct validity are needed to fully evaluate its ability to assess the construct of **CU traits** in ways similar to other measures that have been used in research. Finally, given that the study utilized an adolescent sample that represents mostly non-marital births in large Midwestern cities and was mostly African American and mostly low income, the results may not generalize to other populations.

## Conclusion

Within the context of these limitations, the results of this study provide initial promising psychometric evidence for the use of a semi-structured interview for the assessment of the **LPE** specifier in clinical research. This method provides an alternative to using established behavior rating scales in that it allows for interviewers to ensure that questions are being understood by the informant, but it does not require the same time requirement and level of training as the **CAPE 1.1**, which requires clinician ratings.

Given the recent inclusion of the **LPE** specifier in the major classification systems for childhood and adolescent disorders, there is a great need for multiple methods for assessing the symptoms of the specifier that vary in their time and training requirements in order to meet the needs of various assessment contexts. The **M-LPE** provides a promising approach for use in many clinical research settings.

## Acknowledgments

The research reported in this paper was supported by a grant from the National Institutes of Health R01MH103761 to C.S.M. We would like to acknowledge the past work of the Fragile Families and Child Wellbeing Study, the families for sharing their experiences with us, and the project staff for making the study possible.

## Tables

**Table 1. Zero-order correlations among main study variables.**

|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|---|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| [Content Note: The data within this table was unreadable in the source document and could not be reproduced.] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

\*Note. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . **M-LPE** = CU Traits symptom counts, **ICU** = Inventory of Callous Unemotional Traits, **YSR** = Youth Self Report, **CBCL** = Child Behavior Checklist, **RB** = Rule Breaking subscale, **AGG** = Aggressive Behavior subscale. **ODD** = Oppositional Defiance Disorder, **CD** = Conduct Disorder, **RPQ** = Reactive-Proactive Aggression Questionnaire, **SRD** = Self-Report of Delinquency.

**Table 2. Multiple regressions testing the incremental validity of the M-LPE symptom count.**



|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|---|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
|   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| [Content Note: The data within this table was unreadable in the source document and could not be reproduced.] |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

\*Note. \* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$ . **M-LPE** = CU Traits symptom counts, **ICU** = Inventory of Callous Unemotional Traits, **YSR** = Youth Self Report, **CBCL** = Child Behavior Checklist, **RB** = Rule Breaking subscale, **AGG** = Aggressive Behavior subscale, **ODD** = Oppositional Defiance Disorder, **CD** = Conduct Disorder, **RPQ** = Reactive-Proactive Aggression Questionnaire, **SRD** = Self-Report of Delinquency.

## References

- Achenbach\_1991\_Manual\_for\_the\_Child\_Behavior\_Checklist\_4–18\_and\_1991\_profile
- Achenbach\_1991\_Manual\_for\_the\_Youth\_Self-Report\_and\_1991\_profile
- Ambrosini\_2000\_Historical\_development\_and\_present\_status\_of\_the\_schedule\_for\_affective\_disorders\_and\_schizophrenia\_for\_school-age\_children\_(K-SADS)
- Ansel\_Barry\_Gillen\_&\_Herrington\_2015\_An\_analysis\_of\_four\_self-report\_measures\_of\_adolescent\_callous-unemotional\_traits
- Association\_&American\_Psychiatric\_Association\_2013\_Diagnostic\_and\_statistical\_manual\_of\_mental\_disorders(5th\_ed.)
- Bellina\_Brambilla\_Garzitto\_Negri\_Molteni\_&\_Nobile\_2013\_The\_ability\_of\_CBCL\_DSM-oriented\_scales\_to\_predict\_DSM-IV\_diagnoses\_in\_a\_referred\_sample\_of\_children\_and\_adolescents
- Birmaher\_Khetarpal\_Brend\_Cully\_Balach\_Kaufman\_et\_al\_1997\_The\_Screen\_for\_Child\_Anxiety\_Related\_Emotional\_Disorders\_(SCARED)
- Cardinale\_&\_Marsh\_2017\_The\_reliability\_and\_validity\_of\_the\_Inventory\_of\_Callous\_Unemotional\_Traits\_a\_meta-analytic\_review
- Centifanti\_Shaw\_Atherton\_Thomson\_MacLellan\_&\_Frick\_2019\_CAPE\_for\_measuring\_callous-unemotional\_traits\_in\_disadvantaged\_families
- Charles\_Acheson\_Mathias\_Michael\_Furr\_&\_Dougherty\_2012\_Psychopathic\_traits\_and\_their\_association\_with\_adjustment\_problems\_in\_girls
- Costello\_Edelbrock\_Dulcan\_Kalas\_&Klaric\_1984\_Report\_on\_the\_NIMH\_diagnostic\_interview\_schedule\_for\_children(DISC)
- de\_la\_Peña\_Villavicencio\_Palacio\_Félix\_Larraguibel\_Viola\_...Ulloa\_2018\_Validity\_and\_reliability\_of\_the\_kiddie\_schedule\_for\_affective\_disorders\_and\_schizophrenia\_present\_and\_lifetime\_version\_DSM-5(K-SADS-PL-5)\_Spanish\_version

- Elliott & Huizinga 1984 The relationship between delinquent behavior and ADM problems
- Fanti Frick & Georgiou 2009 Linking callous-unemotional traits to instrumental and non-instrumental forms of aggression
- Feilhauer Cima & Arntz 2012 Assessing callous-unemotional traits across different groups of youths
- Frick 2013 Clinical Assessment of Prosocial Emotions Version 1.1(CAPE 1.1)
- Frick & Ray 2015 Evaluating callous-unemotional traits as a personality construct
- Frick Ray Thornton & Kahn 2014 Can callous-unemotional traits enhance the understanding diagnosis and treatment of serious conduct problems in children and adolescents
- Frick Stickle Dandreaux Farrell & Kimonis 2005 Callous-unemotional traits in predicting the severity and stability of conduct problems and delinquency
- Goetschius Hein Mattson Lopez-Duran Dotterer Welsh ... & Monk 2019 Amygdala-prefrontal cortex white matter tracts are widespread variable and implicated in amygdala modulation in adolescents
- Gomez Vance & Gomez 2014 Analysis of the convergent and discriminant validity of the CBCL TRF and YSR in a clinic-referred sample
- Hare & Neumann 2008 Psychopathy as a clinical and empirical construct
- Hawes Kimonis Mendoza Diaz Frick & Dadds 2020 The Clinical Assessment of Prosocial Emotions(CAPE 11) A multi-informant validation study
- Hawes Price & Dadds 2014 Callous-unemotional traits and the treatment of conduct problems in childhood and adolescence
- Hein Mattson Dotterer Mitchell Lopez-Duran Thomason & Monk 2018 Amygdala habituation and uncinate fasciculus connectivity in adolescence
- Hodges 1993 Structured interviews for assessing children
- Hyde Waller & Burt 2014 Commentary Improving treatment for youth with callous-unemotional traits through the intersection of basic and applied science
- Kaufman Birmaher Brent Rao Flynn Moreci ... & Ryan 1997 Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version(K-SADS-PL)
- Kim Cheon Kim Chang Yoo Kim ... & Noh 2004 The reliability and validity of kiddie-schedule for affective disorders and schizophrenia-present and lifetime version-Korean version(K-SADS-PL-K)

- Kimonis\_Fanti\_Frick\_Moffitt\_Essau\_Bijttebier\_&\_Marsee\_2015\_Using\_self-reported\_callous-unemotional\_traits\_to\_cross-nationally\_assess\_the\_DSM-5\_With\_Limited\_Prosocial\_Emotions\_specifier
- Kimonis\_Frick\_Skeem\_Marsee\_Cruise\_Munoz\_...\_Morris\_2008\_Assessing\_callous-unemotional\_traits\_in\_adolescent\_offenders
- Kotler\_&\_McMahon\_2005\_Child\_psychopathy\_Theories\_measurement\_and\_relations\_with\_the\_development\_and\_persistence\_of\_conduct\_problems
- Lauth\_Magnusson\_Ferrari\_&\_Petursson\_2008\_An\_Icelandic\_version\_of\_the\_Kiddie-SADS-PL
- Lee\_&\_Preacher\_2013\_Calculation\_for\_the\_test\_of\_the\_difference\_between\_two\_dependent\_correlations\_with\_one\_variable\_in\_common
- McHugh\_2018\_Phi\_Correlation\_Coefficient
- Molinuevo\_Martinez-Membrives\_Pera-Guardiola\_Requena\_Torrent\_Bonillo\_Batalla\_Torrubia\_&Frick\_2019\_Psychometric\_properties\_of\_the\_Clinical\_Assessment\_of\_Prosocial\_Emotions\_Version\_1.1(CAPE\_1.1) in\_young\_males\_who\_were\_incarcerated
- Piacentini\_Cohen\_&\_Cohen\_1992\_Combining\_discrepant\_diagnostic\_information\_from\_multiple\_sources
- Raine\_Dodge\_Loeber\_Gatzke-Kopp\_Lynam\_Reynolds\_&\_Liu\_2006\_The\_reactive-proactive\_aggression\_questionnaire
- Reichman\_Teittler\_Garfinkel\_&\_McLanahan\_2001\_Fragile\_families\_Sample\_and\_design
- Tackett\_Krueger\_Sawyer\_&\_Graetz\_2003\_Subfactors\_of\_DSM-IV\_conduct\_disorder
- World\_Health\_Organization\_2018\_International\_statistical\_classification\_of\_diseases\_and\_related\_health\_problems\_(11th\_Revision)