AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Assessment

ISSN: 1040-3590

2023, Vol. 35, No. 12, 1085–1097
https://doi.org/10.1037/pas0001279

Clinical Utility of Diagnosing Limited Prosocial Emotions in Young Children Using the Clinical Assessment of Prosocial Emotions (CAPE)

Bryan Neo[1], Georgette E. Fleming[1], Silvana Kaouar[1], Mei E. Chan[1], Nikki N. Huang[1], David J. Hawes[2, 3], Valsamma Eapen[4, 5, 6], Nancy Briggs[7], and Eva R. Kimonis[1, 2]

[1] Parent-Child Research Clinic, School of Psychology, Faculty of Science, University of New South Wales
[2] Research Department, Karitane, Sydney, Australia
[3] School of Psychology, Faculty of Science, University of Sydney
[4] Discipline of Psychiatry and Mental Health, School of Clinical Medicine, Faculty of Medicine and Health, University of New South Wales
[5] Academic Unit of Infant, Child, and Adolescent Psychiatry Services, South Western Sydney Local Health District, Liverpool, New South Wales, Australia
[6] Ingham Institute for Applied Medical Research, Liverpool, New South Wales, Australia
[7] Mark Wainwright Analytical Centre, Division of Research, University of New South Wales

This study evaluated the interrater reliability, convergent and divergent validity, incremental validity, and clinical prognostic utility of the Clinical Assessment of Prosocial Emotions (CAPE; Frick, 2013) for assessing limited prosocial emotions (LPE). Participants were 232 young children (M_{age} = 3.94 years, SD = 1.46, range = 2–8; 74.6% boys) clinic-referred for conduct problems. We scored the CAPE using binary and dimensional scoring approaches and measured outcomes using parent-report and child laboratory measures. CAPE LPE symptom ratings had good interrater reliability. Children diagnosed with pretreatment LPE had more severe externalizing problems and lower empathy than children without LPE but did not differ in emotion recognition accuracy or anxiety. Dimensional CAPE symptom sum scores were associated with criterion variable scores in expected ways and offered incremental validity beyond scores on the parent-report Inventory of Callous-Unemotional Traits for predicting conduct problem severity, aggression, empathy deficits, and global emotion recognition accuracy. Among children who completed parent management training (n = 44), those diagnosed with LPE ended treatment with more severe aggressive behavior than those without LPE. Overall, children diagnosed with CAPE LPE have severe externalizing problems and achieve reduced benefits from standard parent management training, supporting the need for tailored and intensive interventions to maximize treatment outcomes.

Public Significance Statement
Limited prosocial emotions (LPE) diagnosed using the Clinical Assessment of Prosocial Emotions (CAPE) identifies young clinic-referred children between 2 and 8 years old with more severe externalizing problems that do not normalize with traditional forms of parent management training, supporting the need for tailored and intensive interventions to maximize outcomes.

Keywords: assessment, diagnosis, callous-unemotional traits, limited prosocial emotions, psychopathic personality traits

Supplemental materials: https://doi.org/10.1037/pas0001279.supp

The presence of callous-unemotional (CU) traits identifies a subgroup of antisocial children with unique cognitive and socio-emotional deficits that are thought to explain their severe and aggressive conduct problems (see Frick et al., 2014). Relative to justice-involved youths without elevated CU traits, those with elevated CU traits have an earlier onset to their delinquency, at 7 years old on average (Neo & Kimonis, 2021), thus suggesting a critical need to prioritize the assessment, diagnosis, and treatment of

This article was published Online First September 28, 2023.
Bryan Neo https://orcid.org/0000-0001-6160-0489
Georgette E. Fleming https://orcid.org/0000-0002-9606-066X
Silvana Kaouar https://orcid.org/0000-0003-4837-5103
David J. Hawes https://orcid.org/0000-0003-3260-1225
Valsamma Eapen https://orcid.org/0000-0001-6296-8306
Nancy Briggs https://orcid.org/0000-0002-1134-0807
Eva R. Kimonis https://orcid.org/0000-0002-4325-6108
Eva R. Kimonis received competitive funding from the University of New South Wales for part of this work. Bryan Neo, Georgette E. Fleming, and Silvana Kaouar received Australian Government Research Training Program doctoral scholarships for this work. The authors thank all participating families and clinic staff, especially Sue Morgan from Karitane for assisting with study coordination.

The present study and its hypotheses were not preregistered. Deidentified data and statistical code can be made available for replication purposes from Eva R. Kimonis by request and provided that appropriate institutional agreements are met. This article is based on data from Fleming et al. (2022) and Kimonis et al. (2023).

Bryan Neo played a lead role in conceptualization, data curation, formal analysis, investigation, methodology, project administration, visualization,

continued

children with conduct problems and CU traits before their early elementary school years.

The importance of assessing for CU traits was reflected in the inclusion of the "limited prosocial emotions" (LPE) specifier for children diagnosed with conduct disorder (CD) in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision* (*DSM-5-TR*; American Psychiatric Association, 2022). Although this LPE specifier was only stipulated for CD in *DSM-5-TR*, there is empirical support for using it to specify subgroups of children with conduct problems more generally, including those with oppositional defiant disorder (ODD; Hawes et al., 2020), as is specified in the *International Classification of Diseases and Related Health Problems, Eleventh Edition* (World Health Organization, 2019). The *DSM-5-TR* LPE specifier is operationalized by four symptoms: lack of remorse/guilt, callous-lack of empathy, unconcerned about performance in important activities, and shallow/deficient affect. These four symptoms were selected based on factor analytic research on two widely used CU questionnaire measures, the Antisocial Process Screening Device (APSD; Frick & Hare, 2001) and the Inventory of Callous-Unemotional Traits (ICU; Frick, 2004), which revealed four items that consistently loaded onto a single CU factor that was distinct from a CD factor (Frick & Moffitt, 2010). *DSM-5-TR* diagnostic criteria for LPE are met when two or more of the four LPE symptoms are present across relationships and settings for at least 12 months. This symptom count threshold was selected because children with conduct problems and at least two LPE symptoms had the poorest delinquency trajectory over a 4-year follow-up period (Frick & Moffitt, 2010).

Utility of the *DSM-5-TR* LPE Diagnosis in Young Children

Several studies found that the *DSM-5-TR* LPE diagnosis identifies a subgroup of children with more severe externalizing problems. Relative to boys and girls without an LPE diagnosis, children diagnosed with LPE had more severe externalizing problems (Castagna et al., 2021; Colins et al., 2021; Pardini et al., 2012), aggressive behavior (Bansal et al., 2023; Kahn et al., 2012; Pardini et al., 2012), cruel behavior to others and animals (Kahn et al., 2012), and risk of hurting others using a weapon (Déry et al., 2019). Children diagnosed with LPE also had greater global impairment and required more intensive treatment for their conduct problems (Castagna et al., 2021; Pardini et al., 2012).

However, there are inconsistencies in the literature, with some studies finding nonsignificant differences in externalizing problems between children with versus without LPE (Bansal et al., 2023; Déry et al., 2019; Kahn et al., 2012). Further, pretreatment LPE diagnoses failed to predict externalizing treatment outcomes (Kolko & Pardini, 2010), contrary to prior research using continuous CU scores (Hawes et al., 2014). One reason for these mixed findings could be due to methodological differences in identifying the presence of LPE. Specifically, prior research relied on CU questionnaire item scores that each corresponds to one of the four *DSM-5-TR* LPE symptoms to indicate the presence of at least two LPE symptoms, according to extreme or split coding methods (see Kimonis et al., 2015).

Although using CU item scores is a quick and cost-efficient way to screen for the presence of LPE symptoms, this classification method has not reliably identified children with more severe externalizing behavior (Frick et al., 2014) and poorer treatment responsivity (Hawes et al., 2014). Additionally, the reliance on a single questionnaire item score is not aligned with evidence-based assessment practices that advocate for using a multimethod, multi-informant assessment approach (Frick et al., 2020). Consequently, the most common method used in research to operationalize the *DSM-5-TR* LPE specifier is not suitable on its own for clinical use and greatly limits the generalizability of prior study results.

Using a Structured Professional Judgment Method to Diagnose LPE

The only clinical tool currently available for diagnosing *DSM-5-TR* LPE is the Clinical Assessment of Prosocial Emotions (CAPE; Frick, 2013). The CAPE is a structured professional judgment (SPJ) tool under development to diagnose *DSM-5-TR* LPE in children between ages 3 and 21 years. The SPJ approach requires assessors to consider multiple sources of information to rate the presence of each LPE symptom against evidence-based symptom descriptors: 0 (*not descriptive or mildly descriptive*), 1 (*moderately descriptive*), and 2 (*highly descriptive*). A rating of "2" indicates that the LPE symptom is present and contributes to the symptom count required for diagnosis. To determine the presence of LPE symptoms, assessors must evaluate the credibility of the source, as well as the persistence and pervasiveness of LPE symptoms, consistent with *DSM-5-TR* criteria.

Two prior published studies have evaluated the validity of CAPE LPE diagnoses (Centifanti et al., 2020; Hawes et al., 2020) among children below 8 years old, which is the critical age period for early intervention (Neo & Kimonis, 2021). The one other study of the CAPE was with 72 incarcerated adolescent boys (M_{age} = 17.74 years, SD = 1.20, range = 14–22; Molinuevo et al., 2020). Centifanti et al. (2020) evaluated the validity of CAPE LPE diagnoses in a sample of 34 at-risk youths (M_{age} = 13.5 years, SD = 4.7, range = 5–18; 69% boys) in the United Kingdom. Results showed that children diagnosed with LPE scored significantly higher on parent-rated CU traits than children without an LPE diagnosis but did not differ in conduct problem severity, violence risk, or internalizing problems. Given robust findings of greater conduct problem severity among children high on CU traits (Frick et al., 2014), Centifanti et al. (2020) attributed these unexpected results to small sample sizes (n = 7 diagnosed with LPE), which limited statistical power to detect a

writing–original draft, and writing–review and editing. Georgette E. Fleming played a supporting role in formal analysis and an equal role in data curation, investigation, project administration, and writing–review and editing. Silvana Kaouar played a supporting role in project administration and an equal role in data curation, investigation, and writing–review and editing. Mei E. Chan played a supporting role in investigation. Nikki N. Huang played a supporting role in investigation and an equal role in writing–review and editing. David J. Hawes played an equal role in resources and writing–review and editing. Valsamma Eapen played an equal role in resources and writing–review and editing. Nancy Briggs played a supporting role in formal analysis. Eva R. Kimonis played a lead role in funding acquisition, resources, software, and supervision and an equal role in conceptualization, methodology, and writing–review and editing.

Correspondence concerning this article should be addressed to Eva R. Kimonis, Parent-Child Research Clinic, Faculty of Science, School of Psychology, University of New South Wales, Sydney, NSW 2052, Australia. Email: e.kimonis@unsw.edu.au

significant effect. Critically, this study did not assess for ODD/CD diagnoses and cannot directly inform the utility of the *DSM-5-TR* and *International Classification of Diseases and Related Health Problems, Eleventh Edition* LPE diagnoses.

To date, only one study has evaluated the validity of CAPE LPE diagnoses in young children diagnosed with ODD and/or CD. Using a larger sample ($N = 82$), Hawes et al. (2020) tested whether clinic-referred children ($M_{age} = 7.40$ years, $SD = 2.69$, range = 3–15; 75% boys; 66% diagnosed with ODD/CD) in Australia diagnosed with LPE differed from children without an LPE diagnosis on parent- and teacher-rated CU traits. Results showed that children diagnosed with LPE had higher CU scores than children without an LPE diagnosis. Although children diagnosed with ODD/CD and LPE had lower parent-rated affective empathy than children with ODD/CD only, these groups did not differ in conduct problem severity. Hawes et al. (2020) hypothesized that these nonsignificant findings could be due to the binary operationalization of LPE on the CAPE, as is adopted by *DSM-5-TR*, thus suggesting that further research using dimensional CAPE LPE scores is warranted.

Together, prior research demonstrated that CAPE LPE diagnoses have concurrent validity with informant-reported CU and empathy scores but do not identify a subgroup of antisocial children with more severe conduct problems (Centifanti et al., 2020), even when statistical tests were adequately powered (Hawes et al., 2020). These nonsignificant findings are inconsistent with prior research showing that antisocial children diagnosed with LPE using questionnaire items have more severe conduct problems than those without an LPE diagnosis (Colins et al., 2021; Frick & Moffitt, 2010). One reason for this inconsistency could be due to limitations in the measures of conduct problem severity used in prior CAPE studies. For instance, the majority of the 12-item Antisocial measure used by Hawes et al. (2020) tapped into psychopathy-linked narcissism rather than conduct problems. Also, the Antisocial scale's 3-point response option may have limited the range of scores to adequately differentiate conduct problem severity among children with ODD/CD symptoms that exceed diagnostic thresholds. The present study addresses these limitations by using a measure of conduct problem severity with high face validity and a wider range of possible scores.

The present study extends prior CAPE studies in four ways. First, we tested the interrater reliability of CAPE LPE symptom ratings ($n = 36$; Research Question 1). The only other study to test the interrater reliability of CAPE LPE symptom ratings was conducted with incarcerated adolescent boys in Spain using Catalan and Spanish versions of the CAPE (Molinuevo et al., 2020). Results showed that the callous-lack of empathy (standard Cohen's $\kappa = .51$) and unconcerned about performance in important activities (.43) symptoms had moderate agreement among raters, whereas the lack of remorse/guilt (.30) and shallow/deficient affect (.40) symptoms had fair agreement. The authors attributed these modest findings to the secondary coder relying on audio recordings of CAPE clinical interviews to score most cases. Indeed, it is insufficient to rely only on verbal information to assess the credibility of information (Zuckerman et al., 1982). To address this study limitation, the secondary coder in the present study used audio and video recordings of CAPE clinical interviews, in addition to other sources of information, to score the CAPE.

Second, we extended prior studies evaluating the convergent and divergent validity of CAPE LPE diagnoses, which relied on parent- and teacher-report measures only, by also including child laboratory measures of facial emotion recognition accuracy and further evaluating the construct validity of dimensional CAPE scores ($N = 232$; Research Question 2). The present study is the first, to our knowledge, to evaluate whether children with or without an LPE diagnosis differ in their emotion recognition, consistent with prior research finding that continuous ICU scores were negatively associated with facial emotion recognition accuracy in preschool children (Kimonis et al., 2016). We hypothesized that children diagnosed with pretreatment LPE would have higher CU traits and externalizing problems, and lower empathy and facial emotion recognition accuracy, relative to children without LPE, and that these group differences would be greater than any group differences on internalizing problems. Third, we examined the incremental validity of CAPE LPE binary and dimensional scores, entered as predictors in separate hierarchical regression models, for predicting conduct problem severity, aggression, empathy deficits, and facial emotion recognition accuracy, beyond continuous ICU scores ($N = 232$; Research Question 3).

Fourth, we tested the clinical utility of pretreatment CAPE LPE diagnoses for predicting externalizing treatment outcomes among a subsample of children who received an efficacious parent management training intervention called parent–child interaction therapy (PCIT; Ward et al., 2016; $n = 44$; Research Question 4). This objective is important because the *DSM-5-TR* states that a mental disorder should be defined by its clinical utility for determining prognosis and treatment planning; however, no study has evaluated the clinical prognostic utility of the *DSM-5-TR* LPE specifier using a comprehensive diagnostic index of LPE. We hypothesized that children diagnosed with LPE would have lesser reductions in aggressive behavior across treatment than children without an LPE diagnosis.

Method

Participants

Participants were 232 children ($M_{age} = 3.94$ years, $SD = 1.46$, range = 2–8; 74.6% boys) referred to a university-based clinic ($n = 121$; $M_{age} = 4.71$ years, $SD = 1.50$, range = 2–8; 76.9% boys) or a community-based clinic ($n = 111$, $M_{age} = 3.03$ years, $SD = 0.66$, range = 2–5; 72.1% boys; child age and sex was not available for 10 families) for conduct problems in Sydney, Australia. Children from the university-based clinic were significantly older than children from the community-based clinic, $t(171.42) = 11.09$, $p < .001$, $d = 1.41$. This age difference between clinics was expected as the university-based clinic offered clinical services only to children between 2 and 8 years, whereas the community-based clinic offered clinical services only to children between 2 and 5 years. The proportion of boys and girls did not significantly differ between clinics, $\chi^2(1, N = 222) = 0.18$, $p = .67$. Of the 232 families, both caregivers completed assessment measures for 103 children ($n = 20$ from the community-based clinic, $n = 83$ from the university-based clinic). The remaining 129 families had one caregiver complete assessment measures ($n = 91$ from the community-based clinic, $n = 38$ from the university-based clinic; mothers $n = 126$).

At the university-based clinic, the median annual household income was $150,000 (ranging AUD$20,000–$1,000,000; not reported by 16 families) and 59.5% of children were White ($n = 72$), 24.8% were biracial ($n = 30$), 9.1% were Asian ($n = 11$), 2.5% were Middle Eastern ($n = 3$), and 2.5% ($n = 3$) identified as "other"; two families did not report their race and/or ethnicity. Income and race and/or ethnicity data

were not routinely collected from the community-based clinic and were therefore unavailable. Based on the latest census data from the Australian Bureau of Statistics (2021), the median annual household income in the local government area of the community-based clinic was AUD$72,280. The Australian Bureau of Statistics did not report race and/or ethnicity data. Since there could be potential differences in socioeconomic levels between clinics, we controlled for study site in our analyses.

Measures

Supplemental Materials provide detailed information about each measure. Table 1 reports internal consistency estimates for all study variables. The benchmarks set for acceptable internal consistency were Cronbach's $\alpha > .70$, McDonald's $\omega > .70$, and mean interitem correlation (MIC) > .15 (Clark & Watson, 1995; Nunnally, 1978). We combined questionnaire scores between two caregivers using item-level resolved scoring (i.e., selecting the higher item score between raters to compute scale scores) to circumvent potential underreporting of symptoms (e.g., Kimonis et al., 2016; Pardini et al., 2012). Semistructured interview measures were conducted with both caregivers together, where available, and scored using consensus ratings between caregivers.

LPE/CU Traits

The CAPE (Frick, 2013) is an SPJ tool to diagnose *DSM-5-TR* LPE. The CAPE manual includes a semistructured interview consisting of nine yes-or-no, skip logic questions that capture the four *DSM-5-TR* LPE symptoms. Parent-report interview responses were used as one of multiple information sources to score the CAPE, consistent with administration guidelines in the manual. CAPE symptom sum scores were computed as the sum of CAPE LPE ratings, ranging from 0 to 8. Internal consistency of dimensional CAPE LPE symptom sum scores was acceptable. Using the categorical scoring approach, 50 children (21.6%) were diagnosed with LPE.

The 24-item preschool ICU (Kimonis et al., 2016) is a parent-report measure of CU traits, operationalized in the *DSM-5-TR* as LPE. Items are rated on a 4-point scale, 0 (*not true at all*), 1 (*somewhat true*), 2 (*very true*), and 3 (*definitely true*). ICU scores had good internal consistency. Seven children were missing ICU data.

The five-item CU scale from the Achenbach System of Empirically Based Assessment–Child Behavior Checklist (CBCL CU; Willoughby et al., 2011) is a parent-report measure of CU traits developed using a well-established broadband measure of child psychopathology for children between 1.5 and 5 years old (CBCL preschool version; Achenbach & Rescorla, 2000). Items are rated on a 3-point scale, 0 (*not true*), 1 (*somewhat or sometimes true*), and 2 (*very true or often true*). Although the Cronbach's α metric of the CBCL CU scale in the present study was below the benchmark set for adequate reliability, it was consistent with prior research using this scale (i.e., Cronbach's α ranging .55–.65; Waller et al., 2015; Willoughby et al., 2011, 2014) and the MIC in the present study was acceptable. Twelve children were missing CBCL preschool version data, and one child had incomplete data for scoring this scale. Thirty participants completed the CBCL school-age version because they were over 5 years old. The CBCL school-age version does not include items to compute CBCL CU scores.

Externalizing and Internalizing Problems

The Diagnostic Interview Schedule for Children, Adolescents, and Parents (DISCAP; Johnson et al., 1999) is a semistructured interview to diagnose ODD and CD. Internal consistency for the DISCAP ODD and CD Criteria A symptoms were acceptable (Cronbach's $\alpha = .69–.73$, McDonald's $\omega = .70–.73$, MIC = .25–.36). The DISCAP CD module was not routinely administered at the community-based clinic because of the children's younger age relative to those in the university-based clinic. Of the remaining children who completed DISCAP ODD ($n = 231$) and/or CD ($n = 132$) modules, 113 were diagnosed with ODD only, two were diagnosed with CD only, and 38 were diagnosed with both ODD and CD (i.e., 66.2% of the sample were diagnosed with ODD and/or CD).

The 36-item Eyberg Child Behavior Inventory (ECBI; Eyberg & Pincus, 1999) is a parent-report measure of child conduct problem severity. Items are rated on a 7-point scale ranging from 1 (*never*) to 7 (*always*). ECBI Intensity scores had excellent internal consistency. Ten children were missing ECBI data.

The 19-/18-item Aggressive Behavior and 10-/six-item Anxiety scales of the parent-report CBCL (preschool/school-age versions; Achenbach & Rescorla, 2000, 2001) were used in the present study. Scores between CBCL versions were combined by using age- and sex-based *T* scores. All scales had good internal consistency. The CBCL Aggressive Behavior scale was used over the six-/five-item CBCL Oppositional Defiance scale because aggressive behavior is most relevant to CU traits. Additionally, there is item overlap between the two scales (e.g., "Temper tantrums or hot temper" is included in both scales) and CBCL Oppositional Defiance scores in the present study had poor internal consistency (i.e., Cronbach's $\alpha = .31$ and MIC = .14). Results testing CBCL Oppositional Defiance scores as outcomes are reported in Supplemental Materials. Twelve children were missing CBCL preschool version data. No data were missing for the CBCL school-age version.

The 16-item Brief Proactive Reactive Aggression (BPRA; Brown et al., 1996) is a parent-report measure of proactive (10 items) and reactive (six items) aggression. Items are rated on a 3-point scale, 0 (*never*), 1 (*sometimes*), and 2 (*very often*). Both scales had good internal consistency. BPRA was only administered to community-based clinic participants, and data were missing for 14 children.

Empathy

The nine-item Affective Empathy subscale of the Griffith Empathy Measure (GEM; Dadds et al., 2008) is a parent-report measure of affective empathy. Items are rated on a 9-point scale, −4 (*strongly disagree*) to +4 (*strongly agree*). The GEM Affective Empathy subscale had good internal consistency. These data were missing for 13 children.

Laboratory Emotion Recognition Tasks

Dynamic Adult Facial Stimuli

A computerized facial emotion recognition task was used to measure global and fear-specific emotion recognition accuracy using E-Prime (Version 2). This task includes stimuli from the Pictures of Facial Affect Series (Ekman & Friesen, 1976) and was used previously with preschool children (Kimonis et al., 2016). Children were presented with 40 one-second dynamic visual stimuli depicting

Table 1

Zero-Order Correlations (Pearson's r) Between Study Variables

Study variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. Child's age	—																
2. Child's sex (1 = boy, 2 = girl)	.02	—															
3. Study site (1 = university, 2 = community)	−.58*	−.03	—														
4. ICU total	.25*	−.12	−.30*	—													
5. CBCL CU	.03	−.13	−.09	.59*	—												
6. ECBI Intensity	.33*	−.11	−.34*	.53*	.52*	—											
7. CBCL Oppositional Defiance	.23*	−.06	−.35*	.35*	.45*	.58*	—										
8. CBCL Aggression	.14*	−.15*	−.26*	.49*	.66*	.69*	.81*	—									
9. BPRA Proactive	.25* (.13)	−.03 (−.001)	—	.28* (.18)	.53* (.27*)	.60* (.37*)	.47* (.02)	.66* (.29*)	—								
10. BPRA Reactive	.23* (.07)	−.04 (−.03)	—	.22* (.02)	.51* (.21*)	.53* (.17)	.63* (.48*)	.72* (.45*)	.73*	—							
11. GEM Affective	.02	.01	−.17*	−.38*	−.35*	−.17*	−.05	−.17*	−.17 (−.07)	−.16 (−.06)	—						
12. CBCL Anxiety	−.002	.02	−.04	.14*	.26*	.20*	.29*	.35*	.15 (−.12)	.30* (.29*)	.16*	—					
13. Fear accuracy (static child faces)	.36* [.37*]	.10 [.11]	−.32* [−.32*]	.002 [−.03]	−.20* [−.27*]	.05	−.01 [−.05]	−.13 [−.22*]	−.11 (−.02) [−.18]	−.14 (−.09) [−.20]	.12 [.13]	.08 [.07]	—				
14. Global accuracy (static child faces)	.53* [.51*]	.06 [.08]	−.52* [−.51*]	.08 [.002]	−.15 [−.27*]	.14	.14 [.07]	−.07 [−.23*]	.05 (−.02) [−.05]	.09 (.07) [.01]	.12 [.15]	.04 [.02]	.61* [.61*]	—			
15. Fear accuracy (dynamic adult faces)	.62* [.59*]	.12 [.15]	−.42* [−.37*]	.09 [−.04]	−.21* [−.40*]	.23*	.17 [.04]	.06 [−.14]	.15 (−.20) [.01]	.38* (.40*) [.31]	.08 [.13]	.02 [−.03]	.39* [.39*]	.56* [.54*]	—		
16. Global accuracy (dynamic adult faces)	.62* [.60*]	−.01 [.02]	−.42* [−.38*]	.12 [.02]	−.11 [−.25*]	.20*	.18* [.08]	.01 [−.17]	.20 (.06) [.10]	.21 (.10) [.13]	.03 [.07]	.05 [.02]	.35* [.34*]	.68* [.67*]	.76* [.75*]	—	
17. CAPE LPE symptom sum (ranging from 0 to 8)	.23*	−.003	−.30*	.54*	.43*	.42*	.30*	.41*	.38* (.26*)	.28* (.01)	−.42*	.04	.05 [.02]	−.001 [−.07]	.16 [.06]	.22* [.15]	—
M	3.94	1.22	1.48	32.24	4.08	165.23	69.56	72.53	4.52	5.51	7.44	63.23	.35	.54	.35	.51	2.52
SD	1.46	0.42	0.50	10.48	2.16	31.98	8.53	11.95	4.18	2.89	11.95	11.87	.34	.22	.27	.21	2.23
Cronbach's α	—	—	—	.88	.63	.92	.83 {.31}	.90 {.83}	.79	.86	.88	.82 {.80}	—	—	—	—	.73
McDonald's ω	—	—	—	.87	.60	.91	.82	.90 {.81}	.78	.85	.88	.82 {.79}	—	—	—	—	.75
Mean interitem correlation	—	—	—	.24	.26	.24	.45 {.14}	.33 {.22}	.39	.39	.45	.31 {.40}	—	—	—	—	.40

Note. ICU = Inventory of Callous-Unemotional Traits; CBCL = Child Behavior Checklist; CU = callous-unemotional; ECBI = Eyberg Child Behavior Inventory; BPRA = Brief Proactive Reactive Aggression Scale; GEM = Griffith Empathy Measure; CAPE = Clinical Assessment of Prosocial Emotions; LPE = Limited Prosocial Emotions. Correlations in parentheses ( ) are partial correlations between scores on BPRA Proactive or BPRA Reactive and other study variables after controlling either BPRA aggression type. Correlations in square brackets [ ] are partial correlations controlling for conduct problem severity, measured using ECBI Intensity scores. We reported internal consistency estimates separately for the preschool and school-age versions of the Child Behavior Checklist (CBCL) with the latter presented in { }.

* $p < .05$.

an adult White man or woman's face (20 of each sex) morphing from neutral to one of five facial emotional expressions (happy, sad, angry, fear, pain; eight clips per emotion) in random, sequential order. After each face presentation, children were instructed to identify aloud one of the five emotions. An experimenter entered the child's response on a keyboard using designated keys. Global and fear-specific emotion recognition accuracy scores were calculated as the proportion of correctly identified emotional expressions. There were 76 children from the community-based clinic and 27 from the university-based clinic who did not complete this task due to time constraints during the assessment process, technical difficulties, or the child's inability or unwillingness to complete the task. Children who completed this task were significantly older and had higher CBCL CU scores and externalizing problem scores than children who did not complete the task. Groups did not differ in CBCL CU scores after controlling for study site. Test statistics for score differences between task completers and non-completers are reported in Supplemental Table 1. We covaried child age, study site, and conduct problem severity in our analysis.

Static Child Facial Emotion Stimuli

To account for own-age bias in recognizing facial emotion (i.e., better at recognizing facial emotions closer to one's own age; Riediger et al., 2011), we administered a facial emotion recognition task using child models from the empirically validated Radboud Faces Database (Langner et al., 2010). Children were presented with 24 static facial images of White boys and girls (12 of each sex) displaying six prototypical emotional expressions (sadness, surprise, anger, disgust, happiness, and fear; four per emotion) on a printed booklet in fixed, sequential order. Children were instructed to identify aloud which one of the six emotions were displayed, which the experimenter recorded in the response booklet. Global and fear-specific emotion recognition accuracy scores were calculated as the proportion of correctly identified emotional expressions. These data were not available for six children from the university-based clinic and 71 from the community-based clinic for the same reasons detailed above. Children who completed this task were significantly older and had higher ICU and externalizing problem scores than children who did not complete the task. Groups did not differ in ICU scores after controlling for study site. Test statistics for score differences between task completers and noncompleters are reported in Supplemental Table 2. We covaried child age, study site, and conduct problem severity in our analysis.

Procedure

Detailed descriptions of the study procedure are reported in Supplemental Materials. Research ethical approval was obtained from university and local health department research ethics committees. Children were eligible to participate if they were referred for conduct problems, regardless of meeting diagnostic criteria for conduct disorders. At the university-based clinic, children were excluded from participation if they had a primary mental health diagnosis other than ODD and/or CD (e.g., autism spectrum disorder, intellectual disability) or were deaf. We did not exclude children at the community-based clinic on these criteria because data collection was part of a broader study that examined more heterogeneous populations of children with conduct problems. We statistically controlled for study site in our analyses to account for potential sampling differences. All children received a toy valued at <AUD$5 for their participation. Eligible families completed a pretreatment assessment, including parent-report questionnaires, a general semistructured clinical interview, DISCAP and CAPE semistructured interviews, and child laboratory tasks. To alleviate demand characteristics that could limit the generalizability of our results, all families were offered treatment for conduct problems regardless of their child's LPE status on the CAPE at the pretreatment assessment. Assessors were registered psychologists or master's-/doctoral-level trainees in clinical or forensic psychology and were trained in advanced clinical interviewing skills. We used pretreatment data to evaluate the interrater reliability and construct validity of CAPE LPE diagnoses. Of the 121 families from the university-based clinic, 44 received standard PCIT treatment that was delivered in clinic (n = 31) or over the internet (n = 13), in a fixed session format (14 sessions; see Thomas & Zimmer-Gembeck, 2012). Therapists were registered psychologists (n = 4, 75% female) trained by certified PCIT trainers who are the second and last authors on this article. Families completed six assessments (pretreatment [100% of families], midtreatment [82%], posttreatment [77%], 1.5-month follow-up [MFU; 71%], 3-month follow-up [74%], and 4.5-month follow-up [90%]). Only pretreatment data were collected at the community-based clinic (n = 111).

The present study and its hypotheses were not preregistered. Deidentified data and statistical code can be made available for replication purposes from the senior author (last author of this article) by request and provided appropriate institutional agreements are met.

Statistical Analyses

Planned Analyses

To test the interrater reliability of each CAPE LPE symptom (Research Question 1), CAPE interviews for a randomly selected 36 participants (15%) were secondary coded. The secondary coder was a postbaccalaureate researcher with a psychology degree who was trained by one of the original assessors and was required to complete two criterion videos with the original assessor before coding CAPEs. The secondary coder was provided with identical information to the original assessor for scoring the CAPE (i.e., audio and video recordings of pretreatment clinical assessment, DISCAP and CAPE interviews, and scores on the ICU and CBCL CU) and was masked to the LPE status of all participants. We used quadratically weighted Cohen's κ to estimate interrater reliability. In contrast to standard Cohen's κ, the weighted version accounts for the level of disagreement between raters and applies a larger emphasis on bigger disagreements in LPE ratings (i.e., rating differences between "2" and "0" are weighted more heavily than differences between "2" and "1"; Cohen, 1968). We applied quadratic weighting because the difference between ratings "0" and "1" is less significant than the difference between ratings "1" and "2" (i.e., a rating of "2" indicates the presence of an LPE symptom and contributes to the symptom count to diagnose LPE). Standard Cohen's κ values were reported in the present study for the purposes of comparison with prior research. We interpreted Cohen's κ estimates as follows: 0.80–1.00 (very good), 0.60–0.80 (good),

0.40–0.60 (moderate), 0.20–0.40 (fair), and <0.20 (poor; Altman, 1991). We used McNemar's test to examine whether the proportion of children diagnosed with LPE differed between raters. This repeated measures χ^2 test accounts for sample nonindependence and is suitable for paired data (i.e., both raters scored the same CAPE cases; Tabachnick & Fidell, 2013).

To test whether children with or without a CAPE LPE diagnosis differed on levels of conduct problems, empathy, anxiety, and facial emotion recognition accuracy (Research Question 2), we conducted analysis of covariance (ANCOVA) in IBM SPSS Statistics (Version 27) using pretreatment data. All ANCOVA models controlled for child age and sex, study site, conduct problem severity (ECBI Intensity scores), and anxiety levels (CBCL Anxiety scores). ANCOVA analyses were repeated with the subset of children who met diagnostic criteria for ODD/CD. Effect sizes were estimated using η_p^2 (small = .01, medium = .06, and large = .14; Cohen, 1988). We computed zero-order correlations between CAPE symptom sum scores and all study variables, as well as partial correlations, controlling for conduct problem severity, when examining facial emotion recognition task accuracy scores.

To test whether CAPE binary and symptom sum scores incrementally predicted conduct problem severity, aggression, empathy deficits, and facial emotion recognition accuracy beyond ICU scores (Research Question 3), we conducted hierarchical regression analyses in IBM SPSS Statistics (Version 27) using pretreatment data. All regression models controlled for the same covariates used in Research Question 2.

To test whether children with or without a CAPE LPE diagnosis differed on aggression treatment outcomes (Research Question 4), we conducted mixed model analyses in IBM SPSS Statistics (Version 27) using the MIXED procedure to model nonlinearity. This statistical procedure does not require participants to have the same number of observations at each time point (i.e., no listwise deletion of cases with missing data when conducting the analysis). We used maximum likelihood estimation and the SPSS default covariance structure (i.e., variance components). We also specified random intercept and slope in the model. We controlled for child age and sex, pretreatment conduct problem severity (ECBI Intensity scores) and anxiety (CBCL Anxiety scores), treatment delivery method (internet delivered vs. in clinic), and therapist. We explored the shape of change over time by systematically increasing the power polynomial and comparing model fit statistics between models. We selected the model with the lowest Akaike's information criterion and Bayesian information criterion values. We repeated analyses for the subset of children diagnosed with ODD and/or CD. Effect sizes were estimated using η_p^2 (small = .01, medium = .06, and large = .14; Cohen, 1988).

Covariate Analyses

Table 1 reports correlations between study variables. Detailed descriptions of covariates analyses are reported in Supplemental Materials. ANCOVA and mixed models controlled for child age and sex, study site, conduct problem severity (ECBI Intensity scores), and anxiety (CBCL Anxiety scores) to ensure that significant differences were not due to demographic factors, sampling differences, or pretreatment conduct problem severity or anxiety levels. We did not control for conduct problem severity or anxiety when testing whether LPE groups differ in externalizing problems or anxiety, respectively.

Results

Research Question 1: Are CAPE LPE Symptom Ratings Reliable Across Assessors?

Interrater reliability statistics for CAPE symptom ratings are as follows: lack of remorse/guilt (quadratically weighted Cohen's $\kappa = .80$/standard Cohen's $\kappa = .56$), callous-lack of empathy (.86/.73), unconcerned about performance in important activities (.87/.76), and shallow/deficient affect (.77/.57). Quadratically weighted Cohen's κ values were higher than standard Cohen's κ values across CAPE LPE symptoms and ranged between "very good" and "good," according to Altman's (1991) benchmarks. The proportion of cases diagnosed with LPE on the CAPE did not significantly differ across assessors ($p = .22$).

Research Question 2: Evaluating the Convergent and Divergent Validity of CAPE LPE Diagnoses and CAPE Symptom Sum Scores

Table 2 presents ANCOVA statistics comparing children diagnosed with versus without CAPE LPE on outcomes. Relative to children without an LPE diagnosis, children diagnosed with LPE had significantly higher ICU and CBCL CU scores, controlling for covariates. Children diagnosed with LPE had higher ECBI Intensity, CBCL Aggression, and BPRA Proactive Aggression scores, and lower GEM Affective scores than children without an LPE diagnosis, controlling for covariates. Children with or without an LPE diagnosis did not significantly differ in BPRA Reactive Aggression or CBCL Anxiety scores, controlling for covariates. All significant differences had small to medium effect sizes, except for ICU scores which had large effect sizes. This result is consistent with prior CAPE research finding small to medium effect size differences on conduct problem severity (although not statistically significant; Centifanti et al., 2020) and statistically significant large effect size differences on CU measures (Centifanti et al., 2020; Hawes et al., 2020). Children diagnosed with or without LPE did not differ significantly in fear and global facial emotion recognition of static child or dynamic adult expressions, controlling for covariates. The statistical significance of results did not change when we repeated ANCOVA models using a subset of children diagnosed with ODD and/or CD, except for BPRA Proactive Aggression, $F(1, 47) = 1.41$, $p = .24$, $\eta_p^2 = .03$, 95% confidence interval (CI) [<.01, .14], and CBCL CU, $F(1, 112) = 3.28$, $p = .07$, $\eta_p^2 = .03$, 95% CI [<.01, .10] scores, which were no longer significantly different between ODD- and/or CD-diagnosed children with and without an LPE diagnosis on the CAPE.

Table 1 presents correlations between CAPE symptom sum scores and outcome scores. CAPE symptom sum scores were significantly and positively correlated with ICU and CBCL CU scores, with greater effect sizes for ICU scores. CAPE symptom sum scores were significantly and positively correlated with ECBI Intensity, CBCL Oppositional Defiance and Aggression, and BPRA Proactive Aggression scores, and negatively correlated with GEM Affective Empathy scores. CAPE symptom sum scores were nonsignificantly correlated with BPRA Reactive Aggression scores after controlling for BPRA Proactive Aggression scores; however, CAPE symptom sum scores remained positively correlated with BPRA Proactive

Table 2

Estimated Marginal Means, Standard Errors, and ANCOVA Statistics Comparing Outcomes Between Clinic-Referred Children With and Without Limited Prosocial Emotions (LPE) on the Clinical Assessment of Prosocial Emotions (CAPE)

		LPE diagnosis (≥2 symptoms rated "2") $n = 50$			No LPE diagnosis (<2 symptoms rated "2") $n = 182$					
Multimethod outcome measures	N	n	M	SE	n	M	SE	F	p	η_p^2 [95% CI]
Parent-Report Questionnaires										
ICU total	213	49	37.16	1.29	164	31.21	0.68	$F(1, 206) = 15.85$	<.001	.07 [.02, .13]
CBCL CU preschool version	182	38	4.83	0.31	144	3.90	0.15	$F(1, 175) = 7.14$	.008	.04 [.01, .10]
ECBI Intensity	213	49	178.95	4.15	164	161.51	2.22	$F(1, 207) = 13.34$	<.001	.06 [.02, .12]
CBCL Aggression	220	49	77.40	1.55	171	71.13	0.81	$F(1, 214) = 12.55$	<.001	.06 [.02, .11]
BPRA Proactive	97	9	7.30	1.36	88	4.23	0.43	$F(1, 92) = 4.61$	.04	.05 [<.01, .13]
BPRA Reactive	97	9	6.79	0.90	88	5.31	0.28	$F(1, 92) = 2.49$	.12	.03 [<.01, .10]
GEM Affective	212	49	−0.23	1.60	163	9.96	0.84	$F(1, 205) = 30.23$	<.001	.13 [.06, .20]
CBCL Anxiety	213	49	62.46	1.79	164	63.63	0.94	$F(1, 207) = 0.32$	.57	<.01 [<.01, .02]
Emotion recognition task using static child faces										
Fear accuracy	150	42	.39	.05	108	.36	.03	$F(1, 143) = 0.29$	.59	<.01 [<.01, .03]
Global accuracy	150	42	.53	.03	108	.56	.02	$F(1, 143) = 0.80$	.37	<.01 [<.01, .04]
Emotion recognition task using dynamic adult faces										
Fear accuracy	126	40	.36	.03	86	.35	.02	$F(1, 119) = 0.03$	.86	<.01 [<.01, .01]
Global accuracy	126	40	.54	.03	86	.51	.02	$F(1, 119) = 0.68$	.41	<.01 [<.01, .05]

Note. All ANCOVA models statistically controlled for child age and sex, study site, and CBCL Anxiety scores. We additionally controlled for conduct problem severity using ECBI Intensity scores in ANCOVA models that did not examine conduct problems as outcomes (i.e., all outcomes, except CBCL Aggression, and BPRA Proactive and Reactive). We did not control for CBCL Anxiety scores when testing whether LPE groups differ on CBCL Anxiety scores. We did not control for the study site when examining BPRA scores as outcomes because the BPRA was only administered at the community-based clinic. There was no change in statistical significance levels when we repeated the models with the subset of 153 clinic-referred children who met diagnostic criteria for oppositional defiant disorder (ODD) and/or conduct disorder (CD) on the Diagnostic Interview Schedule for Children, Adolescents, and Parents (DISCAP), except for BPRA Proactive Aggression scores which did not significantly differ between ODD- and/or CD-diagnosed children with and without LPE on the CAPE ($p = .24$), and CBCL CU scores (marginal significance at $p = .07$). Of the 153 clinic-referred children diagnosed with ODD and/or CD, 45 children were diagnosed with LPE on the CAPE. ANCOVA = analysis of covariance; CI = confidence interval; *SE* = standard error; ICU = Inventory of Callous-Unemotional Traits; CBCL = Child Behavior Checklist; CU = callous-unemotional; ECBI = Eyberg Child Behavior Inventory; BPRA = Brief Proactive Reactive Aggression scale; GEM = Griffith Empathy Measure.

Aggression scores after controlling for BPRA Reactive Aggression scores. CAPE symptom sum scores were not significantly correlated with CBCL Anxiety scores, or static child or dynamic adult facial emotion recognition task accuracy scores, controlling for conduct problem severity.

Research Question 3: Do CAPE LPE Diagnoses and CAPE Symptom Sum Scores Offer Incremental Validity Beyond ICU Scores?

CAPE LPE diagnoses predicted higher ECBI Intensity scores (marginal significance at $p = .06$) and lower GEM Affective scores ($p < .001$) beyond ICU scores, controlling for covariates (see Supplemental Table 3). CAPE LPE diagnoses did not significantly predict CBCL Aggression, BPRA Proactive and Reactive Aggression scores, or static child or dynamic adult facial emotion recognition task accuracy scores beyond ICU scores, controlling for covariates ($p > .05$). CAPE LPE symptom sum scores significantly predicted higher ECBI Intensity, CBCL Aggression, and BPRA Proactive Aggression scores, and lower GEM Affective Empathy and static child global emotion recognition task accuracy scores, beyond ICU scores, controlling for covariates (see Supplemental Table 4). CAPE LPE symptom sum scores did not significantly predict CBCL Oppositional Defiance, BPRA Reactive Aggression Scores, or other facial emotion recognition task accuracy scores, beyond ICU scores, controlling for covariates.

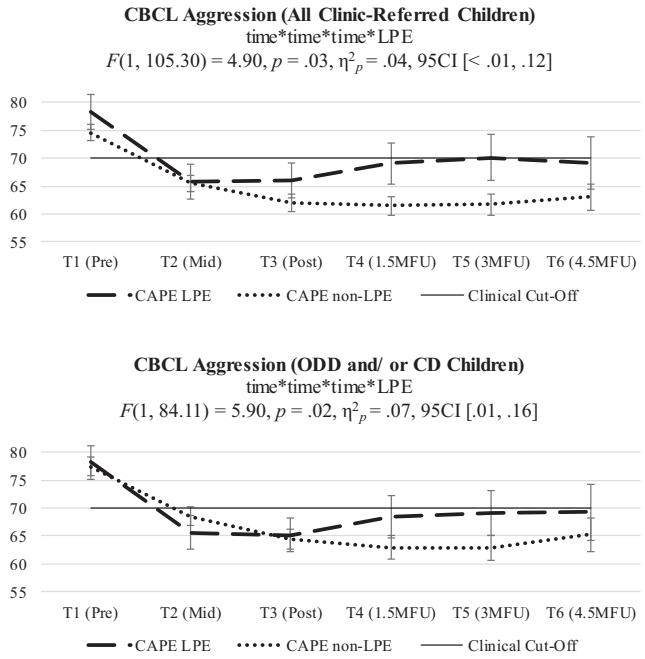
Research Question 4: Do Children Diagnosed With LPE on the CAPE Have Poorer Treatment Prognosis Than Children Without LPE?

Using an iterative process, we retained the model which included the cubic time by LPE group interaction term because it had the lowest Akaike's information criterion and Bayesian information criterion values, relative to other linear, quadratic, and biquadratic models (see Supplemental Table 5). Figure 1 presents mixed model statistics for CBCL Aggression scores across six assessment time points between children with and without an LPE diagnosis ($n = 44$, 10 with LPE). Results indicated a significant cubic time by LPE interaction, showing that children with LPE had poorer treatment outcomes in terms of CBCL Aggression scores than children without LPE (small to medium effect size). When we repeated these analyses with the subset of children with ODD/CD ($n = 34$, 10 with LPE), these significant differences had a medium to large effect size. Supplemental Tables 6 and 7 report estimated marginal means and standard errors for mixed model analysis with all clinic-referred children and a subset of children diagnosed with ODD/CD, respectively. Spline regression analyses indicated that this nonlinear change did not differ between LPE groups after midtreatment ($n = 44$, $\chi^2 = 1.74$, $p = .63$; $n = 34$, $\chi^2 = 1.49$, $p = .69$) or after posttreatment ($n = 44$, $\chi^2 = 1.60$, $p = .50$; $n = 34$, $\chi^2 = 1.48$, $p = .48$).

Supplemental Materials provide results and discussion of additional analyses conducted using CBCL Oppositional Defiance scores as

Figure 1

Mixed Model Analysis Comparing Child Behavior Checklist (CBCL) Aggression Scores Between Clinic-Referred Children With and Without Limited Prosocial Emotions (LPE) on the Clinical Assessment of Prosocial Emotions (CAPE)



Note. ODD = oppositional defiant disorder; CD = conduct disorder; T1 = pretreatment; T2 = midtreatment; T3 = posttreatment; T4 = 1.5-month follow-up; T5 = 3-month follow-up; T6 = 4.5-month follow-up; MFU = month follow-up; 95% CI = 95% confidence interval around η_p^2; Clinical Cut-Off = clinically significant levels of problems based on age- and sex-based norms on the CBCL. All models controlled for child age and sex, treatment delivery method (online or in-person), therapist, and pretreatment Eyberg Child Behavior Inventory (ECBI) Intensity and CBCL Anxiety scores.

treatment outcomes; however, caution is required when interpreting these results due to the poor internal consistency of the CBCL Oppositional Defiance scale in the present study.

Discussion

This study evaluated the interrater reliability, convergent and divergent validity, incremental validity, and clinical prognostic utility of the *DSM-5-TR* "limited prosocial emotions" diagnosis on the CAPE (Frick, 2013) in a large racially and ethnically diverse, mixed-sex sample of clinic-referred children with conduct problems between ages 2 and 8 years. We advanced prior research on the CAPE in four key ways. First, we evaluated the interrater reliability of CAPE LPE symptom ratings with children in a younger age range than previously studied and overcame previous limitations that were hypothesized to explain poor consistency. Second, we extended past findings on the convergent and divergent validity of CAPE LPE diagnoses by using well-validated measures of conduct problem severity and child laboratory task measures of various forms of facial emotion recognition to validate LPE diagnoses against global and fear-specific deficits for the first time. We also evaluated both binary and dimensional approaches to scoring the CAPE. Third, we examined the incremental validity of CAPE LPE binary and dimensional scores beyond ICU scores for predicting scores on theoretically related constructs of CU traits. Fourth, we evaluated the clinical prognostic utility of pretreatment CAPE LPE diagnoses to predict externalizing treatment outcomes across six assessment time points using a subsample of children who received PCIT (n = 44). This study is the first to test treatment trajectories of children diagnosed with versus without *DSM-5-TR* LPE using an assessment tool that aligns with evidence-based clinical assessment practices.

Children Diagnosed With *DSM-5-TR* LPE Have Poorer Treatment Outcomes

Consistent with hypotheses, children diagnosed with LPE differed from children without an LPE diagnosis in their treatment response to an efficacious parent management training intervention, PCIT, that has previously demonstrated large effect size reductions in externalizing problems (d = 1.65; Ward et al., 2016). The change after midtreatment or posttreatment was nonlinear and did not differ between LPE groups. However, children diagnosed with LPE on average ended treatment with higher CBCL Aggression scores that were in the subclinical range (i.e., 65 ≤ CBCL Aggression T score ≤ 69; Achenbach & Rescorla, 2000, 2001), relative to children without an LPE diagnosis whose CBCL Aggression scores were within normal limits (i.e., CBCL Aggression T score ≤ 64). Given we controlled for child age and sex, treatment characteristics (delivery method, therapist), and pretreatment conduct problem severity and anxiety in the analyses, these findings suggest that an LPE diagnosis uniquely predicted poorer treatment response beyond conduct problem severity or comorbid anxiety. These results remained significant when limiting our analyses to children diagnosed with ODD and/or CD, with larger effect sizes observed (η_p^2 = .07 relative to .04 when using the full sample). This finding lends support to the *DSM-5-TR* LPE specifier for identifying young children with conduct disorders who require more targeted interventions than traditional treatments to address their unique needs (Fleming et al., 2022; Kimonis et al., 2019).

Our findings were consistent with prior studies showing that pretreatment CU traits uniquely predicted poorer treatment outcomes in conduct problem severity at posttreatment (Hawes et al., 2014). However, these findings were at odds with the one prior study finding that pretreatment *DSM-5-TR* LPE status, operationalized using questionnaire measures, did not significantly predict externalizing treatment outcomes (Kolko & Pardini, 2010). One explanation for the contradictory findings relates to the method used in Kolko and Pardini (2010) to diagnose LPE, which relied on a few selected APSD CU questionnaire item scores. This method is problematic because prior research that diagnosed LPE using CU questionnaire item scores also did not consistently find significant group differences in key behavioral correlates of CU traits, including the same externalizing problems scale used by Kolko and Pardini (2010; e.g., Kahn et al., 2012). Apart from methodological differences that might explain

mixed results on the impact of LPE status on treatment outcomes, it is also possible that children diagnosed with LPE may respond more to individualized multimodal interventions such as those used by Kolko and Pardini (2010), thus resulting in nonsignificant LPE effects.

LPE Symptoms in Young Children Can Be Reliably Coded Between Assessors

All LPE symptom ratings had adequate interrater reliability and the proportion of children diagnosed with LPE did not differ between the original assessor and the secondary coder. Using Cohen's κ descriptors for the quadratically weighted metric (Altman, 1991), three LPE symptoms had "very good" agreement (lack of remorse/guilt, callous-lack of empathy, and unconcerned about performance in important activities). Only the shallow/deficient affect LPE symptom had slightly poorer but "good" agreement. This may suggest that the shallow/deficient affect symptom could be more difficult to assess using parent-reported information and/or that this symptom manifests differently in young children.

Correspondingly, prior research examining the empirical structure of CU traits in young children found improved model fit when most parent-reported ICU shallow/deficient affect items were removed (Kimonis et al., 2016). Similar to the ICU, the CAPE manual operationalizes the shallow/deficient affect symptom as incongruent affect to situations and a lack of intense emotions except in an ingenuine way or to benefit oneself. A recent study found that 3- to 5-year-old children with elevated LPE had more incongruent facial emotional responses to sad and happy film clips than typically developing children, but this incongruence of emotion to situations did not differ significantly from children with conduct problems but without LPE (Kimonis et al., 2023). It is possible that differences in emotional deficits may be too subtle to be apparent to outward observers. Indeed, most caregivers (89.7%) agreed that their child typically showed their emotions openly to others (i.e., CAPE semistructured interview item 8). Additionally, it is common for young children to express distressed emotions to trigger caregiver responses when they have yet to develop the ability to regulate their emotions (Nelson, 1998). We are unaware of any evidence to suggest that this behavior differs for children with elevated CU traits.

Thus, the clinical description of the shallow/deficient affect symptom may require refinement to include age-appropriate behavioral indicators for young children. Drawing from Bowlby's (1944) historical work on affectionless delinquent children (>5 years old), a lack of affectionate behavior toward important others and unresponsiveness to situations that would typically evoke positive (kindness, affection) or negative (punishment) emotions may be indicators of shallow/deficient affect that are relevant to young children. Promisingly, there is research underway to develop LPE symptom descriptions for young children (Neo et al., 2023), who currently rely on CAPE symptom descriptions that apply to children up to 21 years old. By qualitatively coding parent-report transcripts of CAPE semistructured interviews, the authors identified several LPE indicators in young children: lack of remorse/guilt (e.g., continues playing after they hurt someone else), callous-lack of empathy (e.g., does not get upset when movie/TV show characters are upset), unconcerned about performance in important activities (e.g., does the bare minimum to complete most tasks), and shallow/deficient affect (e.g., pretends to cry when they are in trouble). Further research is needed to test whether these age-appropriate LPE indicators improve the measurement reliability of CAPE LPE scores in young children.

Both standard and quadratically weighted Cohen's κ values in the present study were higher than the standard Cohen's κ values found in the sample of incarcerated adolescent boys (Molinuevo et al., 2020). This difference could suggest that the availability of nonverbal information is important for reliably scoring the CAPE since the secondary coder relied solely on audio recordings of CAPE interviews for most cases in the Molinuevo et al. (2020) study. Prior meta-analytic research found that certain nonverbal characteristics (e.g., low facial pleasantness [frowning, sneering], using few gestures to illustrate speech, and object-/self-fidgeting) are discrepant with truthful responding (DePaulo et al., 2003); however, further research is needed to examine how nonverbal characteristics could impact on the reliability of LPE assessments. Alternatively, this improved interrater reliability relative to prior research was because the present study used all available information on CU traits, including CU questionnaire item ratings and semistructured clinical interviews, to rate LPE symptoms, as recommended in the CAPE manual (Frick, 2013).

In the present study, interrater reliability estimates were higher when using quadratically weighted Cohen's κ statistic than its standard metric. We argue that the former is a more accurate measure of interrater reliability because it accounts for the level of disagreement between raters and the unequal importance between rating categories. This is important because clinicians will be cautious about assigning a symptom rating of "2" because of the clinical impact of assigning an LPE diagnosis. Consequently, the interrater reliability estimates found by Molinuevo et al. (2020) may be underestimates.

LPE Diagnosis Distinguishes Young Children With Severe Conduct Problems

Consistent with prior research (Centifanti et al., 2020; Hawes et al., 2020), children with LPE had higher CU traits, lower empathy, and did not differ in anxiety levels, relative to children without LPE. Importantly, differences between LPE groups remained significant after accounting for child age and sex, sampling differences between clinic sites, anxiety, and conduct problem severity. Effect size differences between children with and without LPE were medium to large for the ICU and small to medium for the CBCL CU, which is consistent with *DSM-5-TR* LPE criteria being determined based on factor analytic studies of the ICU and its precursor, the APSD, whereas CBCL CU items only assess two *DSM-5-TR* LPE symptoms. Using a dimensional scoring approach, CAPE symptom sum scores were positively correlated with CU traits, conduct problem severity, and aggression; negatively correlated with affective empathy; and uncorrelated with anxiety, as expected.

Results showed that CAPE symptom sum scores offered incremental validity over ICU scores for predicting greater conduct problem severity and aggression, and lower affective empathy and global emotional recognition accuracy (static child facial stimuli) scores. This finding suggests that the CAPE captures important clinical information for predicting outcomes that are not captured by CU questionnaire measures (e.g., symptom persistence and pervasiveness, credibility of source). In contrast, CAPE LPE categorical diagnoses offered poorer incremental validity over ICU scores, relative to CAPE dimensional scores, and only predicted greater conduct problem

severity and empathy deficits. Overall, these findings appear to support the clinical utility of dimensional LPE scores, which is important to consider in future iterations of the *DSM-5-TR* and *International Classification of Diseases and Related Health Problems, Eleventh Edition*.

Contrary to prior CAPE research, children diagnosed with LPE in the present study had more severe externalizing problems than children without LPE across various measures, including conduct problem severity and proactive aggression, but did not differ in reactive aggression. These findings support the divergent validity of CAPE LPE diagnoses for identifying children with more severe conduct problems, including premeditated types of aggression that are characteristic of LPE (Muñoz & Frick, 2012). The larger effect size differences between children with and without CAPE LPE for ICU and GEM Affective scores than for conduct problem scores also suggest the CAPE better captures CU traits than conduct problems. Similarly, stronger correlations were observed between CAPE symptom sum scores and CU scores than with conduct problem scores.

Group differences between children with and without CAPE LPE diagnoses remained mostly consistent within the subsample of children diagnosed with ODD/CD. This finding supports the LPE distinction for CD in *DSM-5-TR* and suggests that LPE is relevant to the diagnosis of ODD and to conduct problems more generally (Hawes et al., 2014). The only exception was for proactive aggression, which did not differentiate LPE groups within the ODD/CD sample, likely due to the small sample size that impacted on statistical power (n LPE = 6, n non-LPE = 46), and for the CBCL CU scale that was only marginally significant ($p = .07$). Notably, the CBCL CU scale does not assess two *DSM-5-TR* LPE symptoms (lack of empathy, unconcerned about performance), which highlights the importance of using LPE/CU tools that capture all four LPE symptoms to screen for *DSM-5-TR* LPE.

Unexpectedly, children with and without LPE did not differ in facial emotion recognition accuracy when using the same child laboratory tasks used in a prior study that found negative associations between ICU scores and global and fear-specific emotion recognition accuracy in young children (Kimonis et al., 2016). These mixed findings were not explained by own-race bias effects as there were no statistically significant differences in fear or global emotional facial recognition accuracy between White and non-White children, or between CAPE LPE groups among White children ($p > .05$). These unexpected findings could be due to statistically significant differences between task completers and noncompleters in child age, conduct problem severity, aggression, and CU traits (see Supplemental Tables 1 and 2), thus limiting the generalizability of these findings. We recommend caution in their interpretation and a need for further study replication.

Limitations

There are several study limitations. First, we used ICU and CBCL CU scores as sources of information when scoring the CAPE and as indicators of criterion validity, which may result in predictor-criterion contamination. The rationale for this decision was to follow the administration and scoring recommendations from the CAPE manual as closely as possible by relying on other sources of information about CU traits that would be available to clinicians, thus generalizing findings to clinicians who use the CAPE for clinical decision making. Nonetheless, future research should replicate our findings using LPE measures that are not simultaneously used to score the CAPE. Second, we were unable to covary race and/or ethnicity in analyses because these data were not routinely collected for children recruited from the community-based clinic. It is possible that this may not have impacted findings since children from our university-based clinic diagnosed with and without LPE did not differ in their race and/or ethnicity, $\chi^2(1, N = 119) = 4.26$, $p = .37$. Nonetheless, we controlled for a study site in all analyses to account for sampling differences between clinics. Third, CAPE semistructured interviews were not conducted with the child's teacher, as is recommended in the CAPE manual. This is because many of our child participants were younger than school age. Instead, we emphasized the need for both caregivers to attend assessments to reduce mono-source bias when scoring the CAPE (44% of CAPEs were scored using information from both caregivers). We also asked caregivers to report on their child's behavior at daycare and/or school during the general clinical interview and whether their child's problems interfered with their daycare/school functioning during the DISCAP semi-structured interview. Future research is needed to validate the CAPE using information beyond caregiver reports. Fourth, we did not examine the test–retest reliability of CAPE LPE scores. Fifth, we used a small subsample of 36 children (15%) to evaluate the interrater reliability of the CAPE, and future research with larger sample sizes is recommended. Last, future research is recommended to further evaluate the convergent validity of the CAPE against neurocognitive, biological, and psychophysiological LPE correlates, as well as its divergent validity relative to the current *DSM-5-TR* age-of-onset specifier and other proposed specifiers (see Salekin, 2016) for subtyping antisocial children.

Conclusion

The CAPE is a psychometrically sound tool to assess and diagnose *DSM-5-TR* LPE in young clinic-referred children between ages 2 and 8 years. The presence of an LPE diagnosis suggests that clinicians must tailor their intervention approach to the unique needs of children with LPE to maximize their treatment outcomes. With the availability of novel early interventions for children with LPE (e.g., PCIT adapted for children with CU traits [PCIT-CU]; Fleming et al., 2022; Kimonis et al., 2019), early assessment and intervention for LPE will likely improve outcomes of children diagnosed with LPE.

References

Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*. University of Vermont, Research Center for Children, Youth, and Families.

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. University of Vermont, Research Center for Children, Youth, and Families.

Altman, D. G. (1991). *Practical statistics for medical research*. Chapman and Hall.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Australian Bureau of Statistics. (2021). *Census data*. Retrieved July 18, 2022, from https://abs.gov.au/census/find-census-data/search-by-area

Bansal, P. S., Goh, P. K., Eng, A. G., Elkins, A. R., Thaxton, M., & Martel, M. M. (2023). Utility of the limited prosocial emotions specifier in preschoolers with conduct problems. *Assessment*, *30*(2), 274–286. https://doi.org/10.1177/10731911211051070

Bowlby, J. (1944). Forty-four juvenile thieves: Their characters and home-life. *The International Journal of Psycho-Analysis*, *25*, 19–53.

Brown, K., Atkins, M. S., Osborne, M. L., & Milnamow, M. (1996). A revised teacher rating scale for reactive and proactive aggression. *Journal of Abnormal Child Psychology*, *24*(4), 473–480. https://doi.org/10.1007/BF01441569

Castagna, P. J., Babinski, D. E., Pearl, A. M., Waxmonsky, J. G., & Waschbusch, D. A. (2021). Initial investigation of the psychometric properties of the Limited Prosocial Emotions questionnaire (LPEQ). *Assessment*, *28*(8), 1882–1896. https://doi.org/10.1177/1073191120927782

Centifanti, L. C. M., Shaw, H., Atherton, K. J., Thomson, N. D., MacLellan, S., & Frick, P. J. (2020). CAPE for measuring callous-unemotional traits in disadvantaged families: A cross-sectional validation study. *F1000Research*, *8*, Article 1027. https://doi.org/10.12688/f1000research.19605.1

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*(4), 213–220. https://doi.org/10.1037/h0026256

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Academic Press.

Colins, O. F., Fanti, K. A., & Andershed, H. (2021). The DSM-5 limited prosocial emotions specifier for conduct disorder: Comorbid problems, prognosis, and antecedents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *60*(8), 1020–1029. https://doi.org/10.1016/j.jaac.2020.09.022

Dadds, M. R., Hunter, K., Hawes, D. J., Frost, A. D. J., Vassallo, S., Bunn, P., Merz, S., & Masry, Y. E. (2008). A measure of cognitive and affective empathy in children using parent ratings. *Child Psychiatry and Human Development*, *39*(2), 111–122. https://doi.org/10.1007/s10578-007-0075-4

DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, *129*(1), 74–118. https://doi.org/10.1037/0033-2909.129.1.74

Déry, M., Bégin, V., Toupin, J., & Temcheff, C. (2019). Clinical utility of the limited prosocial emotions specifier in the childhood-onset subtype of conduct disorder. *Canadian Journal of Psychiatry*, *64*(12), 838–845. https://doi.org/10.1177/0706743719885469

Ekman, P., & Friesen, W. V. (1976). Measuring facial movement. *Environmental Psychology and Nonverbal Behavior*, *1*(1), 56–75. https://doi.org/10.1007/BF01115465

Eyberg, S. M., & Pincus, D. (1999). *ECBI & SESBI-R: Eyberg child behavior inventory and sutter-eyberg student behavior inventory-revised: Professional manual*. Psychological Assessment Resources.

Fleming, G. E., Neo, B., Briggs, N. E., Kaouar, S., Frick, P. J., & Kimonis, E. R. (2022). Parent training adapted to the needs of children with callous-unemotional traits: A randomized controlled trial. *Behavior Therapy*, *53*(6), 1265–1281. https://doi.org/10.1016/j.beth.2022.07.001

Frick, P. J. (2004). *Inventory of callous-unemotional traits*. Louisiana State University. https://faculty.lsu.edu/pfricklab/icu.php

Frick, P. J. (2013). *Clinical assessment of prosocial emotions*. Louisiana State University. https://faculty.lsu.edu/pfricklab/cape.php

Frick, P. J., Barry, C. T., & Kamphaus, R. W. (2020). *Clinical assessment of child and adolescent personality and behavior* (4th ed.). Springer. https://doi.org/10.1007/978-3-030-35695-8

Frick, P. J., & Hare, R. D. (2001). *The antisocial process screening device*. Multi-Health Systems.

Frick, P. J., & Moffitt, T. E. (2010). *A proposal to the DSM-V childhood disorders and the ADHD and disruptive behavior disorders work groups to include a specifier to the diagnosis of conduct disorder based on the presence of callous-unemotional traits*. American Psychiatric Association. https://www.researchgate.net/publication/309107121_A_proposal_to_the_DSM-V_Childhood_Disorders_and_the_ADHD_and_Disruptive_Behavior_Disorders_Work_Groups_to_include_a_specifier_to_the_Diagnosis_of_Conduct_Disorder_based_on_the_presence_of_callous-unem

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), 1–57. https://doi.org/10.1037/a0033076

Hawes, D. J., Kimonis, E. R., Mendoza Diaz, A., Frick, P. J., & Dadds, M. R. (2020). The clinical assessment of prosocial emotions (CAPE 1.1): A multi-informant validation study. *Psychological Assessment*, *32*(4), 348–357. https://doi.org/10.1037/pas0000792

Hawes, D. J., Price, M. J., & Dadds, M. R. (2014). Callous-unemotional traits and the treatment of conduct problems in childhood and adolescence: A comprehensive review. *Clinical Child and Family Psychology Review*, *17*(3), 248–267. https://doi.org/10.1007/s10567-014-0167-1

Johnson, S., Barrett, P. M., Dadds, M. R., Fox, T., & Shortt, A. (1999). The diagnostic interview schedule for children, adolescents, and parents: Initial reliability and validity data. *Behaviour Change*, *16*(3), 155–164. https://doi.org/10.1375/bech.16.3.155

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

Kimonis, E. R., Fanti, K. A., Anastassiou-Hadjicharalambous, X., Mertan, B., Goulter, N., & Katsimicha, E. (2016). Can callous-unemotional traits be reliably measured in preschoolers? *Journal of Abnormal Child Psychology*, *44*(4), 625–638. https://doi.org/10.1007/s10802-015-0075-y

Kimonis, E. R., Fanti, K. A., Frick, P. J., Moffitt, T. E., Essau, C., Bijttebier, P., & Marsee, M. A. (2015). Using self-reported callous-unemotional traits to cross-nationally assess the DSM-5 "With limited prosocial emotions" specifier. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *56*(11), 1249–1261. https://doi.org/10.1111/jcpp.12357

Kimonis, E. R., Fleming, G., Briggs, N., Brouwer-French, L., Frick, P. J., Hawes, D. J., Bagner, D. M., Thomas, R., & Dadds, M. (2019). Parent–child interaction therapy adapted for preschoolers with callous-unemotional traits: An open trial pilot study. *Journal of Clinical Child and Adolescent Psychology*, *48*(Supp. 1), S347–S361. https://doi.org/10.1080/15374416.2018.1479966

Kimonis, E. R., Le, B., Fleming, G. E., Kyranides, M. N., Demetriou, C. A., Fanti, K. A., Neo, B., Prasad, A. H., McGregor, K., Hawes, D. J., & Eapen, V. (2023). Facial reactions to emotional films in young children with conduct problems and varying levels of callous-unemotional traits. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *64*(3), 357–366. https://doi.org/10.1111/jcpp.13701

Kolko, D. J., & Pardini, D. A. (2010). ODD dimensions, ADHD, and callous-unemotional traits as predictors of treatment response in children with disruptive behavior disorders. *Journal of Abnormal Psychology*, *119*(4), 713–725. https://doi.org/10.1037/a0020910

Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition and Emotion*, *24*(8), 1377–1388. https://doi.org/10.1080/02699930903485076

Molinuevo, B., Martínez-Membrives, E., Pera-Guardiola, V., Requena, A., Torrent, N., Bonillo, A., Batalla, I., Torrubia, R., & Frick, P. J. (2020). Psychometric properties of the clinical assessment of prosocial emotions: Version 1.1 (CAPE 1.1) in young males who were incarcerated. *Criminal Justice and Behavior*, *47*(5), 547–563. https://doi.org/10.1177/0093854819892931

Muñoz, L. C., & Frick, P. J. (2012). Callous-unemotional traits and their implication for understanding and treating aggressive and violent youths. *Criminal Justice and Behavior*, *39*(6), 794–813. https://doi.org/10.1177/0093854812437019

Nelson, J. K. (1998). The meaning of crying based on attachment theory. *Clinical Social Work Journal*, *26*(1), 9–22. https://doi.org/10.1023/A:1022841427355

Neo, B., & Kimonis, E. R. (2021). Callous-unemotional traits linked to earlier onset of self-reported and official delinquency in incarcerated boys. *Law and Human Behavior*, *45*(6), 554–565. https://doi.org/10.1037/lhb0000472

Neo, B., Koh, L., Lu, T., Fleming, G. F., Kaouar, S., & Kimonis, E. R. (2023). *Behavioral indicators of limited prosocial emotions/callous-unemotional traits in young children* [Manuscript in preparation]. School of Psychology, University of New South Wales.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Pardini, D., Stepp, S., Hipwell, A., Stouthamer-Loeber, M., & Loeber, R. (2012). The clinical utility of the proposed DSM-5 callous-unemotional subtype of conduct disorder in young girls. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(1), 62–73.e4. https://doi.org/10.1016/j.jaac.2011.10.005

Riediger, M., Voelkle, M. C., Ebner, N. C., & Lindenberger, U. (2011). Beyond "happy, angry, or sad?": Age-of-poser and age-of-rater effects on multi-dimensional emotion perception. *Cognition and Emotion*, *25*(6), 968–982. https://doi.org/10.1080/02699931.2010.540812

Salekin, R. T. (2016). Psychopathy in childhood: Toward better informing the DSM-5 and ICD-11 conduct disorder specifiers. *Personality Disorders*, *7*(2), 180–191. https://doi.org/10.1037/per0000150

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Thomas, R., & Zimmer-Gembeck, M. J. (2012). Parent–child interaction therapy: An evidence-based treatment for child maltreatment. *Child Maltreatment*, *17*(3), 253–266. https://doi.org/10.1177/1077559512459555

Waller, R., Hyde, L. W., Grabell, A. S., Alves, M. L., & Olson, S. L. (2015). Differential associations of early callous-unemotional, oppositional, and ADHD behaviors: Multiple domains within early-starting conduct problems? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *56*(6), 657–666. https://doi.org/10.1111/jcpp.12326

Ward, M. A., Theule, J., & Cheung, K. (2016). Parent–Child Interaction Therapy for child disruptive behaviour disorders: A meta-analysis. *Child and Youth Care Forum*, *45*(5), 675–690. https://doi.org/10.1007/s10566-016-9350-5

Willoughby, M. T., Mills-Koonce, W. R., Gottfredson, N. C., & Wagner, N. (2014). Measuring callous unemotional behaviors in early childhood: Factor structure and the prediction of stable aggression in middle childhood. *Journal of Psychopathology and Behavioral Assessment*, *36*(1), 30–42. https://doi.org/10.1007/s10862-013-9379-9

Willoughby, M. T., Waschbusch, D. A., Moore, G. A., & Propper, C. B. (2011). Using the ASEBA to screen for callous unemotional traits in early childhood: Factor structure, temporal stability, and utility. *Journal of Psychopathology and Behavioral Assessment*, *33*(1), 19–30. https://doi.org/10.1007/s10862-010-9195-4

World Health Organization. (2019). *International statistical classification of diseases and related health problems* (11th ed.). https://icd.who.int/en

Zuckerman, M., Driver, R., & Koestner, R. (1982). Discrepancy as a cue to actual and perceived deception. *Journal of Nonverbal Behavior*, *7*(2), 95–100. https://doi.org/10.1007/BF00986871

Received January 19, 2023
Revision received May 24, 2023
Accepted August 4, 2023 ■