Supplementary Materials

The Promoting Empathy and Affiliation in Relationships (PEAR) Study: Protocol for a Longitudinal Study Investigating the Development of Early Childhood Callous-Unemotional Traits

Table of Contents

Questionnaires

The selection of questionnaires used to assess core constructs within the PEAR study was guided by existing literature, information about the measures' validity and reliability, and our own pilot data. The psychometric properties of existing questionnaires will be calculated before their use in analyses, and, in most cases, established procedures for creating composite subscales will be followed. The new measures developed for use in the PEAR study will undergo the same psychometric evaluation. Established data reduction techniques (e.g., exploratory and confirmatory factor analysis, item functioning analysis) will guide the creation of composite subscales. After eligibility has been determined, participating parents will be sent a link to complete the questionnaire battery before their lab visit through REDCap (Patridge & Bardyn, 2018).

Parent-Reported Questionnaires About Themselves

Adverse childhood experiences. We measure parents' own experiences of adversity during their childhood using the 15-item Adverse Childhood Experiences Measure (ACES-parent) (Felitti et al., 1998) (e.g., "Did you live with anyone who was a problem drinker or alcoholic or who used street drugs?"). Items are rated as either "yes" or "no". A total adverse experience score is computed by summing the number of "yes" responses.

Anxiety symptomatology. We measure anxiety symptomatology in parents using the 7-item Generalized Anxiety Disorder (GAD-7) scale (Löwe et al., 2008) (e.g., "Over the last two weeks, how often have you been bothered by not being able to stop or control worrying?"). Items are rated on a 4-point Likert scale from "Not at all" (1) to "Nearly every day" (4). An anxiety severity score is computed by summing all items.

Chaos in the home. To measure chaos and disorganization in the home, we use the 15-item Confusion, Hubbub, and Order Scale (CHAOS) (Matheny Jr et al., 1995) (e.g., "There is very little commotion in our house"). Items are rated on a 4-point Likert scale, ranging from

"Very much like our home" (1) to "Not at all like our home" (4) with items summed to index more household chaos.

COVID-19 exposure and impact on the family. We use the 30-item COVID-19 Exposure and Family Impact Scales (CEFIS) to assess the impact of COVID-19 on family members within households, extended family, and close friends (Kazak et al., 2021). Some items require a "yes/no" response (e.g., "We had a 'stay at home' order"). Other items (e.g., parenting, physical well-being) are rated on a 4-point Likert scale ranging from "[COVID-19] made it a lot better" (1) to "[COVID-19] made it a lot worse" (4), with an option to indicate if the item is not applicable. Still other items are rated on a 10-point Likert scale, from "No distress" (1) to "Extreme distress" (10) (e.g., "Overall, how much distress have you experienced related to COVID-19?"). One item is open response, and probes other positive or negative effects of COVID-19 on the family.

Depression symptomatology. We measure depression symptomatology in parents using the 9-item Patient Health Questionnaire (PHQ-9) (Kroenke et al., 2001), which assesses mood and sleep experiences (e.g., "Over the last two weeks, how often have you been bothered by having little interest or pleasure in doing things?"). Items are rated on a 4-item Likert scale from "Not at all" (1) to "Nearly every day" (4). A depression severity score is obtained by summing all items.

Emotion regulation. To measure emotion regulation difficulties in parents, we use the 36-item Difficulties in Emotion Regulation Scales (DERS-5) (Gratz & Roemer, 2004), which assesses parental emotion regulation efforts and experiences (e.g., "I experience my emotions as overwhelming and out of control"). Items are rated on a 5-point Likert scale from "Almost never (0-10%)" to "Almost always (91-100%)". In addition to a total summed emotion regulation score, there are 6 subscales: Awareness, Clarity, Goals, Impulse, Nonacceptance, and Strategies.

Harsh and ineffective parenting practices. To measure harsh or ineffective parenting practices, we use the 30-item Parenting Scale (PS) (Arnold et al., 1993), which assesses

ineffective parenting behaviors. Items contain a stem followed by a response option rated on a 7-point Likert scale with 3 subscales: Laxness (e.g., "I threaten things that: 'I'm sure I can carry out' (1) to 'I know I won't actually do' (7)"), Over-Reactivity (e.g., "When I'm upset or under stress: I am picky and on my child's back" (1) to "I am not more picky than usual" (7)"), Verbosity (e.g., "Before I do something about a problem: 'I give my child several reminders and warnings' (1) to 'I use only one reminder and warning' (7)). Items within subscales are summed to index harsher or more ineffective parenting.

Instrumental support. To assess parental perceptions of material and tangible support in the home (i.e., instrumental help), we use the 9-item Maternal Social Support Index (MSSI) (Pascoe et al., 1988), which assesses the division of responsibility within the household (e.g., grocery shopping, paying bills). Items are rated as either "I take sole responsibility" (0) or "Someone else does this task or helps me complete the task" (1). Items are summed to create a scale with higher scores indicating more instrumental help with household tasks.

Neighborhood impoverishment. To measure neighborhood impoverishment, we use a validated 17-item scale that assesses difficulties parents perceive in their community and neighborhood (e.g., unemployment, organized crime) (Shaw et al., 2004; Shaw et al., 1998). Items are rated on a 3-point Likert scale, ranging from "Not a problem" (1) to "A big problem" (3), then summed to index more neighborhood impoverishment.

Parenting stress. We measure parenting stress using the 37-item Parental Stress Index (Abidin, 1995), which includes items such as "I often have the feeling that I cannot handle things very well" and "My child seems to cry or fuss more often than most children". Items are rated on a 5-point Likert scale ranging from "Strongly disagree" (1) to "Strongly agree" (5), except for one write-in response. In addition to a total stress sum score, 4 subscales include Defensive Responding, Parental Distress, Parent-Child Dysfunctional Interaction, and Child Difficulty (Haskett et al., 2006).

Parental beliefs about children's emotions. To assess emotion scaffolding and parental understanding of children's emotion, we use the 33-item Parents' Beliefs about Children's Emotions (PBACE) scale (Halberstadt et al., 2013), which assesses parental conceptualizations of children's emotional experiences and states (e.g., "When children are sad, they need to find their own ways to move on"). Items are rated on a 6-point Likert scale from "Strongly disagree" (1) to "Strongly agree" (6) and summed.

Personality. To measure parental personality traits, we use the 20-item Mini International Personality Item Pool (Mini-IPIP) (Donnellan et al., 2006), which provides scores for Extraversion (e.g., "life of the party"), Agreeableness (e.g., "sympathize with others' feelings"), Conscientiousness (e.g., "get chores done right away"), Neuroticism (e.g., "have frequent mood swings"), and Intellect/Imagination (e.g., "have a vivid imagination"). Items are rated on a 5-point Likert scale from "Very inaccurate" (1) to "Very accurate" (5) and summed within subscales.

Positive parenting. To measure positive parenting, we use the 16-item Comprehensive Early Childhood Parenting (CEPAQ) (Verhoeven et al., 2017), which assesses aspects of positive parenting as reported by the parent. Subscales include Sensitivity (e.g., "I notice when my child is sad or doesn't feel good"), Responsiveness (e.g., "When my child is having a hard time, I am able to help him/her"), Affection (e.g., "I hug, kiss, or hold my child for no particular reason"), and Activities (e.g., "I tell my child stories or read books to them"). Items are rated on a 6-point Likert scale from "Never" (1) to "Always" (6) and summed within subscales.

Psychopathic traits. To assess psychopathic traits in parents, we use the Self-Report Psychopathy (SRP-SF) scale (Gordts et al., 2017; Neumann & Pardini, 2014), which is a 29-item measure that assesses psychopathic personality traits and harmful behaviors, including Interpersonal (e.g., "I have pretended to be someone else in order to get something"), Affective (e.g., "Most people are wimps"), Lifestyle (e.g., "I'm a rebellious person"), and Antisocial subscales (e.g., "I have threatened people into giving me money, clothes, or makeup"). Items

are rated on a 5-point Likert scale, from “Disagree Strongly” (1) to “Agree Strongly” (5). In addition to a total sum score, subscale scores are obtained.

Quality of social support. We measure parental perceptions of the quality of their current social support using the Quality of Social Support (QSS) scale (Crnic & Booth, 1991), which is a 16-item including items such as “How satisfied are you with the number of times you talk on the phone with your friends during a typical week?”. Items are rated on a 4-point Likert scale from “Very dissatisfied (I wish things were very different)” (1) to “Very satisfied (I’m really pleased)” (4), with the option to indicate if an item is not applicable. Items are summed to create a total satisfaction score.

Relationship conflict. To measure relationship conflict in the home, we use parent report on the Conflict Tactics Scales Short Form (CTS-2 SF) (Straus & Douglas, 2004), a 20-item measure completed by parents to assess the frequency of different types of conflict in partner/spousal relationships (e.g., “I explained or suggested a compromise for a disagreement with my partner”, “My partner insulted or swore or shouted or yelled at me”). Items are rated on a 6-point Likert scale from “Once in the past year” (1) to “More than 20 times in the past year” (6). In addition to a total conflict score, there are three subscales: Physical Assault, Injury, and Sexual Coercion.

Sensitivity to threat and affiliation. To measure parental sensitivity to threat and affiliation, we use the 28-item self-reported version of the Sensitivity to Threat and Affiliative Reward Scale (STARS) (Perlstein et al., 2022), which assesses sensitivity to social threat (“It would bother me if someone else around me was crying”), non-social threat (“I worry about dangerous things or accidents happening”), physical affiliation (“I like to hug or kiss people to say hello or goodbye”), and non-physical affiliation (“I like to talk about my feelings with people”). The Threat Sensitivity subscale is computed by summing the 13 threat-related items and the Affiliative Reward subscale by summing the 15 affiliation items.

Substance misuse. To measure parental substance misuse, we use the 7-item Substance Use Screener (SUS) (Sullivan et al., 2020), which assesses parental alcohol and drug habits. Some items are rated on 6-point Likert scales (e.g., "During the last 30 days, how often did you have any type of alcoholic beverage?" with response options ranging from "Not at all" to "Every day"), some require yes/no responses (e.g., "Have you ever smoked a cigarette?"), and some require that participant write in their answers (e.g., "About how old were you when you had your first cigarette?).

Parent-Reported Questionnaires About Their Child

Adverse childhood experiences. We assess children's adverse experiences using parent report on the 17-item Pediatric Adverse Childhood Experiences and Related Life-Events Screener (PEARLS) (Ye et al., 2023). Items are rated as either "yes" or "no" (e.g., "Has your child ever lived with a parent/caregiver who went to jail/prison?") and the total number of endorsements is calculated.

Affiliative reward. To measure children's affiliative reward, we use the newly-developed 28-item Components of Affiliative Reward Experiences Scale (CARES) (Paz et al., 2023), which assesses different aspects of affiliation toward peers, family and strangers. Items are rated on a 4-point scale from "Never" (0) to "Always" (3) and summed. There are 4 subscales: Wanting to Affiliate (e.g., "Wants to be liked by other children"), Enjoyment from Experiences of Affiliation (e.g., "Likes to be part of family activities"), Understanding of Affiliative Relationships (e.g., "Understands and explains which children are his/her friends"), and Affiliative Enactment Behavior (e.g., "Plays in groups with (not just beside) other children").

Autism traits. We assess autistic traits in children using parent report on the 50-item Autism Quotient (AQ-Child) (Ashwood et al., 2016) (e.g., "[My child] prefers to do things the same way over and over again", "[My child] often notices small sounds when others do not").

Items are rated on a 4-point Likert scale from “Definitely agree” (1) to “Definitely disagree” (4). A total sum score is obtained, with higher scores indicating more autistic traits.

Behavior problems. We measure child behavior problems and prosocial behavior using parent report on the 33-item Strengths and Difficulties Questionnaire (SDQ) (Goodman et al., 2000), which assesses areas of strength and difficulty for children (e.g., “[My child] often loses temper”). Items 1-25 are rated on a 3-point Likert scale from “Not true” (1) to “Certainly true” (3). The remaining items probe the extent of the difficulty experienced by the child and family, and in what domains (e.g., friendships, classroom learning). In addition to a total difficulties summed score, we compute subscale scores summing items within the Conduct Problems, Emotional Problems, Hyperactivity, Peer Problems, Prosociality, and Impact scales (i.e., how impactful difficulties have been for child and family).

Behavioral inhibition. We measure child behavioral inhibition using the 30-item Behavioral Inhibition Questionnaire (BIQ) (Kim et al., 2011). Parents rate how often a statement (e.g., “[My child] approached new situations or activities very hesitantly”) describes their own child on a 7-point Likert scale from “Hardly ever” (1) to “Almost always” (7). In addition to a total behavioral inhibition sum score, subscale scores can be obtained for Unfamiliar Peer Engagement, Unfamiliar Adult Engagement, Performance Situations, Separation/Preschool, Unfamiliar Situations, and Physical Challenge.

Callous-unemotional traits. Child callous-unemotional (CU) traits are assessed using parent report on the 24-item Inventory of Callous-Unemotional Traits (ICU) (Cardinale & Marsh, 2020; Kimonis et al., 2016; Kimonis et al., 2008), which provides subscale scores for Callousness (e.g., “unconcerned about feelings of others”), Uncaring (e.g., “always tries best”), and Unemotionality (e.g., “hides feelings”). Items are rated on a 4-point scale from “Not at all true” (0) to “Definitely true” (3) and summed.

Child temperament. We measure child temperament using 30-item Colorado Child Temperament Inventory (CCTI) (Plomin & Rowe, 1977). Items are rated on a 5-point Likert

scale from “Strongly disagree” (1) to “Strongly agree” (5). There are 6 sum-score subscales: Sociability (e.g., “[My child] makes friends”), Activity (e.g., “[My child is] energetic”), Emotionality (e.g., “[My child is] upset easily”), Attention Span-Persistence (e.g., “[My child] plays with a single toy”), Reaction to Food (e.g., “[My child] rarely took new food without fussing”), and Soothability (e.g., “[My child is] easily distracted when crying”)

Problematic traits. We measure problematic personality traits and temperament in children using the 28-item Child Problematic Traits Inventory (CPTI) (Colins et al., 2014), which assesses grandiose deceitfulness, callous unemotionality, and need for stimulation. Items are rated on a 40-point Likert scale from “Does not apply at all” (1) to “Applies very well” (4). In addition to a total sum score for problematic traits, subscales may be calculated for Grandiose Deceitfulness (e.g., “[My child] lies often to avoid problems”), Callous Unemotionality (e.g., “[My child] seldom expresses sympathy for others”), and Need for Stimulation (e.g., “[My child] likes change and that things happen all the time”).

Psychopathology. To measure general symptoms of psychopathology in children, we use parent reports on the Child Behavior Checklist (Achenbach & Rescorla, 2000), a widely-used 100-item measure that assesses behavioral difficulties in preschool-aged children. Items (e.g., “[My child] acts too young for age”) are rated on a 3-point Likert scale from “Not true (as far as you know)” (0) to “Very true or often true” (3). We compute total sum scores for Internalizing and Externalizing psychopathology, as well as DSM-5 oriented subscales assessing symptoms of Anxiety, Oppositional Defiant, and Attention-Deficit Hyperactivity Disorders.

Sensitivity to threat and affiliative reward. Child sensitivity to threat and affiliation are measured using the parent-reported version of the Sensitivity to Threat and Affiliative Reward Scale (STARS) (Perlstein et al., 2022), described above.

Sleep. We assess child sleep using parent report on the 20-item Brief Sleep Questionnaire (BISQ-R) (Mindell et al., 2019), which assesses child sleep patterns (e.g., “In a

typical week, how often does your child usually have the exact same bedtime routine?") and caregiver impressions of sleep behavior (e.g., "Typically, how difficult is bedtime?").

Social preferences. We measure children's social preferences using the 11-item Child Social Preference Scale (CSPS) (Coplan et al., 2004), which assesses children's social tendencies. Items are rated on a 5-point Likert scale from "Not a lot" (1) to "A lot" (5). In addition to a total social preference sum score, subscale scores for Shyness (e.g., "My child seems to want to play with other children, but is sometimes nervous to") and Social Disinterest (e.g., "My child often seems content to play alone") can be obtained.

Social Responsiveness. We measure child social responsiveness using the 65-item Social Responsiveness Scale (SRS-2) (Uljarević et al., 2019), which assesses social ability and responsiveness in children aged 2 years, 5 months to 18 years. Items are rated on a 4-point Likert scale from "Not true" (1) to "Almost always true" (4). A total sum score measuring social deficits is obtained, in addition to five subscales that include Social Awareness, Social Cognition, Social Communication, Social Motivation, and Restricted Interests.

Physiology and Eye-tracking

Simultaneous Assessment of Attention and Physiology

A key point of innovation of the PEAR Study is the multi-method approach to assessing risk factors for CU traits. Specifically, we assess multiple indicators of neurophysiological and attentional functioning across contexts. Participating parents and children wear mobile devices to allow for the continuous and time-synchronized collection of cardiac physiology and respiration data, and both stationary and mobile eye-tracking devices will be used to assess various aspects of attention across computer-based and interaction tasks. In addition to collecting these data while participants are engaged in various tasks, we collect tonic or resting data while they watch a 3-minute nature clip. The child watches the video on a touchscreen computer while the parent watches the video concurrently on a desktop monitor in a separate

room. The maintenance of homeostasis is itself a dynamic process that provides insight into an individual's capacity to adaptively navigate a changing environment (Cacioppo & Berntson, 2011), and baseline cardiac functioning is an established biomarker of emotion dysregulation and risk for psychopathology (Beauchaine, 2015). We use resting physiology data to calculate task-related physiological responding. Moreover, the synchronization of the physiological data and video recordings will allow us to model dynamic change in these systems across tasks (e.g., vagal flexibility) (Burt & Obradović, 2013; Wagner et al., 2023).

Autonomic Nervous System Functioning

Participants' electrocardiogram (ECG), electrodermal activity (EDA), cardiac output (NICO), and respiratory effort (RSP) will be collected throughout the visit using Biopac MP160 data acquisition and analysis systems with AcqKnowledge 5 software. Respiration will be collected using an elastic respiration band worn around the participant's chest. We will use two Wireless BioNomadix modules attached to the respiration belt to transmit ECG, NICO, and RSP. One BioNomadix transmitter is worn on the wrist to transmit EDA. We will apply 3 peel-and-stick hypoallergenic electrodes (EL512 1" round foam electrodes) to participants' chests to obtain ECG and 8 electrodes to obtain NICO (4 electrodes on the neck, with two on each side; 4 on the ribs, with two on each side). EDA signal will be obtained with 2 electrodes worn on the palm of the hand. Shielded lead wires will be used to connect electrodes to the transmitters. ECG will be used to derive respiratory sinus arrhythmia (RSA), while NICO data will be used to derive PEP (pre-ejection period), and EDA will be analyzed for skin conductance responses (SCRs). We will use Observer XT (Noldus Information Systems) to synchronize psychophysiology data to the millisecond with video recordings collected in Noldus MediaRecorder6.

Once collected, physiology data will be processed in AcqKnowledge 5 by down-sampling the data, applying a high- or low-pass filter to minimize "noise", and creating timestamps

marking the beginning and end of each experimental task. EDA and NICO artifacts will be edited in AcqKnowledge, while artifacts in ECG data will be edited with CardioEdit software. AcqKnowledge will be used to derive PEP from NICO data and SCRs from EDA. Task-specific RSA estimates will be calculated using CardioBatch software. RSA reactivity during tasks will be computed as an overall change from baseline (collected during 5 mins at the start of the visit when children sit and draw) and/or as a dynamic change across task epochs.

Stationary Eye-Tracking

Participants' eye movements will be measured using a desktop-mounted Eyelink 1000 Plus (SR Research Ltd., Ontario, Canada) during computer tasks designed to assess attentional processes relevant to threat and affiliation. Participants will place their heads on an adjustable tower mount chin rest and wear a bullseye calibration target sticker on their forehead to adjust for head movements. Stimuli will be presented on a desktop monitor approximately 60 cm away from the participant. All stimuli were created using SR Research Experiment Builder (SR Research Ltd., version 2.4.77). Stimuli are presented on a high refresh rate (i.e., 144 Hz) display, allowing for gaze-contingent research. Eye-tracking data are captured using the Eyelink 1000 Plus, which samples binocularly at up to 2000 Hz. Eyelink allows for accurate binocular/monocular, unstabilized tracking with an average accuracy of <0.5 degrees. For adult participants, a 9-point target display is used for calibration and validation of eye position, while a 5-point animated target display (i.e., an image of a spinning beach ball, lion, and star) is used for child calibration and validation. Data Viewer (SR Research Ltd., version 3.2) will be used to create static and dynamic interest areas and output variables such as dwell time, saccadic reaction time, and interest area reports.

Mobile Eye-Tracking

Mobile eye-tracking enables eyeglasses-based video recorders to register the eye gaze of participants within their visual environment. During tasks for which we collect mobile eye-

tracking data, participants will wear Pupil Invisible eye-tracking glasses (Pupil-Labs GmbH, Berlin, Germany). The glasses are equipped with two cameras (sampling frequency of 200Hz@ 192 × 192px) facing inward to capture the participants' eye and record pupil movement, and an outward-facing camera (60Hz@ 1280 × 720px) attached to the earpiece to record a video of the participant's "visual world". An adjustable elastic strap is attached to the back of the glasses to ensure that the glasses fit the participants' heads. A USB cable connects the glasses to a OnePlus 8T Android smartphone that runs the Pupil Invisible Companion app. The app creates a recording overlaying the participant's eye fixations with the "point-of-view" video of the participant's visual surroundings.

Phones will be placed in mesh fanny packs worn by the participants to allow ventilation during recording sessions. Participants will complete a calibration procedure at the beginning of the recording session by looking at five points on a paper bullseye calibration target, which is mounted to the wall approximately two feet away from the participant at their eye level. Participants repeat the calibration procedure once to perform validation. In addition, two drift checks will be performed, one at the end of each set of parent-child interaction activities.

Eye-tracking recordings will be uploaded to the Pupil Player app, where the eye-tracking footage, depicted by a set of circles and cross-hair lines, will be merged with the participant's visual world video. Manual gaze correction will be performed by adjusting the x and y coordinates of the eye gaze to align with the fixation points on the calibration target. The corrected recordings will be exported and coded for the duration of fixation on the toys (e.g., magnet tiles, picture book), objective within the testing room (e.g., fridge, rug, wall decal), and looks to the parent or child, with mutual eye-contact between parent-child dyads during interaction tasks as a variable of interest.

Behavioral Tasks

Behavioral, task-based data will be collected using a touchscreen or a computer via keyboard and mouse input. As with our questionnaires, task selection was guided by existing literature, information about the measures' validity and reliability, and our own pilot data. Tasks were built using Psychopy (Peirce, 2007) and Experiment Builder (SR Research Ltd., version 2.4.77). Other tasks, such as EF Touch and the NIH Toolbox Picture Vocabulary Test, are deployed using existing software (e.g., NIH Toolbox).

Parent Computer Tasks

Behavioral and visual attention bias to threat. To assess behavioral and visual attention to social and nonsocial threat cues in parents, we use an adapted version of a visual search paradigm previously validated in adults (LoBue, 2009; LoBue & DeLoache, 2008). During each trial, participants see a 3 by 3 grid of nine images and are instructed to touch or click on the image that is "different from the others as quickly as possible". During social trials, participants identify the angry, fearful, or happy face from among 8 neutral faces. During non-social trials, participants select the snake or frog from among 8 images of flowers. Accuracy and reaction time are recorded. Following one untimed scaffolding trial and two timed practice trials, adults complete 48 trials on the eye-tracking computer using a desktop monitor and mouse. Half the social trials are with child faces and half with adult faces. The outcome of interest is response time to threatening (angry and fearful faces; snakes) and non-threatening (happy faces and frogs) stimuli.

Emotion induction. To evaluate parental sensitivity to emotion, we use an emotion induction task, the Emotion Video Experience Ratings (EVER) task, which is a passive-viewing and rating task combined with the collection of physiological (e.g., heart rate, skin conductance) data that can give novel insight into the impact of videos that are expected to elicit different emotional responses in adults. During each trial, participants view a 1-minute neutral baseline

clip from a nature documentary overlaid with neutral music, followed by a 2-minute movie clip intended to induce a discrete emotion. The subjective emotional impact of the chosen clips was validated in a separate sample and, in some cases, in prior work (Dadds et al., 2016; Gross & Levenson, 1995; Schaefer et al., 2010). Target emotions are 2-minute clips depicting fear (*The Babadook*), happiness (*Mamma Mia!*), sadness (*The Champ*), and anger (*My Bodyguard*). Following each trial, participants rate how pleasant (i.e., valence) and stimulating (i.e., arousal) they found the clip on 9-point Likert scales guided by the Self-Assessment Manikin (Bradley & Lang, 1994). They also rate the greatest degree they experienced each of six discrete emotions (anger, fear, disgust, happiness/amusement, sadness, surprise) during the clip on a 9-point Likert scale. Physiological and eye-tracking (e.g., saccades, fixations) data will be collected.

Emotion recognition. To assess emotion recognition in parents, we use an adaptation of a task that assesses accuracy in recognizing and labeling facial expressions that convey specific emotions (Brislin & Patrick, 2019). Participants view a photo of an adult expressing a discrete emotion (750 ms) and use a computer mouse to select an emotion label (3 sec). Stimuli consist of images of adults expressing one of six discrete emotions (anger, disgust, fear, happiness, sadness, or surprise) at one of three intensity levels (33% emotional, 66% emotional, 100% emotional) based on morphing with a photo of the same actor displaying a calm expression. Images are from the full-color RADIATE face set (Conley et al., 2018) and comprise one male and one female actor from each of the following racial groups: East Asian, Black, Hispanic, and White. Participants are presented with one image from each actor, emotion, and emotional intensity combination (e.g., one trial of the East Asian female displaying 33% anger) in a randomized order. The primary outcome of interest is accuracy of emotion identification. Eye-tracking data will also be collected to assess eye activity (e.g., saccades, fixations).

*Reward learning***.** To assess reward learning in parents, we use an adapted version of a reward learning task presented on the touchscreen (Wimmer et al., 2018; Wimmer & Poldrack,

2022). During each trial, parents are presented with an image of one of four different colored jars, along with response options of an up and a down arrow. Two of the jars are more rewarding, and two of the jars are more punishing on average. When the up arrow is selected on reward trials, participants have an 80% chance of winning an average of 25 points and a 20% chance of losing an average of 5 points. Conversely, when participants select the up arrow on punishment trials, there is an 80% probability of losing an average of 25 points and a 20% probability of a neutral outcome (0 points on average). When the down arrow is selected on a reward trial, participants have an 80% chance of losing 5 points on average, and a 20% chance of gaining an average of 25 points. When participants select the down arrow on a punishment trial, there is an 80% probability of a neutral outcome (0 points on average) and 20% probability of losing an average of 25 points. Exact point amounts are jittered 5 points around the mean using a flat/universal distribution. As part of the task, there are three phases: learning, rating, and choice. The learning phase consists of 12 practice trials and 40 experimental trials. Participants learn that for half of the stimuli (the “reward” stimuli), it is better, on average, to press the “up” key, and for the other half (the “loss” stimuli), it is better, on average, to press the “down” key. Next is the rating phase, where participants rate how rewarding each stimulus was on a sliding scale, and then the choice phase, where participants are shown two stimuli and must choose which they believe was more rewarding. The primary variable of interest is the number of trials required for participants to learn which jars are associated with reward and which with punishment.

Child Computer Tasks

Behavioral and visual attention bias to threat. Children complete the same attention bias task as adults (i.e., using emotional faces and animals); this task was previously validated in our target age group (3-5 year-olds) (LoBue, 2009; LoBue & DeLoache, 2008). In contrast to the

adult version of the task, children complete 30 experimental trials on a touchscreen, and all social trials contain children's faces.

Emotion induction. To evaluate child sensitivity to emotion, we use a similar task to the EVER paradigm described above, but using developmentally appropriate videos from *The Lion King.* The subjective emotional impact of the chosen clips was validated in prior work (Dadds et al., 2016; Kimonis et al., 2023). Target emotions are in 2-minute clips depicting fear, happiness, sadness, and anger. The videos are passively viewed; there is no rating component to the child version of the task.

Emotion Recognition. Children's emotion recognition is assessed in a newly-developed touchscreen task separated into three parts. The first part of the task is the "evaluation" section, in which children see a grid of four different child faces, each showing a different emotion (happy, angry, sad, or scared). Children are asked how each face feels to get a baseline assessment of whether they can generate the correct emotion word. Next, as a "test", children see a grid of four new emotion faces and are prompted to touch the face that matches a given emotion label provided by the experimenter (e.g., "find the friend who feels happy"). The remaining two sections, "matching" and "receptive labeling," are presented in counterbalanced order. In the "matching" section (adapted from Székely et al., 2011), children are shown an emotion face (happy, angry, sad, or scared) at the top of the screen for two seconds. Then, while the top image remains on the screen, two images of a different child (of the same sex and race) are presented below—one displaying an emotion matching that of the image above, and one "contrast" emotion. Children have up to six seconds to select the new image that matches the emotion of the top image. In the "receptive labeling" section (adapted from Wu et al., 2023) children are presented with two images of the same child, one showing the target emotion, and the other showing a contrast emotion. A voiceover directs the child to select the face showing a certain emotion (happy, angry, sad, or scared). Children have 10 s to choose the correct image. Both the "matching" and "receptive labeling" sections have 24 test trials each— six of each

target emotion. The response options from the "matching" section are the same images used in the "receptive labeling" section, with an additional 24 face images (one per trial) used as the target stimuli in the "matching" section. All stimuli are from the Child Affective Facial Expression set (LoBue & Thrasher, 2015) and reflect equal proportions of male and female actors; they are 50% white, 42% Black, and 8% Asian, to roughly reflect our intended participant sample.

Executive functioning. We assess child executive functioning using two touchscreen tasks from "EF Touch," a battery of seven executive functioning and two non-executive functioning tasks developed specifically for use with 3- to 6-year-old children (Willoughby et al., 2017; Willoughby et al., 2013). The first task, Bubbles, assesses general reaction time (to be used as a covariate in analyses of other behavioral data). Children see blue circles (i.e., "bubbles") appear at different locations on a screen and are instructed to press the bubbles as they appear as fast as possible. Next, the Arrows task measures inhibitory control. Two green buttons appear on the bottom half of the screen. In each trial, an arrow pointing either left or right appears at the top left or right of the screen. Children are instructed to press the green button that the arrow is pointing to as fast as possible. During congruent trials, the arrow appears above the same button it is pointing to (e.g., the arrow appears above the right-hand button and is pointing to the right). During incongruent trials, the arrow appears above the opposite button from the one it is pointing to (e.g., the arrow appears above the right-hand button but is pointing to the left). For the incongruent trials, children must suppress the dominant response of pressing the button that the arrow is above. Children have 2 seconds to respond before the screen advances to the next trial. The primary outcome of interest is accuracy and reaction time on congruent and incongruent trials.

*Reinforcement learning to reward and punishment***.** To assess children's reinforcement learning to reward and punishment, we use an adapted touchscreen version of the Stars in Jars task (Briggs-Gowan et al., 2014). During each trial, children see an illustration of one of four jars of different colors and shapes appear on the touchscreen monitor. Two jars always have stars in

them, and two are always empty. Children learn that if they touch a jar with stars in it, they win points, and if they refrain from touching an empty jar, they also win points. If they touch an empty jar or do not touch a jar with stars in it, they lose points. Thus, children learn to click on the "rewarding" jars and avoid clicking on the "punishment" jars. The task begins with scaffolding, in which the child and experimenter explore if practice jars have stars inside. Then, the child completes up to 5 practice rounds before advancing to the experimental trials, of which there are 36. In both the practice and experimental trials, there is a time limit of 10 seconds.

Social motivation and preference. Children who display similar social behaviors may vary in their affiliative preferences and motivations (Kopala-Sibley & Klein, 2017). To assess children's affiliative and instrumental motivations and preferences for engaging in social behavior, we developed a new digital storybook task, the Child Affiliative Motivations and Preferences (CAMP) task, which is presented on the touchscreen. The task presents drawings of cartoon bears that represent peer-directed social contexts across three types of play (active, associative, and collaborative). The child is told that the drawings of cartoon bears depict the child and their peers. To emphasize that the bear represents the child, children personalize the image of the bear at the beginning of the task by selecting a hat or bow for the bear to wear. In each trial, the child hears a voiceover and sees a drawing on the left side of the screen introducing the setting (e.g., "It's time to play in the sandbox", along with a drawing of a bear in a sandbox) and two drawings on the right side of the screen, one at the top and one at the bottom. The experimenter asks the child which of the two settings they would pick. One of the drawings represents a social setting (e.g., a bear playing with a peer in the sandbox), and the other a less social setting (e.g., a bear playing alone in the sandbox). Once the child selects a response by pressing on the drawing, the task moves on to the next trial. Each type of play is presented across 3 contexts in a random order for a total of 18 trials. Examples of active contexts include playing hopscotch and ball, examples of associative contexts include eating lunch or walking home, and examples of collaborative contexts include gardening and doing a puzzle. The child

is asked, “which one would you pick?” between two options depicting (a) a group of peers vs. one peer, (b) a group of peers vs. playing alone, or (c) one peer vs. playing alone. The selections represent what the child prefers to do in a social situation devoid of social confounds, such as fear of rejection.

Theory of mind. To assess children’s theory of mind, including its cognitive and affective aspects, we use an adapted touchscreen version of the CAToon Task, which was originally developed and validated in an MRI study with 3- to 9-year-olds (Borbás et al., 2021). Our adapted task consists of 24 hand-drawn stories, with 8 stories selected from each of three conditions: affective theory of mind (AT), cognitive theory of mind (CT), and a control condition of physical cause and effect (PC). During AT trials, participants infer how a character would respond to another character’s expressed or inferred emotions. During CT trials, participants choose how a character would behave based on another character’s intentions or beliefs. Finally, PC trials depict basic laws of physics and serve as a control to demonstrate that the participant understands cause and effect. The task begins with one scaffolding trial, followed by 2 practice trials, and 18 experimental trials. The order of the stories is randomized. In each trial, three images from a story are presented one after another on the touch screen. Each image is presented for 3 seconds. Then, three new images are presented on the screen. Participants are instructed to select the ending that “best fits the story” by touching the corresponding image on the screen. CT and PC trial endings consist of one possible, one improbable, and one impossible solution. AT trials have two possible solutions (one positive expectancy and one negative expectancy outcome) and one impossible solution. Accuracy for each condition and positive/negative expectancy for the AT condition are the primary variables of interest.

Verbal ability. We will assess children’s verbal ability using the NIH Toolbox Picture Vocabulary Test, a computerized measure of receptive vocabulary developed to assess auditory comprehension of single words in participants ages three and older (Gershon et al., 2014). On each trial, children are presented with an audio recording of a word and 4

photographic images on an iPad screen. The child is asked to touch the picture that matches the word's meaning. Once the participant completes the first trial, the program selects the subsequent trial based on the accuracy of the previous response. Successive trials are selected using an updated estimate of the participant's proficiency, such that the task continues until participants' standard error of performance falls below 0.3. The child's performance on the task will be used to index verbal ability as a control variable for other tasks (e.g., receptive emotion recognition task).

Observational Tasks

Procedures

Children will complete one interaction task with a stranger to evaluate behavioral inhibition, and parents and children will engage in three interaction tasks to evaluate parental scaffolding and parent-child dyadic quality. The parent and child will sit at a small table sitting on adjacent sides of the table to each other (i.e., both in view of cameras and able to make eye-contact with one another). Cardiac physiology (ECG, CO and RSP) and mobile eye-tracking data also will be collected throughout the tasks.

*Book reading task***.** The parent and child look at a wordless book together for four minutes. The book's illustrations depict scenes that elicit both positive (e.g., a child welcoming a new classmate) and negative emotions (e.g., bullying).

Magnet task. The magnet task is a puzzle game intended to assess parental scaffolding. The experimenter will give the participants a box of magnet tiles and a binder with images of different shapes made from the magnet tiles (e.g., a star made by joining a pentagon and five triangles; a cube created by joining together 6 square magnets). Participants are asked to start on the first page of the binder and create as many of the shapes as they can in four minutes. The experimenter informs the parent that the task is primarily for the child, but that the parent may provide as much help as they think their child needs.

Conversation task. The parent and child are asked to face one another and discuss three different topics. First, they will be asked to discuss a negative (i.e., "sad, scary, or angry") experience they had together. Then, they are asked to discuss a positive (i.e., "silly, funny, or happy") experience they had together. Finally, the parent and child are asked to take turns talking about the reasons they love each other. The experimenter leaves the room during the task and knocks every two minutes to indicate when the participants should switch to the next prompt. This task assesses parent-child dyadic quality.

Behavioral inhibition. We assess children's behavioral inhibition through a stranger approach and robot task, which evaluates children's behaviors in response to a novel social interaction (approach by an unfamiliar adult) and a novel non-social object (a dancing toy robot). During the task, the child engages in 1 min of independent play with toys (a Lego Duplo set with ocean animals) in the middle of the room. Parents are instructed to sit in a chair in the corner and avoid interacting with their child. After 1 min, an unfamiliar adult research assistant enters the room, attempts to engage the child in conversation, and offers a new toy (a Lego toy whale belonging to the Duplo set) to the child. The stranger will follow a standardized script, speaking to the child in a neutral tone and wearing standardized attire (a black hat, black t-shirt, and black disposable face mask). The interaction with the "stranger" lasts 2 min, after which the child will be allowed to return to independent play for a minute. Next, the same research assistant will enter the room with a toy robot. The robot dances and plays loud music. The research assistant turns the toy on and leaves. After 1 min, the research assistant returns to the room and engages the child in conversation about the robot, encouraging the child to touch or hold it. The child will not be required to touch the robot if they do not want to.

Rating Behavior

Parent-child interaction tasks. To derive observational scores from parent-child interactions, we will rate behavior using established coding schemes, which have been

successfully deployed in numerous longitudinal studies including The Study of Early Childcare and Youth Development and the Family Life Project/ECHO Project (Vernon-Feagans et al., 2013; Vernon-Feagans et al., 2008). Parent, child, and dyadic constructs, such as sensitivity, warmth, intrusion, detachment, positive affect, negative affect, boundary dissolution, and dyadic mutuality, will be derived from these coding schemes. Behavior composites broadly indexing sensitive/responsive (positive) and harsh/controlling (negative) parenting have been established using these coding schemes in multiple studies (Barnett et al., 2008; Vernon-Feagans et al., 2013), and their predictive validity has been established in diverse samples (Brown et al., 2017; Clincy & Mills-Koonce, 2013; Holochwost et al., 2020). Video recordings of the stranger and robot task will be coded in Observer 16 XT by at least two independent raters. Emotion scaffolding, including emotion language use and mental-state language, will be derived from transcripts using the Linguistic Inquiry and Word Count (LIWC) software (Boyd et al., 2022), as well as language style matching, a linguistic metric of dyadic coregulation (Gonzales et al., 2010). The System for Coding Affiliation in Lab Assessments (SCALA) will quantify children's affiliative tenor, including attention towards and awareness of the parent, their social communication, warmth and proximity-seeking, and overall affiliative initiation and enjoyment from the interaction. The SCALA is a newly developed, RDoC-informed coding scheme, designed to assess early individual differences in affiliation as observed in social interactions. Finally, parent-child dyadic synchrony will be assessed using linguistic analysis of transcripts to identify language style concordance between the parent and child, as well as physiological measures of RSA synchrony and coding of mutual eye gaze based on the mobile eye-tracking recordings.

Behavioral inhibition tasks. Observational coding of child responses during the stranger and robot tasks has been reliably used to assess behavioral inhibition in early childhood (Buss, 2011; Buss et al., 2013; Fox et al., 2001; White et al., 2011). Video recordings of the stranger and robot task will be coded in Observer 16 XT by at least two independent raters. The primary

variables of interest will be body and facial fear (i.e., slowed rate of play, bodily tension, freezing, and facial expressions), latency to approach and touch the target stimuli (i.e., the toy whale and robot), and proximity to the caregiver and targets (i.e., the stranger and robot). Physiological data (ECG, NICO, and respiration) will be collected for the parent and child throughout the two tasks.

References

Abidin, R. (1995). Parenting Stress Index. *Psychological Assessment Resources*, Odessa, FL.

Achenbach, T., & Rescorla, L. (2000). *Manual for the ASEBA Preschool Forms & Profiles* University of Vermont, Research Center for Children, Youth, and Families.

Arnold, D. S., O'leary, S. G., Wolff, L. S., & Acker, M. M. (1993). The Parenting Scale: a measure of dysfunctional parenting in discipline situations. *Psychological assessment*, *5*(2), 137.

Ashwood, K., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F., Findon, J., Eklund, H., Spain, D., & Wilson, C. (2016). Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychological Medicine*, *46*(12), 2595-2604.

Barnett, M. A., Deng, M., Mills-Koonce, W. R., Willoughby, M., & Cox, M. (2008). Interdependence of parenting of mothers and fathers of infants. *Journal of Family Psychology*, *22*(4), 561.

Beauchaine, T. P. (2015). Respiratory sinus arrhythmia: A transdiagnostic biomarker of emotion dysregulation and psychopathology. *Current opinion in psychology*, *3*, 43-47.

Borbás, R., Fehlbaum, L. V., Rudin, U., Stadler, C., & Raschle, N. M. (2021). Neural correlates of theory of mind in children and adults using CAToon: Introducing an open-source child-friendly neuroimaging task. *Developmental Cognitive Neuroscience*, *49*, 100959.

Boyd, R. L., Ashokkumar, A., Seraj, S., & Pennebaker, J. W. (2022). The development and psychometric properties of LIWC-22. *Austin, TX: University of Texas at Austin*, 1-47.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry*, *25*(1), 49-59.

Briggs-Gowan, M. J., Nichols, S. R., Voss, J., Zobel, E., Carter, A. S., McCarthy, K. J., Pine, D. S., Blair, J., & Wakschlag, L. S. (2014). Punishment insensitivity and impaired reinforcement learning in preschoolers. *Journal of child Psychology and Psychiatry*, *55*(2), 154-161.

Brislin, S. J., & Patrick, C. J. (2019). Callousness and affective face processing: Clarifying the neural basis of behavioral-recognition deficits through the use of brain event-related potentials. *Clinical Psychological Science*, *7*(6), 1389-1402.

Brown, G. L., Gustafsson, H. C., Mills-Koonce, W. R., & Cox, M. J. (2017). Associations between early caregiving and rural, low-SES, African-American children's representations of attachment relationships. *Attachment & human development*, *19*(4), 340-363.

Burt, K. B., & Obradović, J. (2013). The construct of psychophysiological reactivity: Statistical and psychometric issues. *Developmental Review*, *33*(1), 29-57.

Buss, K. A. (2011). Which fearful toddlers should we worry about? Context, fear regulation, and anxiety risk. *Developmental psychology*, *47*(3), 804.

Buss, K. A., Davis, E. L., Kiel, E. J., Brooker, R. J., Beekman, C., & Early, M. C. (2013). Dysregulated fear predicts social wariness and social anxiety symptoms during kindergarten. *Journal of Clinical Child & Adolescent Psychology*, *42*(5), 603-616.

Cacioppo, J. T., & Berntson, G. G. (2011). The brain, homeostasis, and health: Balancing demands of the internal and external milieu. *The Oxford handbook of health psychology*, 121-137.

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

Clincy, A. R., & Mills-Koonce, W. R. (2013). Trajectories of intrusive parenting during infancy and toddlerhood as predictors of rural, low-income African American boys' school-related outcomes. *American Journal of Orthopsychiatry*, *83*(2-3), 194.

Colins, O. F., Andershed, H., Frogner, L., Lopez-Romero, L., Veen, V., & Andershed, A.-K. (2014). A new measure to assess psychopathic personality in children: The Child Problematic Traits Inventory. *Journal of psychopathology and behavioral assessment*, *36*(1), 4-21.

Conley, M. I., Dellarco, D. V., Rubien-Thomas, E., Cohen, A. O., Cervera, A., Tottenham, N., & Casey, B. (2018). The racially diverse affective expression (RADIATE) face stimulus set. *Psychiatry research*, *270*, 1059-1067.

Coplan, R. J., Prakash, K., O'Neil, K., & Armer, M. (2004). Do you" want" to play? Distinguishing between conflicted shyness and social disinterest in early childhood. *Developmental psychology*, *40*(2), 244.

Crnic, K. A., & Booth, C. L. (1991). Mothers' and fathers' perceptions of daily hassles of parenting across early childhood. *Journal of Marriage and the Family*, 1042-1050.

Dadds, M. R., Gale, N., Godbee, M., Moul, C., Pasalich, D. S., Fink, E., & Hawes, D. J. (2016). Expression and regulation of attachment-related emotions in children with conduct problems and callous–unemotional traits. *Child Psychiatry & Human Development*, *47*(4), 647-656.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment*, *18*(2), 192.

Felitti, V. J., Anda, R. F., Nordenberg, D., Williamson, D. F., Spitz, A. M., Edwards, V., & Marks, J. S. (1998). Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: The Adverse Childhood Experiences (ACE) Study. *American journal of preventive medicine*, *14*(4), 245-258.

Fox, N. A., Henderson, H. A., Rubin, K. H., Calkins, S. D., & Schmidt, L. A. (2001). Continuity and discontinuity of behavioral inhibition and exuberance: Psychophysiological and behavioral influences across the first four years of life. *Child development*, *72*(1), 1-21.

Gershon, R. C., Cook, K. F., Mungas, D., Manly, J. J., Slotkin, J., Beaumont, J. L., & Weintraub, S. (2014). Language measures of the NIH toolbox cognition battery. *Journal of the International Neuropsychological Society*, *20*(6), 642-651.

Gonzales, A. L., Hancock, J. T., & Pennebaker, J. W. (2010). Language style matching as a predictor of social dynamics in small groups. *Communication Research*, *37*(1), 3-19.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British journal of psychiatry*, *177*(6), 534-539.

Gordts, S., Uzieblo, K., Neumann, C., Van den Bussche, E., & Rossi, G. (2017). Validity of the Self-Report Psychopathy Scales (SRP-III full and short versions) in a community sample. *Assessment*, *24*(3), 308-325.

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, *26*, 41-54.

Gross, J. J., & Levenson, R. W. (1995). Emotion elicitation using films. *Cognition & emotion*, *9*(1), 87-108.

Halberstadt, A. G., Dunsmore, J. C., Bryant Jr, A., Parker, A. E., Beale, K. S., & Thompson, J. A. (2013). Development and validation of the Parents' Beliefs About Children's Emotions Questionnaire. *Psychological assessment*, *25*(4), 1195.

Haskett, M. E., Ahern, L. S., Ward, C. S., & Allaire, J. C. (2006). Factor structure and validity of the parenting stress index-short form. *Journal of Clinical Child & Adolescent Psychology*, *35*(2), 302-312.

Holochwost, S. J., Volpe, V. V., Iruka, I. U., & Mills-Koonce, W. R. (2020). Maternal warmth, intrusiveness, and executive functions in early childhood: tracing developmental

processes among African American children. *Early Child Development and Care*, *190*(2), 210-218.

Kazak, A. E., Alderfer, M., Enlow, P. T., Lewis, A. M., Vega, G., Barakat, L., Kassam-Adams, N., Pai, A., Canter, K. S., & Hildenbrand, A. K. (2021). COVID-19 exposure and family impact scales: factor structure and initial psychometrics. *Journal of pediatric psychology*, *46*(5), 504-513.

Kim, J., Klein, D. N., Olino, T. M., Dyson, M. W., Dougherty, L. R., & Durbin, C. E. (2011). Psychometric properties of the Behavioral Inhibition Questionnaire in preschool children. *Journal of personality assessment*, *93*(6), 545-555.

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.

Kimonis, E. R., Frick, P. J., Skeem, J. L., Marsee, M. A., Cruise, K., Munoz, L. C., Aucoin, K. J., & Morris, A. S. (2008). Assessing callous–unemotional traits in adolescent offenders: Validation of the Inventory of Callous–Unemotional Traits. *International journal of law and psychiatry*, *31*(3), 241-252.

Kimonis, E. R., Le, B., Fleming, G. E., Kyranides, M. N., Demetriou, C. A., Fanti, K. A., Neo, B., Prasad, A. H., Chan, A., & Hawes, D. J. (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*, *64*(3), 357-366.

Kopala-Sibley, D. C., & Klein, D. N. (2017). Distinguishing types of social withdrawal in children: Internalizing and externalizing outcomes of conflicted shyness versus social disinterest across childhood. *Journal of Research in Personality*, *67*, 27-35.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of general internal medicine*, *16*(9), 606-613.

LoBue, V. (2009). More than just another face in the crowd: Superior detection of threatening facial expressions in children and adults. *Developmental science*, *12*(2), 305-313.

LoBue, V., & DeLoache, J. S. (2008). Detecting the snake in the grass: Attention to fear-relevant stimuli by adults and young children. *Psychological science*, *19*(3), 284-289.

LoBue, V., & Thrasher, C. (2015). The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults. *Frontiers in psychology*, *5*, 1532.

Löwe, B., Decker, O., Müller, S., Brähler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. *Medical care*, 266-274.

Matheny Jr, A. P., Wachs, T. D., Ludwig, J. L., & Phillips, K. (1995). Bringing order out of chaos: Psychometric characteristics of the confusion, hubbub, and order scale. *Journal of applied developmental psychology*, *16*(3), 429-444.

Mindell, J. A., Gould, R. A., Tikotzy, L., Leichman, E. S., & Walters, R. M. (2019). Norm-referenced scoring system for the brief infant sleep questionnaire–revised (BISQ-R). *Sleep medicine*, *63*, 106-114.

Neumann, C. S., & Pardini, D. (2014). Factor structure and construct validity of the Self-Report Psychopathy (SRP) Scale and the Youth Psychopathic Traits Inventory (YPI) in young men. *Journal of personality disorders*, *28*(3), 419.

Pascoe, J., Ialongo, N., Horn, W., Reinhart, M., & Perradatto, D. (1988). The reliability and validity of the maternal social support index. *Family medicine*, *20*(4), 271-276.

Patridge, E. F., & Bardyn, T. P. (2018). Research electronic data capture (REDCap). *Journal of the Medical Library Association: JMLA*, *106*(1), 142.

Paz, Y., Wagner, N., & Waller, R. (2023). *The Components of Affiliative Reward Experiences Scale (CARES)*. University of Pennsylvania

Peirce, J. W. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1-2), 8-13.

Perlstein, S., Wagner, N., Domínguez-Álvarez, B., Gómez-Fraguela, J. A., Romero, E., Lopez-Romero, L., & Waller, R. (2022). Psychometric Properties, Factor Structure, and Validity of the Sensitivity to Threat and Affiliative Reward Scale in Children and Adults. *Assessment*, 10731911221128946.

Plomin, R., & Rowe, D. C. (1977). A twin study of temperament in young children. *The Journal of Psychology*, *97*(1), 107-113.

Schaefer, A., Nils, F., Sanchez, X., & Philippot, P. (2010). Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and emotion*, *24*(7), 1153-1172.

Shaw, D. S., Criss, M. M., Schonberg, M. A., & Beck, J. E. (2004). The development of family hierarchies and their relation to children's conduct problems. *Development and psychopathology*, *16*(3), 483-500.

Shaw, D. S., Winslow, E. B., Owens, E. B., & Hood, N. (1998). Young children's adjustment to chronic family adversity: A longitudinal study of low-income families. *Journal of the American Academy of Child & Adolescent Psychiatry*, *37*(5), 545-553.

Straus, M. A., & Douglas, E. M. (2004). A short form of the Revised Conflict Tactics Scales, and typologies for severity and mutuality. *Violence and victims*, *19*(5), 507-520.

Sullivan, M. C., Strainge, L., Blackmon, J. E., Cruess, S. E., Wheeler, D., & Cruess, D. G. (2020). Assessing an epidemic: Utility of the diagnostic and statistical manual of mental disorders, level 2 substance use screener in adult psychiatric inpatients. *Journal of Addictions Nursing*, *31*(1), 9-16.

Székely, E., Tiemeier, H., Arends, L. R., Jaddoe, V. W., Hofman, A., Verhulst, F. C., & Herba, C. M. (2011). Recognition of facial expressions of emotions by 3-year-olds. *Emotion*, *11*(2), 425.

Uljarević, M., Frazier, T. W., Phillips, J. M., Jo, B., Littlefield, S., & Hardan, A. Y. (2019). Mapping the research domain criteria social processes constructs to the social

responsiveness scale. *Journal of the American Academy of Child & Adolescent Psychiatry*.

Verhoeven, M., Deković, M., Bodden, D., & van Baar, A. L. (2017). Development and initial validation of the comprehensive early childhood parenting questionnaire (CECPAQ) for parents of 1–4 year-olds. *European Journal of Developmental Psychology*, *14*(2), 233-247.

Vernon-Feagans, L., Cox, M., Willoughby, M., Burchinal, M., Garrett-Peters, P., Mills-Koonce, R., Garrett-Peiers, P., Conger, R. D., & Bauer, P. J. (2013). The Family Life Project: An epidemiological and developmental study of young children living in poor rural communities. *Monographs of the Society for Research in Child Development*, i-150.

Vernon-Feagans, L., Pancsofar, N., Willoughby, M., Odom, E., Quade, A., Cox, M., & Investigators, F. L. K. (2008). Predictors of maternal language to infants during a picture book task in the home: Family SES, child characteristics and the parenting environment. *Journal of Applied Developmental Psychology*, *29*(3), 213-226.

Wagner, N. J., Shakiba, N., Bui, H. N., Sem, K., Novick, D. R., Danko, C. M., Dougherty, L. R., Chronis-Tuscano, A., & Rubin, K. H. (2023). Examining the relations between children's vagal flexibility across social stressor tasks and parent-and clinician-rated anxiety using baseline data from an early intervention for inhibited preschoolers. *Research on Child and Adolescent Psychopathology*, 1-12.

White, L. K., McDermott, J. M., Degnan, K. A., Henderson, H. A., & Fox, N. A. (2011). Behavioral inhibition and anxiety: The moderating roles of inhibitory control and attention shifting. *Journal of abnormal child psychology*, *39*, 735-747.

Willoughby, M. T., Kuhn, L. J., Blair, C. B., Samek, A., & List, J. A. (2017). The test–retest reliability of the latent construct of executive function depends on whether tasks are represented as formative or reflective indicators. *Child Neuropsychology*, *23*(7), 822-837.

Willoughby, M. T., Pek, J., & Blair, C. B. (2013). Measuring executive function in early childhood: a focus on maximal reliability and the derivation of short forms. *Psychological assessment*, *25*(2), 664.

Wimmer, G. E., Li, J. K., Gorgolewski, K. J., & Poldrack, R. A. (2018). Reward learning over weeks versus minutes increases the neural representation of value in the human brain. *Journal of Neuroscience*, *38*(35), 7649-7666.

Wimmer, G. E., & Poldrack, R. A. (2022). Reward learning and working memory: Effects of massed versus spaced training and post-learning delay period. *Memory & Cognition*, *50*(2), 312-324.

Wu, Y., Matteson, H. M., Baker, C., & Frank, M. C. (2023). Angry, sad, or scared? Within-valence mapping of emotion words to facial and body cues in 2-to 4-year old children. *Collabra: Psychology*, *9*, 74333.

Ye, M., Hessler, D., Ford, D., Benson, M., Koita, K., Bucci, M., Long, D., Harris, N. B., & Thakur, N. (2023). Pediatric ACEs and related life event screener (PEARLS) latent domains and child health in a safety-net primary care practice. *BMC pediatrics*, *23*(1), 1-12.