

Children's Facial Muscular Movements and Risk for Early Psychopathology: Assessing Clinical Utility

Adam S. Grabell

University of Massachusetts, Amherst

Hannah M. Jones

Aimee E. Wilett

Lisa M. Bemis

University of Pittsburgh School of Medicine

Lauren S. Wakschlag

Northwestern University Feinberg School of Medicine

Susan B. Perlman

Washington University School of Medicine in St. Louis

Standardized developmentally based assessment systems have transformed the capacity to identify transdiagnostic behavioral markers of mental disorder risk in early childhood, notably, clinically significant irritability and externalizing behaviors. However, behavior-based instruments that both differentiate risk for persistent psychopathology from normative misbehavior, and are

feasible for community clinicians to implement, are in nascent phases of development. Young children's facial expressions during frustration challenges may form the basis for novel assessments tools that are flexible, quick, and easy to implement as markers of psychopathology to complement validated questionnaires. However, the accuracy of facial expressions to correctly classify young children falling above and below clinical cut-offs is unknown. Our goal was to test how facial expressions during frustration, defined by different facial muscular movements, related to individual differences in irritability and externalizing behaviors and discriminated children with clinically significant levels from peers. Participants were 79 children (ages 3–7) who completed a short, moderately frustrating computer task while facial expressions were recorded. Only negative facial expressions that included eye constriction related to irritability and externalizing behaviors and were clinically discriminating. Moreover, these expressions significantly discriminated children with and without clinically significant irritability and externalizing symptoms with high Area Under the Curve (AUC) values ($> .75$) indicating good clinical utility. In contrast, expressions without eye constriction showed no clinical utility. The presence of negative expressions with eye constriction in response to a short frustration prompt may serve as an indicator of early psychopathology, raising the potential for novel

Work presented in this manuscript was supported by NIMH grants K01 MH094467 PI: Perlman, R21 MH100189 PI: Perlman, R01 MH107540 PI: Perlman, K23 MH111708 PI: Grabell, & R01MH107652 PI: Wakschlag.

The Laboratory for Child Brain Development (www.lcbd.pitt.edu) wishes to thank the children and families who participated in this study. We also thank Caroline MacGillivray, BA, Meghan Murphy, BA, Brianna Jones, BA, Min Sun, BA, Stephanie Buerk, BA, and Chiamaka Ukoha, BA for their help in recruitment, data collection, and facial expression coding.

Address correspondences to Dr. Adam Grabell, Self-Regulation, Emotions, and Early Development Laboratory, University of Massachusetts, Amherst, Department of Psychological and Brain Sciences, Tobin Hall, 135 Hicks Way, Amherst, MA, 01003; e-mail: grabell@umass.edu.

assessment tools that may enhance precision of early identification.

Keywords: child; facial expression; psychopathology; ROC curve; risk assessment

EARLY CHILDHOOD IS A developmental period when indicators of risk for persistent psychopathology are especially difficult to differentiate from challenging yet developmentally normative variants (Wakschlag, Tolan, & Leventhal, 2010). This is particularly true for early irritability and externalizing behaviors, the most common reasons children are referred for clinical services (Egger & Angold, 2006; Stringaris, 2011; Tremblay et al., 2004; Wakschlag et al., 2012). Irritability comprises annoyance, touchiness, angry mood, and temper outbursts (Wakschlag et al., 2017) and is a feature of over a dozen DSM-5 disorders (Avenevoli, Blader, & Leibenluft, 2015), including disorders often diagnosed in youth such as disruptive mood dysregulation disorder, oppositional-defiant disorder (ODD), generalized anxiety disorder, and major depressive disorder (American Psychiatric Association, 2013). Externalizing behaviors are aggression, defiance, and violation of social norms and typically comprises the early childhood disorders ADHD, ODD, and conduct disorder (Bongers, Koot, Van der Ende, & Verhulst, 2003). While ODD can include irritability in its presentation, it does not require it, and thus irritability is a transdiagnostic symptom of poor frustration tolerance distinct from the broader construct of externalizing problems. However, these behaviors closely resemble elevated yet developmentally normative misbehaviors exhibited in early childhood when emotion regulation is nascent (Wakschlag et al., 2015; Wakschlag et al., 2010). Behavior observation systems, in which clinicians tune in to specific overt behaviors during standardized presses, could potentially assist in detecting clinically concerning levels of early irritability and externalizing behavior given that, in domains such as autism, these tools show impressive diagnostic accuracy (Risi et al., 2006). However, behavior observation-based tools designed to identify clinical irritability and externalizing behavior do not presently exist in forms that could be deployed in wide-ranging community-based settings. This is, in part, due to a paucity of research on very basic overt behaviors that show clinical utility to discriminate between severe vs. normative irritability and externalizing behaviors. In the present study, we examined children's facial expressions

during a 10-minute mildly frustrating computer game and tested the clinical utility of specific facial muscular movement combinations to differentiate children falling above and below clinical cut-offs for irritability and externalizing behavior problems.

Clinicians are increasingly relying on behavior observation-based assessment tools, paired with validated rating scales, with the Autism Diagnostic Observation Schedule (ADOS-2), the most widely disseminated example (Lord et al., 2000; Risi et al., 2006). The use of these behavior assessment systems highlights both the importance of identifying overt behaviors with clinical utility and developing variants of these tools that can be used in different clinical settings in conjunction with rating scale instruments. Consider the millions of children who make initial contact with mental health care services each year (Bitsko, 2016). For example, a parent may decide that their concerns warrant evaluating their child through a local community clinic, or a school psychologist may initially observe a child in the classroom before an office assessment is performed. Community clinicians are often charged with deciding whether irritability and externalizing symptom severity falls within the clinical range, and whether services should be provided or referrals made (Mash & Hunsley, 2005). However, large-scale surveys reveal that community clinicians overwhelmingly rely on their qualitative impressions of behavior and almost never use standardized observation tools (Addis & Krasnow, 2000). Other surveys find that standardized observation tools are valued by child clinicians but are avoided due to practical concerns related to the length, cost, and training requirements needed to implement them in their specific setting (Jensen-Doss & Hawley, 2010; Whiteside, Sattler, Hathaway, & Douglas, 2016). Thus, there is a critical need for more efficient, flexible, standardized observation tools that can be delivered, with a low training burden, across clinical settings and combined with caregiver report instruments. Developing these tools requires investigating the clinical utility of basic overt child behaviors that can be prompted, captured, and assayed in short spans of time.

Decades of research have shown that frequency and intensity of young children's negative facial expressions are associated with their level of irritability and externalizing problems (Southam-Gerow & Kendall, 2002) and can be captured using simple lab-based paradigms or brief naturalistic observations (Breau et al., 2001; Cole, 1986). Preschool- and kindergarten-age children who more often produced expressions of sadness or anger, such as frowning or scowling, during a

blocked goal (Cole, Zahn-Waxler, Fox, Usher, & Welsh, 1996), negative video clip (Cole, Zahn-Waxler, & Smith, 1994), or unstructured classroom interactions (Hanish et al., 2004), had higher parent- and teacher-rated psychopathology symptoms than peers. Moreover, facial expressions are a nonverbal behavior clinicians are trained to notice and describe in their documentation (Kamphaus & Frick, 2005), are innately recognizable by untrained humans (Ekman et al., 1987), can be recorded anywhere, and are the increasing focus of automated recognition efforts (Ryan et al., 2009). Thus, negative facial expressions are a promising candidate for lightweight, next-generation standardized observation tools that can assist in detecting early-onset mental illness in wide-ranging clinical settings. However, while there is robust evidence that facial expressions relate to individual differences in symptoms, there is a paucity of studies gauging the *clinical utility* of facial expressions. That is, to our knowledge, there have been no investigations testing if facial expressions can reliably discriminate children with clinically significant irritability and externalizing problems from peers with a high degree of accuracy. Moreover, this clinical utility may depend on how specific facial muscles contract to form emotional expressions.

Facial coding systems have advanced to the point where the contraction of individual facial muscles can be reliably identified by trained coders (Cohn, Ambadar, & Ekman, 2007). The most widely used anatomically based coding system, the Facial Action Coding System (FACS) (Ekman & Friesen, 1977), has been successfully taught to thousands of researchers in laboratory settings (Ekman & Rosenberg, 1997). The ability to reliably code individual muscle movements has allowed researchers to investigate which specific muscle movements most accurately signal emotion intensity and underlying regulatory processes. Much of this work has focused on the contraction of orbicularis oculi, or eye constriction, the outer ring of muscles around the eye, which has been hypothesized to convey more intense emotional salience (Duchenne, 1990; Mattson, Cohn, Mahoor, Gangi, & Messinger, 2013). Messinger and colleagues found eye constriction was associated with stronger smile and cry faces in infants that were rated as more emotionally intense by independent observers (2012). More recently, we found that 4- to 6-year-old children who produced more frequent negative expressions with eye constriction, but not without eye constriction, had weaker concurrent lateral prefrontal cortex (LPFC) activation and were rated by caregivers as having more difficulty recovering from emotional challenges.

Children's negative expressions with eye constriction may therefore signal individual differences in the development of critical regulatory abilities underpinning externalizing and irritability problems, raising the possibility that eye constriction can discriminate clinical vs. normative levels of symptoms. Indeed, researchers examining baseball card photographs and social media posts found that adults whose smiles included eye constriction had better social relationships, life satisfaction, and longer life expectancy compared to adults whose smiles lacked eye constriction (Abel & Kruger, 2010; Harker & Keltner, 2001; Seder & Oishi, 2012). These findings suggest even brief instances of eye constriction in emotional expressions may auger long-term physical health, well-being, and psychological outcomes.

The present study examined 79 3- to 7-year-old children who completed a short, well-validated and developmentally sensitive frustration task while facial expressions were recorded and coded via human raters. Caregivers rated their children's irritability and externalizing symptoms using normed scales prior to the task. We hypothesized that negative expressions with eye constriction, but not without eye constriction, during frustration, would (a) relate to continuous externalizing and irritability scores; and (b) accurately discriminate children above the clinical cutoff for externalizing and irritability scales. We also examined positive expressions with and without eye constriction, during winning and frustration events, for the purposes of assessing whether effects were specific to negative affect vs. overall expressivity, and therefore did not have a directional hypothesis regarding positive expressions.

Methods

PARTICIPANTS

We collected data on 85 children between 3 and 7 years recruited from the community using paper and internet advertisements (this comprises 65 children previously reported in Grabel et al., 2018, and 20 additional children). Children were reported by their parents, via a phone screener, to have no psychiatric diagnoses and no first-degree relative with a severe psychiatric diagnosis (e.g., psychosis). However, level of irritability and externalizing symptoms were allowed to freely vary. One child declined to play the frustration task and 5 children were excluded due to technical problems (camera did not record correctly, video file was corrupted, or child moved out of frame). The final sample therefore comprised 79 children ($M = 67.3$ months, $SD = 15.6$), who were 52% male, 69% Caucasian, 27% African American, and

4% Asian. Children came from families ranging from very low to very high yearly incomes ($M = \$93,373$ per year, $SD = \$149,869$; range = \$10,000–\$1,250,000 per year). All children completed the Peabody Picture Vocabulary Test (Dunn & Dunn, 2007) and in general performed in the average range with scores ranging from low average to very superior ($M = 113$, $SD = 14.8$; range = 80–155). Experimental procedures were approved by the local Institutional Review Board.

PARENT-REPORTED PSYCHOPATHOLOGY SYMPTOMS

Dimensional Spectrum of Irritability

Parents rated their child's irritability using the Temper Loss subscale from the Multidimensional Assessment of Profile of Disruptive Behavior (MAP-DB; Wakschlag et al., 2014). The MAP-DB was designed to differentiate normal:abnormal irritability in early childhood (Camacho & Wakschlag, 2019; Wakschlag et al., 2014). The Temper Loss scale comprises 22 items (e.g., "Act irritable," "Stay angry for a long time") rated on a 6-point objective frequency scale (0 = *Never*, 6 = *Many times each day*). This subscale has shown good reliability and validity and has been validated in a sociodemographically diverse set of participants (Wakschlag et al., 2014). The MAP-DB has been used to distinguish young children with clinically significant irritability from those whose irritability falls into the normative range concurrently and longitudinally (Wakschlag et al., 2015; Wiggins et al., 2017) and to examine the relationship between irritability and neural responses during affectively challenging tasks in an independent study (Grabell et al., 2018; Perlman et al., 2015). As described in Appendix A, these cut-points showed strong sensitivity and specificity in classifying the presence of clinical impairment in the MAP-DB calibration sample. Given the translational premise of the present study, we used the MAP-DB Temper Loss cut-off for clinically significant irritability, a sum score of 42.5 (see Appendix A).

Externalizing Symptoms

Parents completed DSM-IV symptom indices, the Early Childhood Inventory (ECI), or Child Symptom Inventory (CSI), depending on whether their child was between ages 3–5 years or 6–7 years, respectively (Sprafkin, Gadow, Salisbury, Schneider, & Loney, 2002; Sprafkin, Volpe, Gadow, Nolan, & Kelly, 2002). The ECI and CSI are analogous instruments that assess early childhood behaviors conforming to traditional DSM diagnostic categories. Both measures assess the same DSM-IV based symptoms, and the vast majority of items are worded identically across measures, with

occasional phrasing changes for some CSI items to reflect that older children are in school (e.g., "plays hooky from school"). Researchers commonly use the ECI and CSI together to assess mood and behavior symptoms when samples comprise children ranging from preschool to early school age (Gadow, DeVincent, Pomeroy, & Azizian, 2004; McMahon & Frick, 2005). Similar to prior work with this measure (Gadow & Sprafkin, 1987), we averaged the attention-deficit/hyperactivity combined-type, oppositional-defiance disorder, and conduct disorder scale T -scores to create an externalizing behavior problems composite. We used a T -score of 65 as the clinical cut-off as this represents 1.5 standard deviations from the mean and commonly denotes nonclinical from borderline-to-clinical scores in widely used child psychopathology instruments (Achenbach & Rescorla, 2000; Hanish et al., 2004). Initial descriptive statistics revealed only two children (2.6% of the sample) fell above the T -score of 65 cut-off for internalizing problems. The percentage of children who fell within the clinical range on this scale was comparable to base rates of broad internalizing problems reported in larger, epidemiological samples (Egger & Angold, 2006). However, this prevented us from including internalizing behavior problems in the present analyses.

FRUSTRATION TASK

Children completed the Frustration Emotion Task for Children (FETCH), a validated frustration induction task developed for young children (Grabell et al., 2018; Grabell et al., 2017; Li, Grabell, Wakschlag, Huppert, & Perlman, 2017; Perlman, Luna, Hein, & Huppert, 2014). Prior to the task, children were shown three boxes: a blue box containing exciting and attractive toys, a red box containing small stickers, and a yellow box containing a single broken crayon. Children ranked the three boxes by preference and were told their performance on the game would determine from which box they would choose their final prize at the end. Next, the child interacted with a touchscreen monitor to compete against Sparky, "a very sneaky dog," to fetch bones by touching the bone as it appeared on the screen. Unbeknownst to the child, each trial was fixed where sometimes the child could fetch the bone before Sparky (win trials), and sometimes Sparky would fetch the bone before the child's possible reaction time (frustration trials; see Figure 1). Win trials depicted an animated drawing showing the child grabbing the bone and placing it within one of five boxes, indicating progression towards the most desired reward (the blue box).

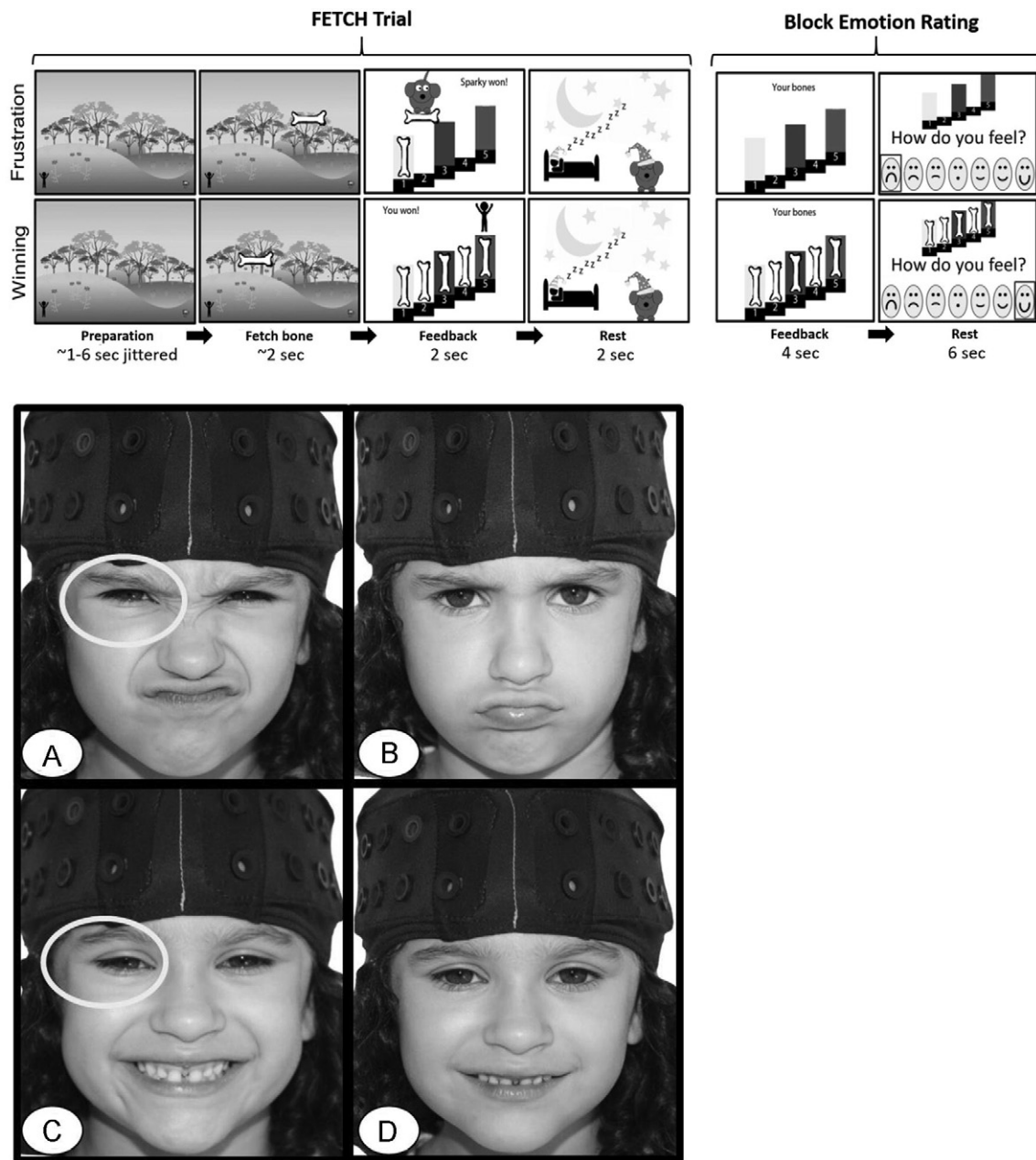


FIG. 1 TOP: Depiction of the Frustration Emotion Task for Children (FETCH). BOTTOM: Examples of negative expressions with (A) and without (B) eye constriction (highlighted with oval), and positive expressions with (C) and without (D) eye constriction.

Frustration trials showed Sparky grabbing the bone and then taking a bone out of the previously won box, indicating the child was getting further away from the most desired reward. Five bones had to be accumulated to win a prize from the large (blue) box. Each trial comprised 2 seconds in which the bone appeared on the screen for the child to fetch, 2 seconds of feedback in which a bone was earned or removed, and a 2-second inter-stimulus interval in which the child was told to rest. The task was animated and contained engaging sound effects.

Trials were grouped into three win and two frustration blocks. Win blocks comprised five win and one frustration trial, except for the final win block, which had an extra win trial so the child would beat the game. Frustration blocks comprised five frustration and one win trial. Children completed an online emotion rating after each block by choosing from seven cartoon faces ranging from negative to positive to indicate their current mood state. The entire task lasted approximately 10 minutes. As described elsewhere (Grabell et al.,

2018; Grabbell et al., 2017), children's prefrontal activation was recorded during the game via functional near-infrared spectroscopy (fNIRS). The fNIRS cap did not obstruct children's faces and these data were not relevant to the present study.

FACIAL CODING

Facial expressions were recorded throughout the FETCH task using a high-definition camcorder mounted on a platform directly above the touch-screen monitor. Facial coders coded both win and frustration blocks, and epochs comprising winning and frustration trials were denoted in the video file using ELAN software (Brugman, Russel, & Nijmegen, 2004). Facial codes comprised a subset of facial movements, or Action Units (AU), from the Facial Actions Coding System (FACS) (Ekman & Friesen, 1977). Consistent with our prior work (Grabbell et al., 2018), the following AUs were coded: brow lowerer (corrugator supercilli; FACS AU 4), eye constriction (orbicularis oculi; FACS AU 6), nose wrinkle (levator labii superioris alaque nasi; AU 9), upper lip raiser (levator labii superioris; FACS AU 10), lip corner puller (zygomaticus major, FACS AU 12), and lip corner depressor (depressor anguli oris; FACS AU 15). Onset and duration of facial codes were continuously denoted in ELAN at the frame-by-frame level using multiple tiers by clicking and dragging through the video feed. Coders were five FACS-trained laboratory members who had passed the FACS certification test and a test custom-designed by our laboratory for coding videos of children (Grabbell et al., 2018). To ensure coders were blind to whether epochs were win or frustration blocks, win/frustration labels were hidden in the files, separated into distinct files and assigned in a random order, across subjects, and coded on mute. Coders were only assigned subjects with whom they had no previous interaction (i.e., were not present when the child was tested) and thus had no previous knowledge of the child's temperament. To assess reliability, 43% of epochs were double coded (coded independently by two individual coders). This percentage was determined by how many videos our undergraduate FACS certified coders could double code during the time they were in the lab across the duration of the project and is similar to double-coding percentages reported in other studies of children's facial expressions (Breau et al., 2001; Cole et al., 1994; Hanish et al., 2004). After each epoch was double coded, the two coders re-watched the footage together to resolve discrepancies, add codes that were originally missed, or remove codes that both had originally denoted but were deemed false positives upon review. The purpose of this intensive review was to establish how reliable each

coder was with a highly accurate "consensus" code. If the two coders had a disagreement about a discrepancy that they could not resolve themselves, a third independent FACS certified coder served as a tiebreaker. After double coding, a consensus code file was created and reliability was calculated as the agreement between each coder and the consensus code using the formula described in the FACS manual: (Number of codes agreed upon by coder and consensus code) \times 2, divided by (total number of facial codes scored between the coder and consensus code). Overall agreement was excellent (85%). In addition, Cohen's Kappa was calculated for each AU and aggregated across subjects, using a strategy similar to Sayette, Cohn, Wertz, Perrott, and Parrott (2001). Kappa values ranged from moderate to excellent as follows: AU 4 = .85, AU 6 = .77, AU 9 = .79, AU 10 = .70, AU 12 = .50, AU 15 = .80. Notably, the percent agreement for AU12 was 82%, and Cohen's Kappa may be an overly conservative estimate of agreement, particularly in cases where it is unlikely that raters engaged in "guessing" (McHugh, 2012) such as in FACS coding (Ekman & Friesen, 1977).

In order to quantify the ELAN clicked-and-dragged events for statistical analyses, ELAN files were exported and parsed into 100 ms bins that denoted whether each FACS AU was present or absent, and whether the concurrent trial was a winning or frustration trial. These bins were aggregated in order to calculate the percentage of time four mutually exclusive expressions (depicted in Figure 1) occurred for each child as follows. Negative expressions with eye constriction were defined as the percentage of time during frustration brow furrowing, nose wrinkling, lip raising, or frowning co-occurred with presence of eye constriction, or the presence of eye constriction on its own without smiling (i.e., wincing) out of the total codeable portion of the video. Negative expressions without eye constriction were defined as the percentage during frustration these movements occurred without presence of eye constriction. Positive expression variables included percentage of smiling during frustration with eye constriction (lip corner raising with eye constriction), and without eye constriction (just lip corner raising).

TESTS OF CLINICAL UTILITY

We first examined distributions of variables of interest and tested whether positive and negative expressions differed by trial type, valence type, and presence/absence of eye constriction via repeated measures ANOVA. Next, we used bivariate correlations to test associations between frequency of facial expressions and continuous irritability and

Table 1
Descriptive Statistics and Intercorrelations Between Expressions, Irritability, and Externalizing Problems

Panel A: Descriptive Statistics			
Variable	Mean	SD	Range
Negative Expressions (% during frustration trials)			
With Eye Constriction	0.74	1.73	0-9.3
Without Constriction	3.96	6.77	0-42.5
Positive Expressions (% during frustration trials)			
With Eye Constriction	7.67	10.47	0-41.0
Without Eye Constriction	27.53	20.52	0-89.3
Negative Expressions (% during win trials)			
With Eye Constriction	0.37	0.69	0-2.86
Without Constriction	0.04	0.21	0-1.93
Positive Expressions (% during win trials)			
With Eye Constriction	3.47	6.46	0-40.6
Without Eye Constriction	0.26	.017	0-0.77
MAP DB Temper Loss Scale	19.86	18.99	0-107
ECI/CSI Externalizing T-score	50.53	8.70	39-74

Panel B: Correlations

Variable	1	2	3	4	5	6
1. Negative Expressions with Eye Constriction	-					
2. Negative Expressions without Constriction	.072	-				
3. Positive Expressions with Eye Constriction	.345**	.028	-			
4. Positive Expressions without Constriction	.020	-.100	.378**	-		
5. MAP DB Temper Loss Scale	.379**	-.076	.196	.082	-	
6. ECI/CSI Externalizing T-score	.404***	.076	.154	-.111	.546***	-

** $p < .01$, *** $p < .001$. Note. MAP DB = Multidimensional Assessment of Preschool Disruptive Behavior; ECI = Early Childhood Inventory
CSI = Childhood Symptom Inventory

externalizing scores. We then compared the frequency of facial expressions in children who fell above and below clinical cut-offs using Receiver Operator Characteristic (ROC) curves. We calculated the area under the ROC curve (AUC) in order to assess the clinical utility of individual facial expressions to correctly classify the presence of early psychopathology. Consistent with the existing literature (Raiker et al., 2017; Swets et al., 1979), we classified the clinical utility of AUC values as follows: $.60 < \text{“fail”}$, $< .70 \text{ “poor”}$, $\geq .70 \text{ “fair”}$, $\geq .80 \text{ “good”}$, and $\geq .90 \text{ “excellent.”}$ Notably, AUCs above .90 are exceptionally rare in the behavioral health sciences, and values between .70 and .80 are considered indicative of instruments that would perform well in realistic clinical conditions (Youngstrom, 2013). For each AUC that surpassed fair clinical utility, we examined the sensitivity and specificity of both the most optimal, and most practical, cut-points. The most optimal cut-point was calculated using Youden’s index (Youden, 1950) as the point along the ROC curve where the summed sensitivity and specificity values were greatest. The most practical clinical cut-point was > 0 , in other words, whether the child produced or did not produce the facial expression at all, the diagnostic rule easiest to implement in a real-

world setting. The DeLong test was used to compare AUCs between ROC curves that surpassed adequate clinical utility (DeLong, DeLong, & Clarke-Pearson, 1988). Finally, for facial expressions that showed clinical utility, odds ratios were calculated to estimate the relative risk a child would meet criteria for disorder if they ever made the expression.

Results

DESCRIPTIVE STATISTICS AND REPEATED MEASURES ANOVA

Descriptive statistics of study variables are shown in Table 1. Sixty-seven percent of the sample produced a negative expression without eye constriction, and 40% produced a negative expression with eye constriction, in response to at least one frustration trial during the FETCH task. Negative expressions with and without eye constriction, during frustration trials, were marginally related to children’s averaged emotion self-ratings following negative blocks (with eye constriction: $r = -.22$, $p = .08$; without eye constriction: $r = -.23$, $p = .07$) such that children with more negative expressions reported lower moods. Positive expressions during frustration or win trials were unrelated to emotion self-ratings (all $p > .11$). MAP DB Temper

Loss scores ranged from low (0) to severe (107) (maximum possible score was 110; $M = 19.9$, $SD = 18.9$). In the present sample, the 42.5 clinical cut-off fell at the 89th percentile and comprised nine children. ECI externalizing T scores ranged from 39 to 75 ($M = 50$, $SD = 8.7$). Eight children (10.3% of the sample) fell above the ECI externalizing clinical cut-off. Continuous irritability and externalizing scores were moderately correlated ($r = .55$, $p < .001$). Four children exceeded the clinical cut-off for both irritability and externalizing behaviors, suggesting clinical groups had some overlap but were not redundant.

A 2 (positive, negative trial) by 2 (positive, negative valence expression) by 2 (presence, absence of eye constriction) repeated measures ANOVA revealed main effects of trial type, $F(1, 78) = 167.28$, $p < .001$, such that children made more expressions during frustration trials than winning trials; valence, $F(1, 78) = 96.98$, $p < .001$, such that positive expressions were more common than negative expressions; and eye constriction, $F(1, 78) = 64.93$, $p < .001$, such that expressions without eye constriction were more common than expressions with eye constriction. Results also revealed a significant Trial \times Valence interaction, $F(1, 78) = 86.99$, $p < .001$, such that the pattern of children making more expressions during negative trials than positive trials was stronger for positive expressions than negative expressions. There was also a Trial \times Eye Constriction, $F(1, 78) = 140.15$, $p < .001$, Valence \times Eye Constriction, $F(1, 78) = 39.30$, $p < .001$, and Trial \times Valence \times Eye Constriction interaction, $F(1, 78) = 70.27$, $p < .001$, such that negative trials and negative expressions were more likely to include eye constriction than positive trials or positive expressions. Because our hypotheses were guided by developmental theory that irritability and externalizing behaviors are linked to frustrating events (Deater-Deckard, Petrill, & Thompson, 2007; Leibenluft, 2017), and data showed that win trials had significantly fewer expressions than frustration trials, correlations and ROC analyses focused solely on frustration trial expressions. Analyses between win trial expressions, irritability, and externalizing behavior are presented in Appendix A (see supplementary Table B).

FACIAL EXPRESSIONS AND CONTINUOUS PSYCHOPATHOLOGY SCORES

As shown in Table 1, frequency of negative expressions with eye constriction were significantly positively correlated with parent-rated irritability ($r = .37$, $p < .01$) and externalizing problems ($r = .40$, $p < .001$), such that children who more frequently produced these expressions had more severe symptoms than peers. All other expressions were unrelated to irritability and externalizing behavior problems.

CLINICAL UTILITY OF FACIAL EXPRESSIONS

Irritability

As shown in Figure 2, negative expressions with eye constriction discriminated children with and without clinical irritability significantly better than chance ($z = 3.2$, $p < .01$), with an AUC of .80, suggesting good clinical utility. Negative faces without eye constriction did not discriminate children with and without clinically significant irritability and demonstrated poor clinical utility. Positive faces with eye constriction also discriminated children with and without clinical irritability better than chance ($z = 2.2$, $p < .05$) and showed fair clinical utility (AUC = .72). Positive faces without eye constriction did not perform better than chance and demonstrated no clinical utility. The DeLong Test revealed that the AUC associated with negative faces with eye constriction was significantly larger than negative expressions ($z = 2.04$, $p < .05$) and positive expressions without eye constriction ($z = 2.83$, $p < .01$), but not positive expressions with eye constriction. Youden's index revealed the cut-point for negative expressions with eye constriction that maximized both sensitivity and specificity was 0.55%, which correctly identified 78% of true positives and 84% of true negatives. The most clinically practical cut-point, > 0 , correctly identified 78% of true positives and 64% of true negatives.

Externalizing Problems

Negative expressions with eye constriction discriminated children with and without clinical externalizing problems significantly better than chance ($z = 2.69$, $p < .01$) with an AUC of .79, suggesting good clinical utility. All other facial expressions did not perform better than chance and showed no clinical utility. The DeLong Test revealed that the AUC associated with negative faces with eye constriction was significantly greater than negative expressions without eye constriction ($z = 1.96$, $p = .05$), and positive expressions with eye constriction ($z = 2.85$, $p < .01$). Youden's index revealed the cut-point for negative expressions with eye constriction that maximized both sensitivity and specificity was 1.48%, which correctly identified 63% of true positives and 93% of true negatives. The most clinically practical cut-point, > 0 , correctly identified 75% of true positives and 63% of true negatives.

ODDS RATIOS

Odds ratios revealed that children who made a negative expression with eye constriction were 6.3 times more likely to meet criteria for clinically significant irritability ($df = 1$, $\chi^2 = 5.85$, $p < .05$) and 5.2 times more likely to meet criteria for clinically significant externalizing problems ($df = 1$,

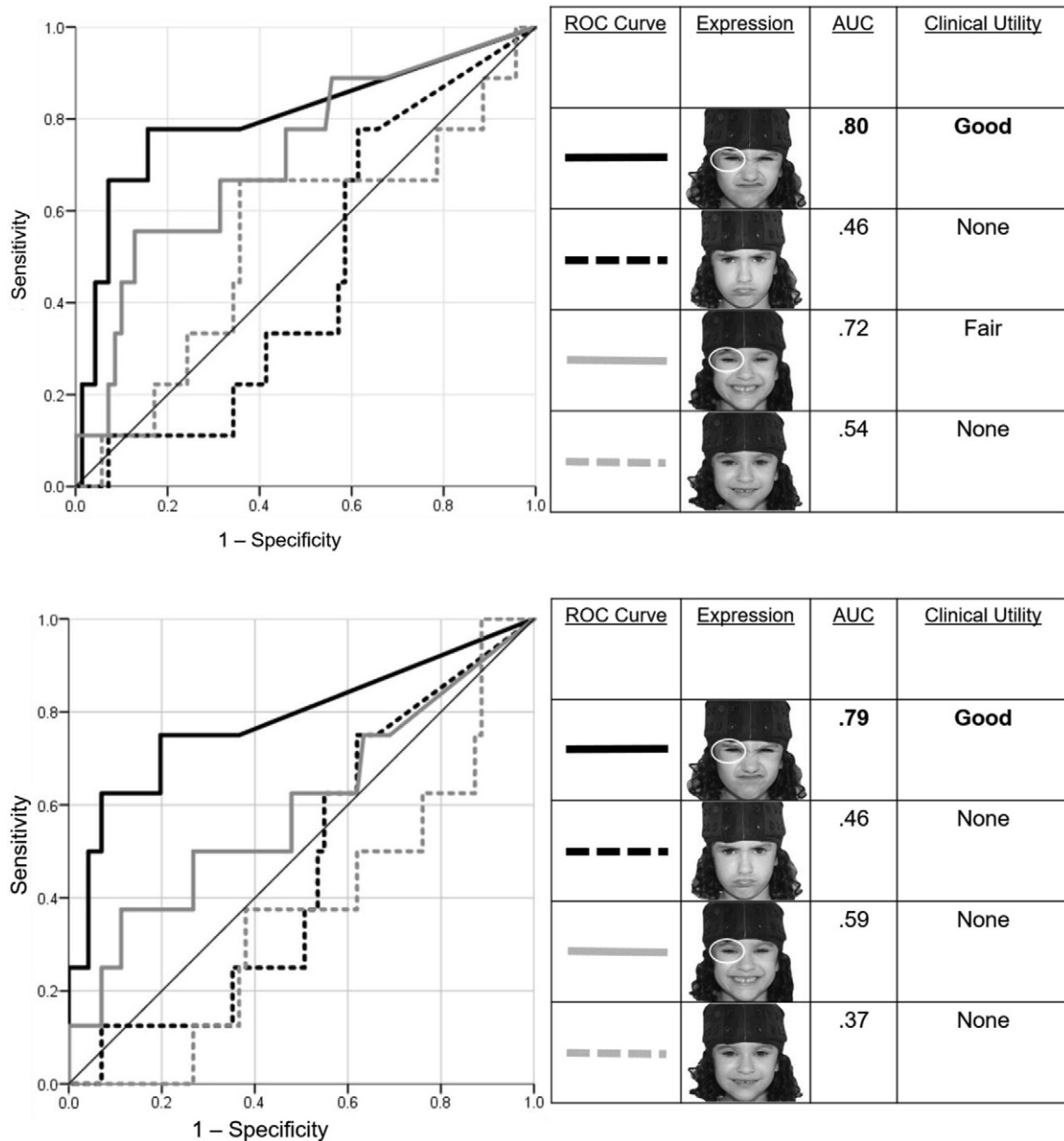


FIG. 2 ROC curves, with corresponding Area Under the Curve (AUC) values and clinical utility, showing the accuracy of negative (black lines) and positive (gray lines) expressions with (solid lines) and without (dashed lines) eye constriction, during frustration trials, in predicting parent-rated irritability (TOP) and externalizing problems (BOTTOM). Diagonal segments produced by ties.

$\chi^2 = 4.34, p < .05$), compared to peers who never made this expression.

Discussion

Our goal was to test how young children's facial expressions in response to brief, moderate frustration related to irritability and externalizing problems and distinguished normative versus clinically significant levels. To our knowledge, the present study is the first to examine the clinical utility of facial expressions to classify the presence or absence of psychopathology. Using the Facial Action Coding System, we calculated how frequently children produced negative and

positive expressions with and without eye constriction during a 10-minute computer game in which progress toward a goal was occasionally blocked. Correlations with continuous scores revealed only negative expression with eye constriction were associated with irritability and externalizing behavior problems. Receiver Operator Characteristic curves revealed that frequency of negative expressions with eye constriction significantly discriminated children with and without clinically significant irritability and externalizing problems, with AUC values suggesting "good" clinical utility. While positive expressions with eye constriction also significantly discriminated

children with and without clinically significant irritability, clinical utility was fair, and this expression did not distinguish children with significant externalizing behaviors or relate to continuous scores. Negative and positive facial expressions without eye constriction were unrelated to any clinical scale.

Negative expressions with eye constriction correctly classified true positives and true negatives, for both the optimal and practical cut-points, at a level of accuracy commensurate with recent studies that investigated the diagnostic efficacy of widely used psychopathology scales in children (Raiker et al., 2017). While our sample of 79 children was relatively smaller than other highly cited clinical pediatric ROC curve studies (Biederman, Monuteaux, Kendrick, Klein, & Faraone, 2005; Chen, Faraone, Biederman, & Tsuang, 1994; Doyle, Ostrander, Skare, Crosby, & August, 1997; Hudziak, Copeland, Stanger, & Wadsworth, 2004), which had samples ranging from around 100 to 300 children, these studies aimed to test the clinical utility of widely adapted scales such as the CBCL and BASC. In contrast, the present study employed an intensive facial coding methodology and results serve as a “proof of concept” that highly operationalized expressions, particularly if combined with existing questionnaires, may help detect clinically significant irritability and externalizing behaviors, and, crucially, that some expressions classify better than others. Moreover, meaningful variation in negative expressions with eye constriction were captured using a short, simple, moderately frustrating computer game requiring minimal guidance from a laboratory research assistant. This raises the exciting possibility that by tracking a single, specific facial expression during a brief emotional challenge, future iterative research may lead to much needed standardized diagnostic instruments with ultra-low time and training burdens. Clinicians working in community settings typically have high caseloads of children with myriad mood and behavior problems and, particularly in low-resource clinics, short turnaround times to make initial clinical decisions (Glisson, Dukes, & Green, 2006). While community clinicians report valuing standardized behavioral observation systems, they rarely use them due to logistical obstacles related to training and implementation (Jensen-Doss & Hawley, 2010; Whiteside et al., 2016). Rather, clinicians typically supplement parent and teacher reports with their own qualitative observations of the child, which may be subject to bias (Meyer et al., 2001). Instruments that can be quickly and easily implemented in clinics, schools, or homes by various mental health providers, or caregivers themselves, could substantially improve the efficiency and accuracy of initial mental health assessments when combined with standardized rating scales and interviews.

The present findings also suggest an important shift in our understanding of how young children with varying levels of transdiagnostic symptoms such as irritability and externalizing behaviors express their emotional states. While many studies have found that children with high levels of mood and behavior problems over- or underproduce negative affect relative to peers (Cole et al., 1994), other studies have not found associations (Camras et al., 1990). Correlations between expressions and continuous irritability and externalizing behaviors suggest that links between facial expressions and early psychopathology may be strongly moderated by the presence or absence of eye constriction. While negative expressions without eye constriction may be much more common, at least during lab-based emotional challenges, our data suggest only individual differences in the frequency of negative expressions with eye constriction predict psychopathology symptoms. Literature on eye constriction as a signaler of general emotion intensification (Messinger, Mattson, Mahoor, & Cohn, 2012) suggests the FETCH task, designed to be a short, moderate frustration challenge, may elicit unusually intense levels of negative emotion in children with clinical levels of irritability and externalizing problems. Moreover, a recent study by our group (Grabell et al., 2018) found that young children who more frequently produced negative expressions with eye constriction during the FETCH task had weaker concurrent lateral prefrontal cortex activation, an area implicated in down-regulating negative emotion (Davidson, Putnam, & Larson, 2000; Grabell et al., 2017), compared to peers. Thus, even infrequent instances of eye constriction during short challenges may reflect strong emotion reactivity coupled with poor regulation, indicative of risk for early onset mental disorder.

STRENGTHS, LIMITATIONS AND FUTURE DIRECTIONS

Notable strengths of the present study included the use of predictor and criterion variables occurring at different levels of analysis, and thus shared no method variance, use of a developmentally sensitive computer task to capture young children’s facial expressions in a short window of time, and use of a popular anatomical coding system to operationalize highly specific kinds of expressions. It bears noting that how to define and assay human emotions is one of the most debated issues in the social sciences and philosophy (Barrett, 2015) and all available methods to measure emotion, including facial coding, inherently contain some bias (Schorr, 2001). Here, we used an anatomical coding system in which facial movements, rather than emotion

states, were coded in an attempt to reduce these biases. Findings from the present study also highlight important next steps in advancing our understanding of clinical utility of facial expressions. Because a relatively smaller number of children fell above the clinical cut-off in ROC analyses, a firm consensus on the clinical utility of negative expressions with eye constriction depends on replication with larger samples. Given that the present sample had significant overlap with the sample reported in our previous investigation of facial expressions and neural activation (Grabell et al., 2018), replication with other, independent samples is particularly important. The present study contributes to the literature by providing novel evidence that pursuing expression-based classification instruments is worthwhile and should be guided by a new understanding of the importance of eye constriction. We recruited children from the community, rather than clinics, in order to test whether facial expressions were sensitive enough to identify the presence of psychopathology consistent with population base rates. A low frequency of children with clinically significant internalizing problems was commensurate with population base rates in the early childhood period (Egger & Angold, 2006) and prevented us from testing whether the clinical utility of facial expressions was specific to irritability and externalizing symptoms or whether it may predict internalizing symptoms as well. Testing whether negative expressions with eye constriction discriminate between clinical phenotypes, or serve as a general marker of psychiatric risk, is an important question to address in future research. Notably, we screened for child and family history of psychiatric illness via a telephone screener, which could have resulted in biased or inaccurate reporting. The present findings set the stage for future work testing the efficacy of negative expressions with eye constriction to detect psychopathology in different clinical samples and identified using different clinical metrics.

While our primary goal was to investigate the clinical utility of facial expression for assessment purposes, the present study has potential connections to early childhood psychotherapy as well. In widely used treatments for early irritability and externalizing behavior, such as Parent-Child Interaction Therapy (PCIT), clinicians rely on weekly parent ratings of child behavior to track treatment progress and make decisions about treatment sequencing (Lyon & Budd, 2010). Tracking changes in children's expressions with eye constriction via short frustration probes could, in parallel, inform clinicians of a treatment's ongoing impact on

underlying emotion regulation systems. Further, emotion theorists have long postulated that facial expressions not only reflect internal emotion states, but cause them (for example, feeling happy because you are smiling, or angry because you are frowning; Lange & James, 1922; Tomkins, 1984; Zajonc, Murphy, & Inglehart, 1989). These findings suggest that, in the same way that clinicians teach children to calm themselves by engaging in slow, diaphragmatic breathing, learning to relax facial expressions could serve as a behaviorally based, developmentally appropriate emotion regulation strategy.

In the present study we examined eye constriction cross-sectionally, commensurate with the extant literature on eye constriction to date (Mattson et al., 2013; Messinger et al., 2012). A fascinating but unexplored question is whether individual differences in frustration-related eye constriction remain stable over time. Finally, the impetus for the present study was the need for objective, behavior-based assessment tools that are easy and quick to administer and can be integrated into multiple clinical settings. Our findings suggest future studies could further push the limits of how well facial expressions can predict mental illness by examining shorter spans of time and more automated coding systems. While full FACS coding is time and labor intensive, our results suggest that, as an assessment tool, clinicians may only need to reliably identify if eye constriction occurs or does not occur during a brief frustration probe, rather than code the full array of FACS codes at a frame by frame level. There have been major efforts to develop automated facial recognition systems capable of identifying highly specific expressions, including whether or not eye constriction is present (McDuff et al., 2016; Zeng et al., 2018), and can detect emotions in naturalistic photographs and videos, such as social media posts (McDuff, El Kaliouby, & Picard, 2015). Given the increased ubiquity with which children's facial expressions are captured with various mobile devices (Radesky, Schumacher, & Zuckerman, 2015), automated instruments optimized to detect negative expressions with eye constriction could eliminate the need for human coding and assist caregivers and clinicians with tracking shifts in children's level of risk and/or progress over time. Advances in psychopathology assessment and screening tools based on specific negative expressions could be scalable to large segments of the population and integrated into daily life in ways that are not currently possible.

Conflict of Interest Statement

The authors declare that there are no conflicts of interest.

Appendix A

The Multidimensional Assessment of Preschoolers Study (MAPS) is a large longitudinal study of preschoolers at-risk for disruptive behavior and other forms of psychopathology (Wakschlag et al., 2015). MAPS recruited a diverse set of preschoolers (3-5 years) from urban areas at risk for disruptive behavior and other forms of psychopathology. Participants were recruited from the waiting rooms of pediatric clinics located in the Greater Chicago area. Eligibility criteria for the MAPS study included: (1) ages 3 to 5 years old; (2) were accompanied by a parent or a legal guardian who could complete the Multidimensional Assessment Profile of Disruptive Behavior (MAP-DB) – a parent questionnaire assessing aspects of disruptive behavior in children, and (3) able to speak English or Spanish. An initial sample of young children (N=1,490) was recruited for initial validation of the MAP-DB. A second group of children (N=1,857) was recruited to validate the MAP-DB in an independent sample.

Using the initial sample of 1,490 children, a Temper Loss score of 42.5 indicated a score at the 95th percentile. The relative sensitivity, specificity, positive predictive value, and odds ratio of this Temper Loss score as a marker of clinical impairment was tested using an independent sample of children ($n=390$) who completed the MAP-DB and the Children's Global Assessment Scale (CGAS) (Shaffer et al., 1983). Specifically, we evaluated whether Temper Loss scores above 42.5 predicted significant impairment on the CGAS (CGAS score ≤ 60 vs > 60). These analyses revealed that a Temper Loss score above 42.5 had strong sensitivity (67%) and excellent specificity (90%) as well as a significant odds ratio (17.26; 95% CI: 4.94-60.25; $p < .001$) in predicting significant impairment. In addition, the continuous MAP-DB Temper Loss score correlated highly ($r = .70$) with the PAPA irritability index created by Dougherty and colleagues (2013).

Table A
Correlations between expressions during win trials, irritability, and externalizing symptoms

Variable	1	2	3	4	5	6
1. Negative Expressions with Eye Constriction	-					
2. Negative Expressions without Constriction	-.037	-				
3. Positive Expressions with Eye Constriction	.056	-.055	-			
4. Positive Expressions without Constriction	-.071	-.163	.281*	-		
5. MAP DB Temper Loss Scale	.103	-.013	.208†	.163	-	
6. ECI/CSI Externalizing T-score	.174	-.046	.205†	-.025	.546***	-

† $p < .10$, * $p < .05$, *** $p < .001$. Note. MAP DB = Multidimensional Assessment of Preschool Disruptive Behavior; ECI = Early Childhood Inventory CSI = Childhood Symptom Inventory

Table B
Results of ROC curve analyses for win trial facial expressions

Expression (win trials)	Irritability AUC	Irritability Clinical Utility	Externalizing AUC	Externalizing Clinical Utility
Negative with Eye Constriction	.69	Poor	.70	Fair
Negative without Eye Constriction	.53	None	.54	None
Positive with Eye Constriction	.64	Poor	.69	Poor
Positive Without Eye Constriction	.54	None	.40	None

Note. AUC = Area Under the Curve

References

- Abel, E. L., & Kruger, M. L. (2010). Smile intensity in photographs predicts longevity. *Psychological Science*, 21(4), 542–544. <https://doi.org/10.1177/0956797610363775>
- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms and profiles*. (Vol. 30): Burlington, VT: University of Vermont, Research center for children, youth
- Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology*, 68(2), 331–339. <https://doi.org/10.1037/0022-006X.68.2.331>
- American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Washington, DC: Author.
- Avenevoli, S., Blader, J. C., & Leibenluft, E. (2015). Irritability in youth: an update. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54(11), 881–883. <https://doi.org/10.1016/j.jaac.2015.08.012>
- Barrett, L. F. (2015). Construction as an integrative framework for the science of emotion. *The Psychological Construction of Emotion*, 448–458.
- Biederman, J., Monuteaux, M., Kendrick, E., Klein, K., & Faraone, S. (2005). The CBCL as a screen for psychiatric comorbidity in paediatric patients with ADHD. *Archives of Disease in Childhood*, 90(10), 1010–1015. <https://doi.org/10.1136/adc.2004.056937>
- Bitsko, R. H. (2016). Health care, family, and community factors associated with mental, behavioral, and developmental disorders in early childhood—United States, 2011–2012. *MMWR. Morbidity and Mortality Weekly Report*, 65.
- Bongers, I. L., Koot, H. M., Van der Ende, J., & Verhulst, F. C. (2003). The normative development of child and adolescent problem behavior. *Journal of Abnormal Psychology*, 112(2), 179–192. <https://doi.org/10.1037/0021-843X.112.2.179>
- Breau, L. M., McGrath, P. J., Craig, K. D., Santor, D., Cassidy, K. -L., & Reid, G. J. (2001). Facial expression of children receiving immunizations: a principal components analysis of the child facial coding system. *The Clinical Journal of Pain*, 17(2), 178–186. <https://journals.lww.com/clinicalpain/pages/default.aspx>.
- Brugman, H., Russel, A., & Nijmegen, X. (2004). Annotating Multi-media/Multi-modal Resources with ELAN. *Paper presented at the LREC*
- Camacho, M. C., & Wakschlag, L. S. (2019). Early Childhood Irritability. *Irritability in Pediatric Psychopathology*, 73
- Camras, L. A., Ribordy, S., Hill, J., Martino, S., Sachs, V., Spaccarelli, S., & Stefani, R. (1990). Maternal facial behavior and the recognition and production of emotional expression by maltreated and nonmaltreated children. *Developmental Psychology*, 26(2), 304–312. <https://doi.org/10.1037/0012-1649.26.2.304>
- Chen, W. J., Faraone, S. V., Biederman, J., & Tsuang, M. T. (1994). Diagnostic accuracy of the Child Behavior Checklist scales for attention-deficit hyperactivity disorder: a receiver-operating characteristic analysis. *Journal of Consulting and Clinical Psychology*, 62(5), 1017–1025. <https://doi.org/10.1037/0022-006X.62.5.1017>
- Cohn, J. F., Ambadar, Z., & Ekman, P. (2007). Observer-based measurement of facial expression with the Facial Action Coding System. *The handbook of emotion elicitation and assessment*, 203–221.
- Cole, P. M. (1986). Children's spontaneous control of facial expression. *Child Development*, 57, 1309–1321. <https://doi.org/10.2307/1130411>
- Cole, P. M., Zahn-Waxler, C., Fox, N. A., Usher, B. A., & Welsh, J. D. (1996). Individual differences in emotion regulation and behavior problems in preschool children. *Journal of Abnormal Psychology*, 105(4), 518–529. <https://doi.org/10.1037/0021-843X.105.4.518>
- Cole, P. M., Zahn-Waxler, C., & Smith, K. D. (1994). Expressive control during a disappointment: Variations related to preschoolers' behavior problems. *Developmental Psychology*, 30(6), 835–846. <https://doi.org/10.1037/0012-1649.30.6.835>
- Davidson, R. J., Putnam, K. M., & Larson, C. L. (2000). Dysfunction in the neural circuitry of emotion regulation—a possible prelude to violence. *Science*, 289(5479), 591–594. <https://doi.org/10.1126/science.289.5479.591>
- Deater-Deckard, K., Petrill, S. A., & Thompson, L. A. (2007). Anger/frustration, task persistence, and conduct problems in childhood: A behavioral genetic analysis. *Journal of Child Psychology and Psychiatry*, 48(1), 80–87. <https://doi.org/10.1111/j.1469-7610.2006.01653.x>
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44(3), 837–845. <https://doi.org/10.2307/2531595>
- Doyle, A., Ostrander, R., Skare, S., Crosby, R. D., & August, G. J. (1997). Convergent and criterion-related validity of the behavior assessment system for children-parent rating scale. *Journal of Clinical Child Psychology*, 26(3), 276–284. https://doi.org/10.1207/s15374424jccp2603_6
- Duchenne, G. -B. (1990). *The mechanism of human facial expression*. New York: Cambridge University Press.
- Dunn, D. M., & Dunn, L. M. (2007). *Peabody picture vocabulary test: Manual*. New York: Pearson.
- Egger, H. L., & Angold, A. (2006). Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, 47(3-4), 313–337. <https://doi.org/10.1111/j.1469-7610.2006.01618.x>
- Ekman, P., & Friesen, W. V. (1977). Facial action coding system.
- Ekman, P., Friesen, W. V., O'sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K. ... Ricci-Bitti, P. E. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4), 712–717. <https://doi.org/10.1037/0022-3514.53.4.712>
- Ekman, P., & Rosenberg, E. L. (1997). *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. New York: Oxford University Press.
- Gadow, K. D., DeVincent, C. J., Pomeroy, J., & Azizian, A. (2004). Psychiatric symptoms in preschool children with PDD and clinic and comparison samples. *Journal of Autism and Developmental Disorders*, 34(4), 379–393. <https://doi.org/10.1023/B:JADD.0000037415.21458.93>
- Gadow, K. D., & Sprafkin, J. (1987). Effects of Viewing High Versus Low Aggression Cartoons on Emotionally Disturbed Children. *Journal of Pediatric Psychiatry*, 12(3), 413–427. <https://doi.org/10.1093/jpepsy/12.3.413>
- Glisson, C., Dukes, D., & Green, P. (2006). The effects of the ARC organizational intervention on caseworker turnover, climate, and culture in children's service systems. *Child Abuse & Neglect*, 30(8), 855–880. <https://doi.org/10.1016/j.chiabu.2005.12.010>

- Grabell, A. S., Huppert, T. J., Fishburn, F. A., Li, Y., Jones, H. M., Wilett, A. E. ... Perlman, S. B. (2018). Using facial muscular movements to understand young children's emotion regulation and concurrent neural activation. *Developmental Science*, 21(5)e12628. <https://doi.org/10.1111/desc.12628>
- Grabell, A. S., Li, Y., Barker, J. W., Wakschlag, L. S., Huppert, T. J., & Perlman, S. B. (2017). Evidence of non-linear associations between frustration-related prefrontal cortex activation and the normal: Abnormal spectrum of irritability in young children. *Journal of Abnormal Child Psychology*, 46(1), 1–11. <https://doi.org/10.1007/s10802-017-0286-5>
- Hanish, L. D., Eisenberg, N., Fabes, R. A., Spinrad, T. L., Ryan, P., & Schmidt, S. (2004). The expression and regulation of negative emotions: Risk factors for young children's peer victimization. *Development and Psychopathology*, 16(2), 335–353. <https://doi.org/10.1017/S0954579404044542>
- Harker, L., & Keltner, D. (2001). Expressions of positive emotion in women's college yearbook pictures and their relationship to personality and life outcomes across adulthood. *Journal of Personality and Social Psychology*, 80(1), 112–124. <https://doi.org/10.1037/0022-3514.80.1.112>
- Hudziak, J. J., Copeland, W., Stanger, C., & Wadsworth, M. (2004). Screening for DSM-IV externalizing disorders with the Child Behavior Checklist: a receiver-operating characteristic analysis. *Journal of Child Psychology and Psychiatry*, 45(7), 1299–1307. <https://doi.org/10.1111/j.1469-7610.2004.00314.x>
- Jensen-Doss, A., & Hawley, K. M. (2010). Understanding barriers to evidence-based assessment: Clinician attitudes toward standardized assessment tools. *Journal of Clinical Child and Adolescent Psychology*, 39(6), 885–896. <https://doi.org/10.1080/15374416.2010.517169>
- Kamphaus, R. W., & Frick, P. J. (2005). *Clinical assessment of child and adolescent personality and behavior*. New York: Springer Science & Business Media.
- Lange, C. G., & James, W. (1922). *The emotions* (Vol. 1). New York: Williams & Wilkins.
- Leibenluft, E. (2017). Pediatric irritability: a systems neuroscience approach. *Trends in Cognitive Sciences*, 21(4), 277–289. <https://doi.org/10.1016/j.tics.2017.02.002>
- Li, Y., Grabell, A. S., Wakschlag, L. S., Huppert, T. J., & Perlman, S. B. (2017). The neural substrates of cognitive flexibility are related to individual differences in preschool irritability: a fNIRS investigation. *Developmental Cognitive Neuroscience*, 25, 138–144. <https://doi.org/10.1016/j.dcn.2016.07.002>
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C. ... Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, 30(3), 205–223. <https://doi.org/10.1023/A:1005592401947>
- Lyon, A. R., & Budd, K. S. (2010). A community mental health implementation of parent–child interaction therapy (PCIT). *Journal of Child and Family Studies*, 19(5), 654–668. <https://doi.org/10.1007/s10826-010-9353-z>
- Mash, E. J., & Hunsley, J. (2005). Evidence-based assessment of child and adolescent disorders: Issues and challenges. *Journal of Clinical Child and Adolescent Psychology*, 34(3), 362–379. https://doi.org/10.1207/s15374424jccp3403_1
- Mattson, W. I., Cohn, J. F., Mahoor, M. H., Gangi, D. N., & Messinger, D. S. (2013). Darwin's Duchenne: Eye constriction during infant joy and distress. *PLoS One*, 8(11)e80161. <https://doi.org/10.1371/journal.pone.0080161>
- McDuff, D., El Kaliouby, R., & Picard, R. W. (2015). Crowdsourcing facial responses to online videos. *Paper presented at the Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on Affective Computing and Intelligent Interaction (ACII), Xi'An, China*
- McDuff, D., Mahmoud, A., Mavadati, M., Amr, M., Turcot, J., & Kaliouby, R. e. (2016). AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit. *Paper presented at the Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*
- McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, 22(3), 276–282.
- McMahon, R. J., & Frick, P. J. (2005). Evidence-based assessment of conduct problems in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, 34(3), 477–505. https://doi.org/10.1207/s15374424jccp3403_6
- Messinger, D. S., Mattson, W. I., Mahoor, M. H., & Cohn, J. F. (2012). The eyes have it: Making positive expressions more positive and negative expressions more negative. *Emotion*, 12(3), 430–436. <https://doi.org/10.1037/a0026498>
- Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R. ... Reed, G. M. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist*, 56(2), 128–156. <https://doi.org/10.1037/0003-066X.56.2.128>
- Perlman, S. B., Jones, B. M., Wakschlag, L. S., Axelson, D., Birmaher, B., & Phillips, M. L. (2015). Neural substrates of child irritability in typically developing and psychiatric populations. *Developmental Cognitive Neuroscience*, 14, 71–80. <https://doi.org/10.1016/j.dcn.2015.07.003>
- Perlman, S. B., Luna, B., Hein, T. C., & Huppert, T. J. (2014). fNIRS evidence of prefrontal regulation of frustration in early childhood. *NeuroImage*, 85, 326–334. <https://doi.org/10.1016/j.neuroimage.2013.04.057>
- Radesky, J. S., Schumacher, J., & Zuckerman, B. (2015). Mobile and interactive media use by young children: the good, the bad, and the unknown. *Pediatrics*, 135(1), 1–3. <https://doi.org/10.1542/peds.2014-2251>
- Raiker, J. S., Freeman, A. J., Perez-Algorta, G., Frazier, T. W., Findling, R. L., & Youngstrom, E. A. (2017). Accuracy of Achenbach scales in the screening of attention-deficit/hyperactivity disorder in a community mental health clinic. *Journal of the American Academy of Child and Adolescent Psychiatry*, 56(5), 401–409. <https://doi.org/10.1016/j.jaac.2017.02.007>
- Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P. ... Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, 45(9), 1094–1103. <https://doi.org/10.1097/01.chi.0000227880.42780.0e>
- Ryan, A., Cohn, J. F., Lucey, S., Saragih, J., Lucey, P., De la Torre, F., & Rossi, A. (2009). Automated facial expression recognition system. *Paper presented at the 43rd Annual International Carnahan Conference on Security Technology*
- Sayette, M. A., Cohn, J. F., Wertz, J. M., Perrott, M. A., & Parrott, D. J. (2001). A psychometric evaluation of the facial action coding system for assessing spontaneous expression. *Journal of Nonverbal Behavior*, 25(3), 167–185.
- Schorr, A. (2001). Subjective measurement in appraisal research. *Appraisal processes in emotion: Theory, methods, research*. 331–349.
- Seder, J. P., & Oishi, S. (2012). Intensity of smiling in Facebook photos predicts future life satisfaction. *Social Psychological and Personality Science*, 3(4), 407–413. <https://doi.org/10.1177/1948550611424968>

- Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A children's global assessment scale (CGAS). *Archives of General Psychiatry*, 40(11), 1228–1231. <https://doi.org/10.1001/archpsyc.1983.01790100074010>
- Southam-Gerow, M. A., & Kendall, P. C. (2002). Emotion regulation and understanding: Implications for child psychopathology and therapy. *Clinical Psychology Review*, 22(2), 189–222. [https://doi.org/10.1016/S0272-7358\(01\)00087-3](https://doi.org/10.1016/S0272-7358(01)00087-3)
- Sprafkin, J., Gadow, K. D., Salisbury, H., Schneider, J., & Loney, J. (2002). Further evidence of reliability and validity of the Child Symptom Inventory-4: Parent checklist in clinically referred boys. *Journal of Clinical Child and Adolescent Psychology*, 31(4), 513–524. https://doi.org/10.1207/S15374424JCCP3104_10
- Sprafkin, J., Volpe, R. J., Gadow, K. D., Nolan, E. E., & Kelly, K. (2002). A DSM-IV–referenced screening instrument for preschool children: The Early Childhood Inventory-4. *Journal of the American Academy of Child and Adolescent Psychiatry*, 41(5), 604–612. <https://doi.org/10.1097/00004583-200205000-00018>
- Stringaris, A. (2011). Irritability in children and adolescents: a challenge for DSM-5. *European Child and Adolescent Psychiatry*, 20(2), 61–66. <https://doi.org/10.1007/s00787-010-0150-4>
- Swets, J. A., Pickett, R. M., Whitehead, S. F., Getty, D. J., Schnur, J. A., Swets, J. B., & Freeman, B. A. (1979). Assessment of diagnostic technologies. *Science*, 205(4408), 753–759. <https://doi.org/10.1126/science.462188>
- Tomkins, S. S. (1984). Affect theory. *Approaches to emotion*, 163, 163–195.
- Tremblay, R. E., Nagin, D. S., Séguin, J. R., Zoccolillo, M., Zelazo, P. D., Boivin, M. ... Japel, C. (2004). Physical aggression during early childhood: Trajectories and predictors. *Pediatrics*, 114(1), e43–e50. <https://doi.org/10.1542/peds.114.1.e43>
- Wakschlag, L. S., Briggs-Gowan, M. J., Choi, S. W., Nichols, S. R., Kestler, J., Burns, J. L. ... Henry, D. (2014). Advancing a multidimensional, developmental spectrum approach to preschool disruptive behavior. *Journal of the American Academy of Child and Adolescent Psychiatry*, 53(1), 82–96. e83. <https://doi.org/10.1016/j.jaac.2013.10.011>
- Wakschlag, L. S., Choi, S. W., Carter, A. S., Hullsiek, H., Burns, J., McCarthy, K. ... Briggs-Gowan, M. J. (2012). Defining the developmental parameters of temper loss in early childhood: implications for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, 53(11), 1099–1108.
- Wakschlag, L. S., Estabrook, R., Pettilerc, A., Henry, D., Burns, J. L., Perlman, S. B. ... Briggs-Gowan, M. L. (2015). Clinical implications of a dimensional approach: the normal: abnormal spectrum of early irritability. *Journal of the American Academy of Child and Adolescent Psychiatry*, 54(8), 626–634.
- Wakschlag, L. S., Perlman, S. B., Blair, R. J., Leibenluft, E., Briggs-Gowan, M. J., & Pine, D. S. (2017). The neurodevelopmental basis of early childhood disruptive behavior: irritable and callous phenotypes as exemplars. *American Journal of Psychiatry*, 175(2), 114–130. <https://doi.org/10.1176/appi.ajp.2017.17010045>
- Wakschlag, L. S., Tolan, P. H., & Leventhal, B. L. (2010). Research Review: 'Ain't misbehavin': Towards a developmentally-specified nosology for preschool disruptive behavior. *Journal of Child Psychology and Psychiatry*, 51(1), 3–22. <https://doi.org/10.1111/j.1469-7610.2009.02184.x>
- Whiteside, S. P., Sattler, A. F., Hathaway, J., & Douglas, K. V. (2016). Use of evidence-based assessment for childhood anxiety disorders in community practice. *Journal of Anxiety Disorders*, 39, 65–70. doi:Use of evidence-based assessment for childhood anxiety disorders in community practice
- Wiggins, J. L., Briggs-Gowan, M. J., Estabrook, R., Brotman, M. A., Pine, D. S., Leibenluft, E., & Wakschlag, L. S. (2017). Identifying Clinically Significant Irritability in Early Childhood. *Journal of the American Academy of Child and Adolescent Psychiatry*, 57(3), 191–199. <https://doi.org/10.1016/j.jaac.2017.12.008>
- Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer*, 3(1), 32–35. [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3)
- Youngstrom, E. A. (2013). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: We are ready to ROC. *Journal of Pediatric Psychology*, 39(2), 204–221. <https://doi.org/10.1093/jpepsy/jst062>
- Zajonc, R. B., Murphy, S. T., & Inglehart, M. (1989). Feeling and facial efference: implications of the vascular theory of emotion. *Psychological Review*, 96(3), 395–416. <https://doi.org/10.1037/0033-295X.96.3.395>
- Zeng, N., Zhang, H., Song, B., Liu, W., Li, Y., & Dobaie, A. M. (2018). Facial expression recognition via learning deep sparse autoencoders. *Neurocomputing*, 273, 643–649. <https://doi.org/10.1016/j.neucom.2017.08.043>

RECEIVED: December 17, 2018

ACCEPTED: August 12, 2019

AVAILABLE ONLINE: 17 August 2019