Received: 13 November 2016 | Revised: 20 June 2017 | Accepted: 11 July 2017

DOI: 10.1111/psyp.12977

ORIGINAL ARTICLE

WILEY PSYCHOPHYSIOLOGY SPR

Emotion processing in female youth: Testing the stability of the late positive potential

Erin Bondy[1,2] | Jeremy G. Stewart[1,2] | Greg Hajcak[3] | Anna Weinberg[4] | Naomi Tarlow[5] | Vijay A. Mittal[6] | Randy P. Auerbach[1,2]

[1]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA

[2]Center for Depression, Anxiety and Stress Research, McLean Hospital, Belmont, Massachusetts, USA

[3]Department of Psychology, Stony Brook University, Stony Brook, New York, USA

[4]Department of Psychology, McGill University, Montreal, Quebec, Canada

[5]Department of Psychology, University of Miami, Coral Gables, Florida, USA

[6]Department of Psychology, Department of Psychiatry, Institute for Policy Research, Department of Medical Social Sciences, Northwestern University, Evanston, Illinois, USA

Correspondence
Randy P. Auerbach, Ph.D., ABPP, McLean Hospital, 115 Mill Street, deMarneffe, Room 240, Belmont, MA 02478, USA.
Email: rauerbach@mclean.harvard.edu

Funding information
National Institute of Mental Health (K23MH097786), Rolfe Fund, Tommy Fuss Fund (to R. P. A.)

Abstract

The Emotional Interrupt Task (EIT) has been used to probe emotion processing in healthy and clinical samples; however, research exploring the stability and reliability of behavioral measures and ERPs elicited from this task is limited. Establishing the psychometric properties of the EIT is critical, particularly as phenotypes and biological indicators may represent traitlike characteristics that underlie psychiatric illness. To address this gap, test-retest stability and internal consistency of behavioral indices and ERPs resulting from the EIT in healthy, female youth ($n = 28$) were examined. At baseline, participants were administered the EIT while high-density 128-channel EEG data were recorded to probe the late positive potential (LPP). One month later, participants were readministered the EIT. Four principal findings emerged. First, there is evidence of an interference effect at baseline, as participants showed a slower reaction time for unpleasant and pleasant images relative to neutral images, and test-retest of behavioral measures was relatively stable over time. Second, participants showed a potentiated LPP to unpleasant and pleasant images compared to neutral images, and these effects were stable over time. Moreover, in a test of the difference waves (unpleasant-neutral vs. pleasant-neutral), there was sustained positivity for unpleasant images. Third, behavioral measures and LPP demonstrated excellent internal consistency (odd/even correlations) across conditions. Fourth, highlighting important age-related differences in LPP activity, younger age was associated with larger LPP amplitudes across conditions. Overall, these findings suggest that the LPP following emotional images is a stable and reliable marker of emotion processing in healthy youth.

KEYWORDS
adolescents, emotion, ERPs

1 | INTRODUCTION

Emotions are complex psychological and physiological processes that shape perception and understanding of the environment (Dolan, 2002; Hajcak, Weinberg, MacNamara, & Foti, 2012; Lang, 1984). As EEG provides excellent temporal resolution in the milliseconds range, it is an excellent tool to probe the time course of neural responses to emotional stimuli. An improved understanding of emotion processing is essential, as this may provide insight into deficits that characterize psychiatric disorders.

Recent research exploring electrophysiological correlates of emotion processing has focused on the late positive potential (LPP). The LPP is a slow-wave ERP sensitive to emotional arousal that reflects elaborated attention toward affective stimuli (Auerbach, Stanton, Proudfit, & Pizzagalli,

2015; Auerbach et al., 2016; Hajcak et al., 2012). The LPP is larger following unpleasant and pleasant stimuli compared to neutral stimuli (Cuthbert, Schupp, Bradley, Birbaumer, & Lang, 2000; Foti & Hajcak, 2008; Foti, Hajcak, & Dien, 2009; Schupp et al., 2000). Generally, the LPP emerges as early as 300 ms poststimulus (Cuthbert et al., 2000) and spans several hundred milliseconds to seconds (Foti & Hajcak, 2008). The LPP is initially maximal over parietal sites (Schupp et al., 2000) and, later in its time course, often propagates to frontocentral regions (Auerbach et al., 2015; Foti & Hajcak, 2008). The LPP also appears to be a variant of the P300, a component that is maximal in parietal regions and emerges between 300–500 ms poststimulus (Olofsson, Nordin, Sequeria, & Polich, 2008; Sutton, Braren, Zubin, & John, 1965). The P300, or P3, indexes increased attention toward salient stimuli, and, similar to the LPP, prior research investigating the P300 has shown increased amplitudes for emotional relative to neutral stimuli (Hajcak, MacNamara, & Olvet, 2010; Polich & Kok, 1995). Taken together, whereas the P300 reflects processes underlying preferential attention to target stimuli, the longer time course of the LPP indexes processes related to sustained attention and encoding of emotional information (Dolcos & Cabeza, 2002).

Evidence from studies using emotional images shows greater LPP amplitude during passive viewing of unpleasant and pleasant images compared to neutral images (Foti et al., 2009; Hajcak & Olvet, 2008; Weinberg & Hajcak, 2010). Although some studies have shown a larger LPP following unpleasant relative to pleasant images (Hajcak & Olvet, 2008; Kujawa, Klein, & Hajcak, 2013), this may reflect differences in the arousal ratings of the images (see Weinberg & Hajcak, 2010). Collectively, these findings suggest that the LPP reflects increased attention toward emotional stimuli (e.g., words, images) and, in some instances, is modulated by valence (Auerbach et al., 2015, 2016; Speed, Nelson, Auerbach, Klein, & Hajcak, 2016).

To improve our understanding of pathophysiological processes underlying emotion processing, research has probed the LPP using the Emotional Interrupt Task (EIT; Mitchell, Richell, Leonard, & Blair, 2006). During the EIT, participants identify a target stimulus that is preceded and followed by task-irrelevant neutral, unpleasant, and pleasant images selected from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 2008). Using the EIT, we can examine (a) behavioral measures of interference and task engagement, and (b) LPP as a function of stimulus valence. In an undergraduate sample, Weinberg and Hajcak (2011a) found a larger LPP following unpleasant and pleasant images compared to neutral images; however, no differences emerged in activity elicited by pleasant and unpleasant images. Interestingly, the LPP following emotional images was associated with slower reaction times when identifying target stimuli. Similarly, among 8- to 13-year-old children, increased LPP following pleasant and unpleasant images was related to faster reaction times to targets (Kujawa, Klein, & Hajcak, 2012). In a study of at-risk youth (i.e., parental history of depression), Nelson and colleagues (Nelson, Perlman, Hajcak, Klein, & Kotov, 2015) showed that parental depression history was associated with attenuated LPP to neutral and emotional stimuli. Further, among healthy youth, Speed and colleagues (2015) demonstrated that higher extraversion was associated with increased LPP to both pleasant and unpleasant images. Together, these findings suggest that emotional stimuli result in increased LPP activity, which at times may interfere with task performance.

In line with the Precision Medicine Initiative (Insel, Amara, & Baschke, 2015), identifying biological indicators that predict treatment response is critical. Prior to testing predictors, research must first examine core psychometric properties, including test-retest stability (i.e., invariance over time) and internal consistency (i.e., odd/even reliability; see Auerbach et al., 2016; Cassidy, Robertson, & O'Connell, 2012; Hess et al., 2017; Olvet & Hajcak, 2009; Tenke et al., 2017; Weinberg & Hajcak, 2011b). In research testing the EIT, Kujawa and colleagues (2013) showed 2-year stability of LPP amplitude in children 8 to 13 years old. At both the baseline and follow-up assessments, there was a greater LPP following pleasant and unpleasant images compared to neutral images, and unpleasant images elicited a greater LPP relative to pleasant images. Age-related differences also emerged, as the LPP was maximal at occipital sites at the first assessment, when participants were younger, whereas the effects were maximal in parietal electrodes 2 years later.

Together, these findings provide initial evidence for stability of the LPP in a limited range of children and early adolescents. Further research, however, is needed to explore psychometric properties across broader developmental periods. To build on prior research, the current study examined the stability and reliability of behavioral indices and the LPP. Healthy, female youth aged 13 to 22 years completed a baseline and 1-month follow-up assessment. The following a priori hypotheses were tested. First, consistent with prior EIT research (Kujawa et al., 2013; Weinberg & Hajcak, 2011a), we expected that healthy youth would (a) show greater accuracy and faster reaction times for neutral relative to pleasant and unpleasant images, and (b) exhibit greater LPP amplitudes to pleasant and unpleasant compared to neutral images. Second, we hypothesized that behavioral and ERP effects would demonstrate stability (i.e., strong test-retest correlations) at the 1-month follow-up assessment. Last, we expected age-related differences in LPP activity. Given prior work demonstrating greater LPP in younger compared to older youth (MacNamara et al., 2016), we expected that younger participants would demonstrate enhanced LPPs in parietal regions compared to older participants.

2 | METHOD

2.1 | Procedure

The Institutional Review Board provided approval for the study. Assent was obtained from youth aged 13 to 17 years, and participants 18 years and older and legal guardians provided written consent. Participants were recruited from the greater Boston area through flyers, online advertisements, and direct mailing. Eligibility criteria included English fluency, right-handedness, and female sex. Exclusion criteria were any history of psychiatric illness, psychotropic medication use, organic brain syndrome, neurologic disorders, or seizures. At baseline, participants were administered a clinical interview assessing lifetime mental illness and a depressive symptom self-report measure. Within 1 to 2 weeks, participants completed the EIT while EEG data were recorded. The mean length of time between the clinical and EEG assessment was 5.32 ± 5.11 days. For the 1-month follow-up assessment, participants returned to the lab and were readministered the self-report measure and EIT (while EEG data were recorded). The length of time between the clinical and EEG follow-up assessment was brief, 0.54 ± 1.57 days. Participants were remunerated $100.

2.2 | Participants

The sample included healthy, female youth aged 13 to 22 years with no lifetime psychopathology. Only female participants were included to reduce heterogeneity. Thirty-three participants were enrolled in the study; however, five youth were excluded from analyses given poor EEG data quality ($n = 1$) and lack of follow-up EEG data ($n = 4$). Excluded participants ($n = 5$) and the final sample ($n = 28$) did not differ in age, $t(31) = 0.91$, $p = .52$, $d = 0.38$, race, $\chi^2(4) = 7.54$, $p = .11$, $\varphi = 0.48$, or family income, $\chi^2(4) = 4.20$, $p = .38$, $\varphi = 0.37$. The final sample ($M_{age} = 17.61$, $SD_{age} = 2.95$) reported the following racial distribution: 21.4% Asian, 3.6% Black or African American, 64.3% White, and 10.7% more than one race. The family income distribution included 14.3% less than $10,000, 7.1% $25,000 to $50,000, 14.3% $50,000 to $75,000, 3.6% $75,000 to $100,000, 50% more than $100,000, and 10.7% not reported.

2.3 | Instruments

2.3.1 | M.I.N.I. International Neuropsychiatric Interview for Child and Adolescents (MINI-KID)

Participants were administered the MINI-KID (Sheehan et al., 2010), a structured diagnostic interview used to assess Axis I psychopathology in children and adolescents. The MINI-KID has shown good reliability and validity (Sheehan et al., 2010). Postbaccalaureate research assistants, graduate students, and postdoctoral fellows administered the interviews after receiving approximately 50 hr of training, which included didactics, listening to past interviews, role play, mock interviews, and direct supervision.

2.3.2 | Beck Depression Inventory (BDI-II)

The BDI-II (Beck, Steer, & Brown, 1996) is a 21-item self-report questionnaire assessing depressive symptom severity over the past 2 weeks. Items range from 0 to 3, with higher scores indicating greater depression severity. The Cronbach's alpha for the BDI-II in the current study was 0.74 at the baseline assessment and 0.88 at the follow-up assessment, suggesting good internal consistency.

2.4 | Experimental task

The EIT (Mitchell et al., 2006) included 60 images selected from the IAPS (Lang et al., 2008): 20 neutral, 20 unpleasant, and 20 pleasant.[1] According to IAPS normative adult ratings (9-point rating scale; Lang et al., 2008), pleasant images used were rated as more positive in valence ($M = 7.51$, $SD = 0.51$) than neutral images ($M = 5.27$, $SD = 0.35$), which were rated as more positive than unpleasant images ($M = 3.09$, $SD = 0.76$). Additionally, both pleasant ($M = 5.03$, $SD = 0.77$) and unpleasant images ($M = 6.12$, $SD = 0.57$) were rated as more arousing than neutral images ($M = 2.99$, $SD = 0.68$), though unpleasant images were rated as more arousing than pleasant images. Images were presented twice over three blocks for a total of 120 trials. Each trial began with a fixation cross presented for 800 ms. Then, an image was displayed for 1,000 ms, followed by a target (< or >) that was presented for 150 ms. Finally, the same image appeared on screen for an additional 400 ms. Participants indicated whether the target arrow was pointing to the right or left by pressing the corresponding button on a response box. The intertrial interval was jittered between 1,500 and 2,000 ms.

EIT behavioral outcomes included reaction time (RT) indexes for correct trials and overall accuracy for each stimulus type (pleasant, unpleasant, neutral). Previous research indicates that RT distributions are not Gaussian (normal)

[1]IAPS images used. Practice images: 9421, 7140, 6260, 7460, 2206, 2750, 9584, 7550, 9160, 2384; pleasant images: 1463, 1710, 1750, 1811, 2070, 2091, 2092, 2224, 2340, 2345, 2347, 7325, 7330, 7400, 8031, 8200, 8370, 8461, 8496, 8497; unpleasant images: 1050, 1052, 1200, 1205, 1300, 1304, 1930, 2458, 2691, 2703, 2800, 2811, 2900, 3022, 6190, 6213, 6231, 6510, 6571, 9600; neutral images: 2514, 2580, 5390, 5395, 5500, 5731, 5740, 5900, 7000, 7002, 7009, 7010, 7026, 7038, 7039, 7090, 7100, 7130, 7175, 7190.

distributions; instead, ex-Gaussian distributions—a combination of Gaussian and exponential distributions that rises rapidly on the left side of the distribution and has a long tail to the right—fit RT distributions optimally (e.g., Balota & Spieler, 1999). The main drawback of a central tendency approach (i.e., computing mean reaction time assuming a normal distribution), especially when untransformed RT data are analyzed, is reduced power. Further, although using cutoffs (e.g., removing RTs longer than a certain absolute value; Mitchell et al., 2006) may improve power in some cases, when the true effect is actually in the long tail of the distribution, cutoff methods produce Type II errors (see Whelan, 2008, for a discussion).

Consequently, prior to analyzing the RT data, we fit an ex-Gaussian distribution to each participant's data using the R package retimes (Massidda, 2013). These distributions have three parameters: the mean (mu) and standard deviation (sigma) of the Gaussian portion of the distribution, and the mean of the exponential portion of the distribution (tau). Briefly, these three parameters were estimated using maximum likelihood (ML) and implementing the simplex method to establish the minimum of the objective function. We used a bootstrapping approach (5,000 samples with replacement) given our sample was relatively small. First, mu and sigma were obtained with a Gaussian kernel estimator (see Van Zandt, 2000), then tau was chosen within the bootstrapped values based on ML criterion.

Classically, mu and tau were proposed to reflect distinct processes where the former is influenced by individual differences in perception and response execution, whereas the latter is more likely to reflect central decision-making processes (Hohle, 1965). Critically, individual differences in attention selection tasks involving choices based on earlier level visual codes (e.g., direction of target) like the EIT typically involve shifts in the entire distribution and may primarily affect mu (e.g., Spieler, Balota, & Faust, 1996). Among other factors, tau is influenced by lapses of attention that produce more frequent longer RT trials, thus creating changes in the left tail of the ex-Gaussian distribution (e.g., Hervey et al., 2006). Given evidence that ex-Gaussian parameters capture separable attentional and cognitive processes, we analyzed mu and tau independently in primary RT analyses.[2]

To analyze accuracy for each stimulus type (pleasant, unpleasant, neutral), we first computed a count of errors made in each condition. To test the effects of stimulus type and time on number of errors, we fit a within-subject Poisson regression model using generalized estimating equation (GEE). In the model, we used an autoregression correlation structure and robust standard errors to control for overdispersion in our data, which is in line with current recommendations (e.g., Cameron & Trivedi, 2009). For all correlational analyses involving accuracy (e.g., test-retest estimates), we fit Poisson regression models with robust standard errors and report betas and standard errors to capture associations.[3]

2.5 | EEG recording, data reduction, and analysis

The EEG was recorded using a 128-channel HydroCel GSN (Electrical Geodesics, Inc., Eugene, OR). Continuous EEG data, referenced to Cz, were sampled at 250 Hz. Electrode impedances were kept below 65 kΩ, and offline analyses were performed using BrainVision Analyzer 2.1 software (Brain Products, Germany). EEG data were rereferenced to the average reference, and low- and high-pass filters were applied at 0.1 and 30 Hz. An independent component analysis (ICA) transform was implemented to identify and remove eye movement artifacts and eyeblinks using the following criteria: whole data, classic PCA sphering, infomax ICA, energy ordering, and 512 convergence steps. For each trial, EEG data were segmented 200 ms before the initial image onset and continued for 1,200 ms. A semiautomated procedure to reject intervals for individual channels used the following criteria: (a) a voltage step > 50 μV between sample rates, (b) a voltage difference > 300 μV within a trial, and (c) a maximum voltage difference of < 0.50 μV within a 100-ms interval. All trials were also visually inspected, and further artifacts were rejected manually. After completing the data reduction steps, we then examined the average number of trials retained per condition (i.e., 40 trials/condition) across assessments: (a) neutral trials: 35.55 ± 4.07, (b) unpleasant trials: 35.96 ± 3.91, (c) pleasant trials: 35.75 ± 4.23.

ERPs were computed time-locked to pretarget neutral, unpleasant, and pleasant images, and the average amplitude 200 ms before the pretarget stimulus onset was used as the baseline. Only trials with a correct response between 150–1,500 ms after target onset were included in ERP averages. The LPP component was calculated as the mean activity at electrode site Pz where the component was maximal across participants for the 400–1,000 ms poststimulus time window.[4] Difference waves also were computed to examine discrepancies between activity during emotional and neutral conditions (difference waves: unpleasant minus neutral vs.

[2]Analyses of sigma yielded nonsignificant results (i.e., no effect of condition, time, or their interaction; no evidence of test-retest reliability) and are available from the authors by request.

[3]In preliminary model building, we also fit negative binomial distribution with log link to the number of errors and found that these models fit slightly less closely than Poisson models.

[4]Given prior work demonstrating that an occipitally maximal LPP characterizes children and early adolescents (e.g., Kujawa et al., 2013), we explored scalp topography maps comparing younger and older participants. Participants exhibited similar parietal distribution across ages.

pleasant minus neutral). In addition to probing subtraction-based difference scores, we computed standardized residuals (i.e., regressed neutral on unpleasant and pleasant images) to extract the unique variance in the emotional images after accounting for the neutral images, an alternative to computing difference scores (see Levinson, Speed, Infantolino, & Hajcak, 2017; Meyer, Lerner, De Los Reyes, Laird, & Hajcak, 2017). All analyses were conducted with SPSS 22.0 (IBM Corp., Armonk, NY). A repeated measures analysis of variance (rmANOVA) tested main effects of time (baseline, follow-up) and condition (neutral, unpleasant, pleasant) as well as a Time × Condition interaction using a Greenhouse-Geisser correction. We expected a significant main effect of condition. To demonstrate stability of a given effect over time, we anticipated that the main effect of time and the Time × Condition interaction would be nonsignificant. To test whether age moderated LPP stability, we conducted an Age (continuous measure) × Time × Condition rmANOVA and, additionally, to determine whether younger participants show greater activity in occipital versus parietal regions (e.g., Hajcak & Dennis, 2009; Kujawa et al., 2013), we conducted Age (continuous measure) × Condition × Electrode (Pz, Oz) rmANOVAs at each time point.

We computed effect sizes (η_p^2) for all analyses, where .02–.12 = small, .13–.25 = medium, and ≥ .26 = large. Test-retest stability for behavioral and ERP measures was assessed using Pearson product-moment correlations with the following criteria: .10–.29 = small, .30–.49 = moderate, ≥ .50 = large (Cohen, 1988). The internal consistency of behavioral measures and the LPP were evaluated through testing the correlation of the odd and even trials. The Spearman-Brown prophecy formula (Nunnally, Bernstein, & Berge, 1967) was used to correct these correlations because the total number of items included in the averages is split in half (reliability = 2 * $r_{odd/even}$/1 + $r_{odd/even}$). Spearman-Brown coefficients were evaluated where > .80 = good/excellent, .70–.79 = acceptable, and < .60 = poor.

TABLE 1 Behavioral data from the Emotional Interrupt Task

	Baseline (n = 28)		Follow-up (n = 28)	
Measure	**Mean**	***SD***	**Mean**	***SD***
Accuracy				
Pleasant	0.96	0.06	0.96	0.04
Unpleasant	0.96	0.05	0.97	0.03
Neutral	0.97	0.05	0.97	0.03
Reaction time—mu (ms)				
Pleasant	390.50	128.93	395.69	166.25
Unpleasant	383.09	136.41	397.02	171.85
Neutral	362.84	133.36	388.26	174.03
Reaction time—tau (ms)				
Pleasant	44.21	23.42	60.31	43.38
Unpleasant	53.77	28.10	60.94	43.33
Neutral	65.55	43.90	61.64	38.72

Note. Reaction time measures are given for correct trials only.

3 | RESULTS

3.1 | Descriptive statistics

Depressive symptoms were assessed at the baseline and follow-up assessment (test-retest r = .82, p < .001) to ensure the nonclinical status across assessments. As expected, depressive symptom scores were low (baseline: 0.86 ± 1.99; follow-up: 1.21 ± 2.87) and did not differ over time, t (27) = −1.12, p = .27, d = −0.24.

3.2 | Behavioral data

Behavioral data from the EIT are summarized in Table 1.

3.2.1 | Number of errors

In our omnibus GEE analysis, the effect of condition was not significant for number of errors, b = −0.02, SE = 0.44, χ^2(1, N = 168) = 0.003, p = .96, OR = 0.98, 95% CI [0.41, 2.32], which may reflect high rates of accuracy across conditions. Additionally, neither the main effect of time, b = −0.18, SE = 0.53, χ^2(1, N = 168) = 0.12, p = .73, OR = 0.83, 95% CI [0.30, 2.33], nor the Time × Condition interaction, b = 0.05, SE = 0.25, χ^2(1, N = 168) = 0.05, p = .83, OR = 1.05, 95% CI [0.65, 1.72], was significant. Test-retest analyses using a series of Poisson regression models revealed an association over time for neutral images, b = 0.14, SE = 0.05, χ^2(1, N = 28) = 6.71, p = .01, OR = 1.15, 95% CI [1.03, 1.27], but not for pleasant, b = 0.03, SE = 0.08, χ^2(1, N = 28) = 0.18, p = .68, OR = 1.03, 95% CI [0.89, 1.20], or unpleasant, b = 0.07, SE = 0.10, χ^2(1, N = 28) = 0.56, p = .45, OR = 1.08, 95% CI [0.89, 1.30], images. There also was strong internal consistency in accuracy at the baseline (Spearman-Brown odd/even corrected reliability = .85) and follow-up (Spearman-Brown odd/even corrected reliability = .78) assessments.

3.2.2 | Reaction time—Mu

In the rmANOVA model for mu, we found a main effect of condition, F(2, 54) = 7.21, p = .003, η_p^2 = .21, and unexpectedly this main effect was qualified by a significant Time × Condition interaction, F(2, 54) = 3.56, p = .04, η_p^2 = .12. As hypothesized, the main effect of time was nonsignificant, F (1, 27) = 0.42, p = .52, η_p^2 = .02. To decompose the interaction, we conducted follow-up simple effects analyses. In the

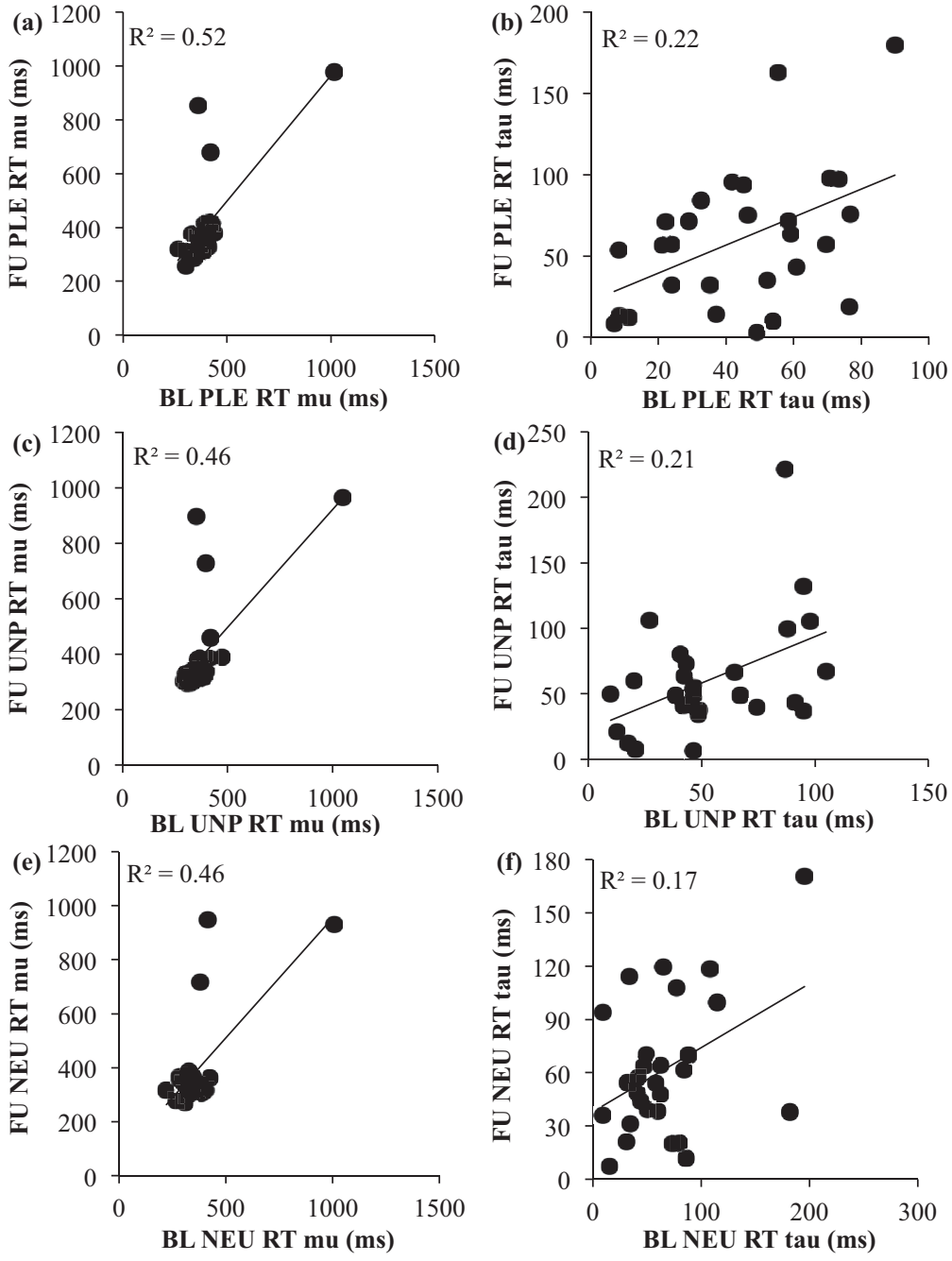

FIGURE 1 Test-retest for reaction time measures at the baseline and follow-up assessments. BL = baseline; FU = follow-up; PLE = pleasant images; UNP = unpleasant images; NEU = neutral images. Correlations depicted include all participants

baseline simple effects model, the effect of condition was significant, $F(2, 54) = 10.22$, $p < .001$, $\eta_p^2 = .27$, such that participants had slower RTs for both pleasant ($M = 362.84$, $SE = 24.37$) and unpleasant ($M = 383.09$, $SE = 25.78$) stimuli compared to neutral ($M = 362.84$, $SE = 25.20$) stimuli, $ps < .02$, $ds > 0.49$. In contrast, RTs for pleasant and unpleasant trials did not significantly differ, $p = .11$, $d = 0.33$. However, in the simple effects model for the follow-up assessment, the effect of condition was nonsignificant, $F(2, 54) = 1.19$, $p = .31$, $\eta_p^2 = .04$. Further, for mu, there were significant associations over time for pleasant ($r = .76$, $p < .001$), unpleasant ($r = .72$, $p < .001$), and neutral ($r = .68$, $p < .001$) images (see Figure 1a,c,e).[5] When

[5]Two participants were identified as univariate outliers (i.e., mu values for all three conditions were 3 *SD* above the mean at the follow-up assessment). All reaction time analyses were run with and without these participants. The results from the rmANOVA models did not change appreciably. For test re-test correlations, the correlation between baseline and follow-up mu in the neutral condition was nonsignificant when the outliers were removed (r = .23, p = .25). In contrast, the correlations for unpleasant (r = .53, p = .005) and pleasant (r = .44, p = .02) stimuli were reduced, but remained statistically significant.

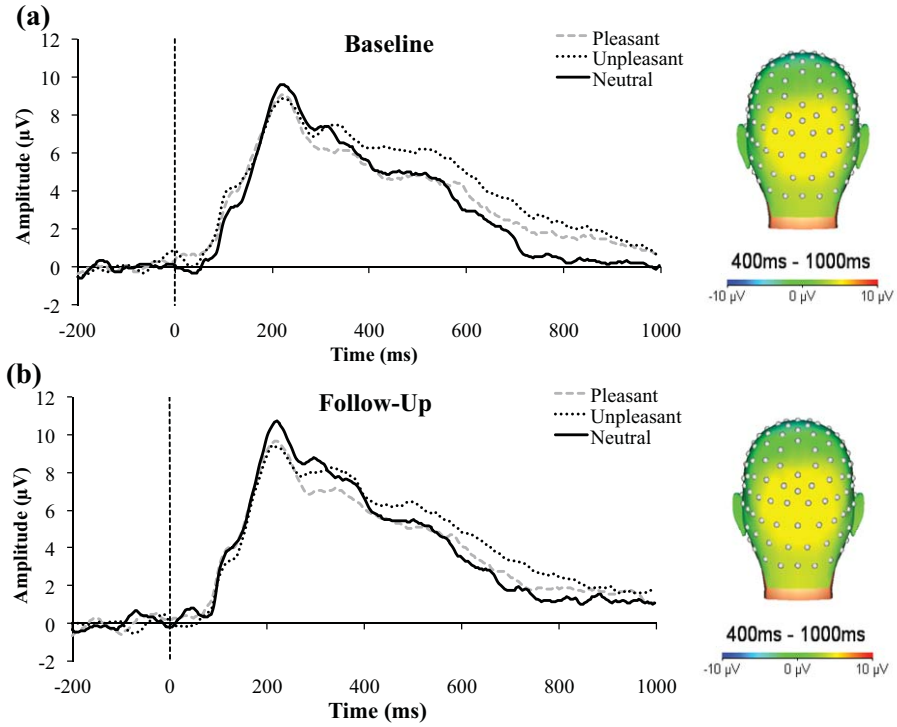

FIGURE 2 LPP activity during the emotional interrupt task. LPP activity at Pz in response to pleasant, unpleasant, and neutral images during the (a) baseline and (b) follow-up assessment. Scalp topographies reflect mean activation across conditions (pleasant, unpleasant, neutral images) at each assessment between 400–1,000 ms poststimulus

examining raw mean reaction time, the Spearman-Brown corrected odd/even reliability was .99 at both the baseline and follow-up assessments, suggesting excellent internal consistency.

3.2.3 | Reaction time—Tau

In the rmANOVA model for tau, the main effects of condition, $F(2, 54) = 2.00$, $p = .14$, $\eta_p^2 = .07$, and time, $F(1, 27) = 1.28$, $p = .27$, $\eta_p^2 = .05$, were nonsignificant, as well as the Time × Condition interaction, $F(2, 54) = 2.42$, $p = .10$, $\eta_p^2 = .08$. However, for tau, there were significant associations over time for pleasant ($r = .46$, $p = .01$), unpleasant ($r = .46$, $p = .01$), and neutral ($r = .41$, $p = .03$) images (see Figure 1b,d,f).

3.3 | ERPs

3.3.1 | LPP

In line with our hypothesis, there was a main effect of condition, $F(2, 54) = 8.78$, $p = .001$, $\eta_p^2 = .25$. Participants had greater sustained positivity to unpleasant and pleasant images compared to neutral images at both time points (Figure 2). As hypothesized, the time, $F(1, 27) = 2.09$, $p = .16$, $\eta_p^2 = .07$, and Time × Condition, $F(2, 54) = 0.39$, $p = .64$, $\eta_p^2 = .01$, effects were nonsignificant. Additionally, there were significant test-retest associations for neutral ($r = .54$, $p = .003$), unpleasant ($r = .87$, $p < .001$), pleasant ($r = .73$, $p < .001$; Figure 3) images. To demonstrate the internal reliability of the LPP, odd/even trial correlations were evaluated. Spearman-Brown corrected odd/even reliability suggests excellent internal consistency at each assessment (baseline: .93; follow-up: .92). Internal consistency also was examined as a function of image valence: baseline (neutral: .84; unpleasant: .89; pleasant: .79) and follow-up (neutral: .90; unpleasant: .82; pleasant: .57).

3.3.2 | Difference waves

The Time × Condition rmANOVA using difference wave scores (unpleasant-neutral and pleasant-neutral) revealed a significant main effect of condition, $F(1, 27) = 6.70$, $p = .02$, $\eta_p^2 = .20$. The unpleasant-neutral difference wave had greater sustained positivity compared to the pleasant-neutral difference. Neither the main effect of time, $F(1, 27) = 0.49$, $p = .49$, $\eta_p^2 = .02$, nor the Time × Condition interaction, $F(1, 27) = 0.13$, $p = .72$, $\eta_p^2 = .01$, was significant. The test-retest correlational analyses did not show significant associations over time for difference scores ($ps > .35$). The internal consistency of the pleasant-neutral difference score at baseline was modest (Spearman-Brown corrected odd/even reliability = .63), but the unpleasant-neutral difference score was poor (Spearman-Brown corrected odd/even

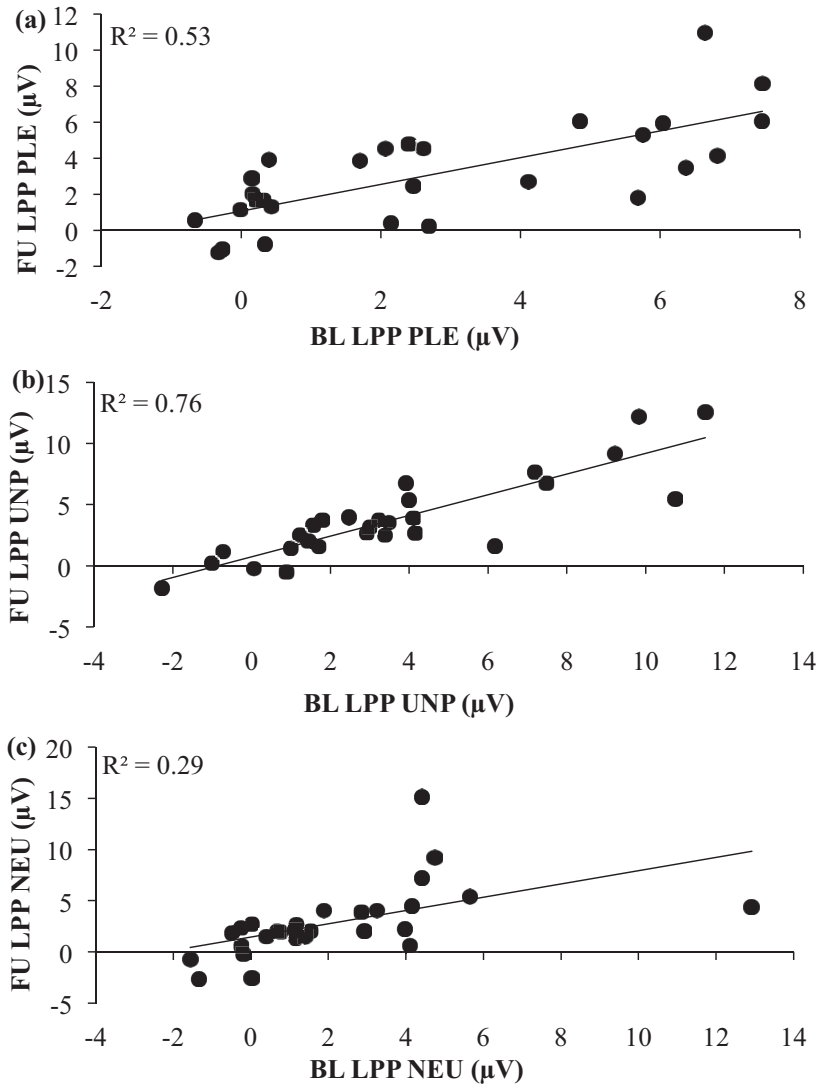

FIGURE 3 Test-retest for LPP during baseline and follow-up assessments. BL = baseline; FU = follow-up; PLE = pleasant; UNP = unpleasant; NEU = neutral

reliability = .35). For the follow-up assessment, the internal reliability of the differences scores also was poor (Spearman-Brown corrected odd/even reliability ≤ .47).

We also examined difference scores using a residual-based method. Similar to the subtraction-based difference scores, test-retest stability analyses of residual-based scores also were not significant (*p*s > .36). Additionally, at baseline the internal reliability of the unpleasant residuals (Spearman-Brown corrected odd/even reliability = .65) and pleasant residuals (Spearman-Brown corrected reliability = .64) were modest. The internal consistency was poor at the follow-up assessments (Spearman-Brown corrected reliability ≤ .52).

3.4 | Age-related differences

To test whether participant age in years (continuous measure) moderated LPP stability, we conducted an Age × Time × Condition rmANOVA. The Age × Time × Condition interaction was nonsignificant, $F(2, 52) = 0.35$, $p = .67$, $\eta_p^2 = .01$, indicating that the LPP over time did not vary as a function of participant age. Additionally, all lower-order two-way interactions were nonsignificant (*p*s > .20, $\eta_p^2 < .06$), and the main effects (time, condition) as reported above did not change when age was included in the model. Interestingly, a significant main effect of age emerged, $F(1, 26) = 17.39$, $p < .001$, $\eta_p^2 = .40$, such that younger participants showed greater positivity compared to older participants. A Pearson's product-moment correlation indicated that age was inversely correlated with the LPP amplitude across conditions and assessments ($r = -.63$, $p < .001$, Figure 4a).

As prior work has shown that the LPP tends to be maximal in occipital regions in younger individuals (Hajcak & Dennis, 2009; Kujawa et al., 2013), we conducted additional analyses testing differential activity in parietal and occipital regions. Specifically, Age × Condition × Electrode (Pz, Oz)

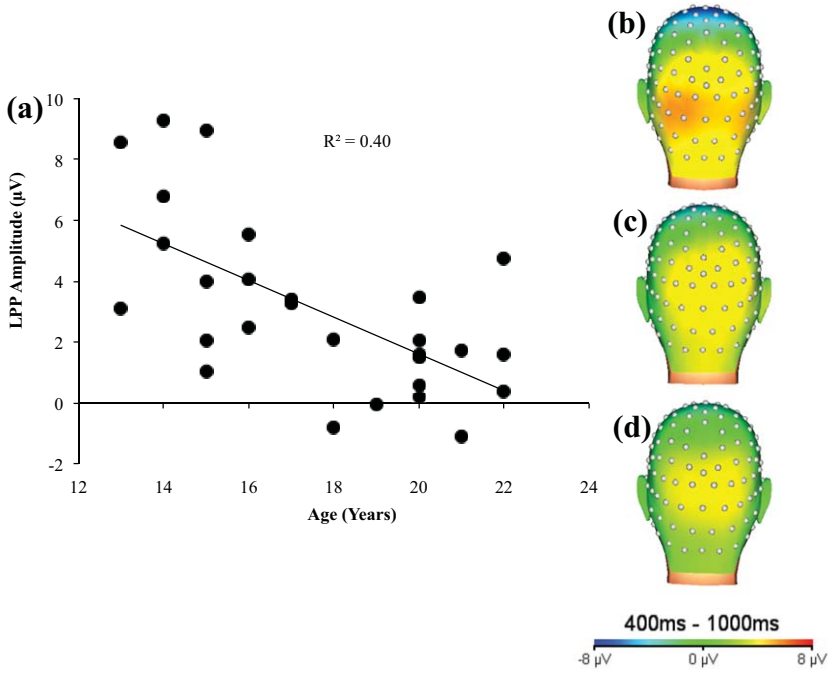

FIGURE 4 Relationship between LPP and age. (a) Average LPP amplitude across image valence (unpleasant, pleasant, neutral) and time (baseline, follow-up). Topographic map of average LPP across image conditions and time for (b) 13- to 15-year-olds ($n = 9$), (c) 16- to 18-year-olds ($n = 7$), and (d) 19- to 22-year-olds ($n = 12$)

rmANOVAs were conducted for Time 1 and Time 2. At each time point, the main effects of condition and electrode as well as all interactions were not significant ($ps > .10$, $\eta_p^2 < .10$). These null effects also reflect the similar topographical map activity for younger versus older participants (see Figure 4b–d).

3.5 | Correlational analyses

Correlational analyses for baseline and follow-up EIT reaction time and ERP indices were conducted (see Table 2). Accuracy measures were not included in these correlation tables because the variables were not linear. At both time points, mu was significantly associated across each condition. Associations among LPP amplitudes were significant across conditions at the baseline and follow-up assessment. There were no significant correlations between reaction time measures and LPP.

4 | DISCUSSION

To improve our understanding of emotion processing, the present study tested the 1-month stability of the EIT among healthy, female youth. Four principal findings emerged. First, there is evidence of an interference effect at baseline (but not at the follow-up assessment), as participants showed a slower reaction time for unpleasant and pleasant images relative to neutral images. Additionally, test-retest of behavioral measures was relatively stable over time. Second, there was greater LPP for unpleasant and pleasant images relative to neutral images, and the test-retest stability was excellent. Third, behavioral measures and LPP demonstrated strong internal consistency (odd/even reliability) across conditions; however, the internal consistency of the difference waves ranged from modest to poor. Last, younger participants exhibited greater LPP amplitudes compared to older participants across conditions. As a whole, these results support the growing literature assessing stability of the LPP over time during emotion processing.

We found no significant effects of task-irrelevant emotional images on accuracy in the current sample. These findings may be due in part to a ceiling effect, as participants performed well across conditions (average accuracy rates $> 95\%$). The lack of accuracy findings may reflect sex-related decreased variability on task performance, though previous studies of female-only early adolescent samples have demonstrated variability in behavioral outcomes (Nelson et al., 2015; Speed et al., 2015). Reaction time was assessed using an ex-Gaussian approach (Spieler et al., 1996). Mu, which is thought to reflect perception and response execution (Hohle, 1965), was slower for pleasant and unpleasant trials compared to neutral trials at baseline and is consistent with prior research demonstrating slower reaction times for emotional versus neutral images (Mitchell et al., 2006; Weinberg & Hajcak, 2011a). At the same time, this difference was not significant at the follow-up assessment. This null effect may reflect our study design, as (a)

TABLE 2 Pearson product-moment correlations among reaction time measures and ERPs at the baseline and follow-up assessment

Baseline	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1. PLE RT mu									
2. UNP RT mu	.99**								
3. NEU RT mu	.97**	.95**							
4. PLE RT tau	.30	.34	.33						
5. UNP RT tau	−.01	−.09	.02	.49**					
6. NEU RT tau	.29	.32	.10	.45*	.21				
7. PLE LPP	−.09	−.10	−.07	.02	.07	−.03			
8. UNP LPP	.06	.04	.04	.21	.21	.25	.82**		
9. NEU LPP	−.05	−.07	−.07	.10	.004	.11	.70**	.74**	
Follow-up	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
1. PLE RT mu									
2. UNP RT mu	.98**								
3. NEU RT mu	.98**	.99**							
4. PLE RT tau	.18	.29	.25						
5. UNP RT tau	.48**	.49**	.54**	.57**					
6. NEU RT tau	.34	.38*	.28	.65**	.52**				
7. PLE LPP	−.23	−.20	−.25	.16	−.23	.16			
8. UNP LPP	−.09	−.09	−.14	−.09	−.29	.09	.82**		
9. NEU LPP	−.27	−.26	−.27	.03	−.29	.01	.76**	.70**	

Note. PLE = pleasant images; UNP = unpleasant images; NEU = neutral images; RT = reaction time; LPP = late positive potential.
[a]**$p < .01$. *$p < .05$.

test-retest was conducted during a 1-month time span, and (b) the second assessment included the third and fourth viewing of the same images (during a relatively brief time window). Unfortunately, this may have unduly influenced the likelihood of instigating an interference effect at the behavioral level; yet, as our findings show, electrocortical differences persisted. Finally, task-irrelevant images did not significantly impact tau, which may not be surprising as tau indexes lapses in attention (e.g., Hervey et al., 2006). Overall, the ex-Gaussian approach to model reaction time demonstrated a preliminary interference effect, and, more broadly, the EIT showed promising psychometric properties.

Consistent with prior work (Kujawa et al., 2012; Weinberg & Hajcak, 2011a), participants exhibited greater positivity to unpleasant and pleasant images compared to neutral images across assessments. Further, the unpleasant-neutral difference score was potentiated relative to pleasant-neutral images. This suggests that unpleasant images elicited greater positivity than pleasant images when compared to neutral images, which may be a result of unpleasant images being significantly more arousing than pleasant images. The LPP was remarkably stable across conditions with large effect sizes, which is in line with previous work demonstrating strong test-retest stability of the LPP (Auerbach et al., 2016; Kujawa et al., 2013). Similar stability estimates have been shown for the P1, a component indexing semantic monitoring of emotional information, during a self-referential encoding task (Auerbach et al., 2016). By contrast, alternative approaches to probe emotion processing using fMRI may be less stable. For example, in a sample of healthy adolescents viewing fearful, happy, and neutral faces, activation in the prefrontal cortex and amygdala (two regions implicated in emotion face processing) demonstrated poor-to-modest stability (van den Bulk et al., 2013). Taken together, ERPs may be a more stable and reliable tool to detect pathophysiological mechanisms associated with emotion processing.

Test-retest reliability and internal consistency of the ERPs was assessed using three different approaches. When probing odd/even correlations and internal consistency for each valence separately, results were in the excellent range. However, for difference scores and standardized residuals, neither approach showed significant test-retest reliability. This is consistent with prior research (Kujawa et al., 2013; Levinson et al., 2017), and it is not necessarily surprising as test-retest reliability of the difference and residualized scores is often lower than the reliability of the constituent components (e.g., unpleasant, pleasant, neutral; see Meyer et al., 2017; but also see Edwards, 2001). Similarly, the internal consistency using difference scores and residualized approaches were not as strong relative to testing each valence separately. These null effects raise important questions, as ERPs tested by using change or residualized scores often rely on neutral stimuli to help interpret the amplitudes of affective stimuli (e.g., unpleasant, pleasant images). Despite this potential reliability problem, these findings (and other similar results) do not necessarily suggest that researchers should avoid difference or residualized scores. Rather, it is important to determine why the reliability may be suboptimal and if this may impact reproducibility. With regard to our study, the LPP amplitudes are positively correlated, which may in part account for the poor reliability (Edwards, 2001). In other research there may be very clear mandates as to why it would be important to use data reduction techniques. Thus, rather than having a blanket "should" or "should not" statement about the use of difference or residualized scores, we believe it is more important to (a) be mindful of the EEG/ERP psychometric properties (and account for potential reliability issues), (b) tailor the data analytic approach to the central research question, and (c) determine whether findings can be replicated (even in the absence of strong test-retest reliability of difference or residualized scores).

The study also sought to address important developmental issues, particularly as it related to determining whether age impacts LPP activity. Results indicated that younger age was associated with greater LPP activity in parietal regions across conditions. These findings support prior work testing age-related electrocortical effects following emotion-based images. Specifically, MacNamara and colleagues (2016) tested age effects on the LPP during the presentation of emotional faces among youth aged 7 to 19 years. Relative to younger children, older participants showed decreased positivity following emotional but not neutral stimuli (geometric shapes). Age-related decreases in the LPP also have been demonstrated in other tasks; when asked to attend to either pain or nonpain cues in images, adolescents exhibited a potentiated LPP compared to young adults (Mella, Studer, Gilet, & Labourie-Vief, 2012). Additionally, age-related differences in the LPP scalp distribution were explored, and results indicated that LPP activity did not vary as a function of electrode site (parietal vs. occipital) for younger versus older youth (see Figure 4b–d). At the same time, this conflicts with prior work in younger individuals that has often shown the LPP is maximal in occipital regions (e.g., Hajcak & Dennis, 2009; Kujawa et al., 2013). Together, these findings underscore the importance of testing whether age influences LPP activity across development, as this may have important implications for interpreting ERP effects.

Our results should be interpreted in light of several limitations. First, prior research using IAPS stimuli has demonstrated greater electrophysiological reactivity in females relative to males during passive viewing of unpleasant stimuli (Lithari et al., 2010), which underscores the importance of testing sex-specific effects. However, as the current study only included female participants, we cannot determine whether our effects generalize to males. Second, prior research has demonstrated that pubertal status influences electrophysiological responses in youth (e.g., processing of emotional faces, fear-potentiated startle; Ferri, Bress, Eaton, & Proudfit, 2014; Schmitz, Grillon, Avenevoli, Cui, & Merikangas, 2014). At the same time, the present study did not assess pubertal status, and thus we cannot determine how pubertal status affects LPP stability. Third, this study examined the stability of ERPs over a 1-month period, and, consequently, it is important to confirm these effects over a longer period of time. Fourth, IAPs images were used to standardize the emotional stimuli across participants. However, the current study did not obtain subjective arousal ratings for each image, which may have facilitated an enhanced interpretation of our ERP effects. Additionally, images were presented twice during each administration of the EIT. Notably, we used the same paradigm as other published studies (e.g., Kujawa et al., 2012, 2013; Nelson et al., 2015; Speed et al., 2015; Weinberg & Hajcak, 2011a), which then allowed us to compare our behavioral and ERP effects to the extant literature. At the same time, reviewing images multiple times during the trial may have an unmeasured impact on core psychometric properties. Last, a number of factors, including menstrual cycle and circadian rhythm, may influence ERP amplitude (Polich & Kok, 1995), and thus future research should account for these potential effects.

In summary, prior research has shown that the LPP is a stable and reliable marker of processing emotional words (Auerbach et al., 2016) and images (Kujawa et al., 2013). Toward the goal of identifying indicators of emotion processing, the current findings provide further support for the stability and reliability of the LPP over time in healthy youth. Ultimately, an improved understanding of electrophysiological correlates of emotion processing may lead to insight regarding the onset and maintenance of debilitating psychiatric symptoms.

ACKNOWLEDGMENTS

Randy P. Auerbach was partially supported through funding from the National Institute of Mental Health (NIMH) K23MH097786, the Rolfe Fund, and the Tommy Fuss Fund. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or NIMH. Dr. Mittal is a consultant for Takeda Pharmaceuticals. No other authors report any conflicts of interest.

REFERENCES

Auerbach, R. P., Bondy, E., Stanton, C. H., Webb, C. A., Shankman, S. A., & Pizzagalli, D. A. (2016). Self-referential processing in adolescents: Stability of behavioral and ERP markers. *Psychophysiology*, *53*(9), 1398–1406. https://doi.org/10.1111/psyp.12686

Auerbach, R. P., Stanton, C. H., Proudfit, G. H., & Pizzagalli, D. A. (2015). Self-referential processing in depressed adolescents: A high-density event-related potential study. *Journal of Abnormal Psychology*, *124*(2), 233–245. https://doi.org/10.1037/abn0000023

Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, *128*(1), 32–55.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory manual* (2nd ed.). San Antonio, TX: Psychological Corporation.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using STATA* (Vol. *5*). College Station, TX: Stata Press.

Cassidy, S. M., Robertson, I. H., & O'Connell, R. G. (2012). Retest reliability of event-related potentials: Evidence from a variety of paradigms. *Psychophysiology*, *49*(5), 659–664. https://doi.org/10.1111/j.1469-8986.2011.01349.x

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cuthbert, B. N., Schupp, H. T., Bradley, M. M., Birbaumer, N., & Lang, P. J. (2000). Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biological Psychology*, *52*(2), 95–111.

Dolan, R. J. (2002). Emotion, cognition, and behavior. *Science*, *298* (5596), 1191–1194. https://doi.org/10.1126/science.1076358

Dolcos, F., & Cabeza, R. (2002). Event-related potentials of emotional memory: Encoding pleasant, unpleasant, and neutral pictures. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(3), 252–263. https://doi.org/10.3758/CABN.2.3.252

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, *4*(3), 265–287.

Ferri, J., Bress, J. N., Eaton, N. R., & Proudfit, G. H. (2014). The impact of puberty and social anxiety on amygdala activation to faces in adolescence. *Developmental Neuroscience*, *36*(3–4), 239–249. https://doi.org/10.1159/000363736

Foti, D., & Hajcak, G. (2008). Deconstructing reappraisal: Descriptions preceding arousing pictures modulate the subsequent neural response. *Journal of Cognitive Neuroscience*, *20*(6), 977–988. https://doi.org/10.1162/jocn.2008.20066

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: Evidence from temporal-spatial PCA. *Psychophysiology*, *46*(3), 521–530. https://doi.org/10.1111/j.1469-8986.2009.00796.x

Hajcak, G., & Dennis, T. A. (2009). Brain potentials during affective picture processing in children. *Biological Psychology*, *80*(3), 333–338. https://doi.org/10.1016/j.biopsycho.2008.11.006

Hajcak, G., & Olvet, D. M. (2008). The persistence of attention to emotion: Brain potentials during and after picture presentation. *Emotion*, *8*(2), 250–255. https://doi.org/10.1037/1528-3542.8.2.250

Hajcak, G., MacNamara, A., & Olvet, D. M. (2010). Event-related potentials, emotion, and emotion regulation: An integrative review. *Developmental Neuropsychology*, *35*(2), 129–155. https://doi.org/10.1080/87565640903526504

Hajcak, G., Weinberg, A., MacNamara, A., & Foti, D. (2012). ERPs and the study of emotion. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 441–474). New York, NY: Oxford University Press.

Hervey, A. S., Epstein, J. N., Curry, J. F., Tonev, S., Eugene Arnold, L., Keith Conners, C., … Hechtman, L. (2006). Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychology*, *12*(2), 125–140.

Hess, U., Arslan, R., Mauersberger, H., Blaison, C., Dufner, M., Denissen, J. J., & Ziegler, M. (2017). Reliability of surface facial electromyography. *Psychophysiology*, *54*(1), 12–23. https://doi.org/10.1111/psyp.12676

Hohle, R. H. (1965). Inferred components of reaction times as functions of foreperiod duration. *Journal of Experimental Psychology*, *69*(4), 382–386.

Insel, P. A., Amara, S. G., & Blaschke, T. F. (2015). Introduction to the theme "precision medicine and prediction in pharmacology." *Annual Review of Pharmacology and Toxicology*, *55*, 11–14. https://doi.org/10.1146/annurev-pharmtox-101714-123102

Kujawa, A., Klein, D. N., & Hajcak, G. (2012). Electrocortical reactivity to emotional images and faces in middle childhood to early adolescence. *Developmental Cognitive Neuroscience*, *2*(4), 458–467. https://doi.org/10.1016/j.dcn.2012.03.005

Kujawa, A., Klein, D. N., & Proudfit, G. H. (2013). Two-year stability of the late positive potential across middle childhood and adolescence. *Biological Psychology*, *94*(2), 290–296. https://doi.org/10.1016/j.biopsycho.2013.07.002

Lang, P. J. (1984). Cognition in emotion: Concept and action. In C. E. Izard, J. Kagan, & R. B. Zajonc (Eds.). *Emotions, cognition, and behavior* (pp. 192–226). Cambridge, UK: Cambridge University Press.

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (2008). International Affective Picture System (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-8. University of Florida, Gainesville, FL.

Levinson, A. R., Speed, B. C., Infantolino, Z. P., & Hajcak, G. (2017). Reliability of the electrocortical response to gains and losses in the doors task. *Psychophysiology*, *54*(4), 601–607. https://doi.org/10.1111/psyp.12813

Lithari, C., Frantzidis, C. A., Papadelis, C., Vivas, A. B., Klados, M. A., Kourtidou-Papadeli, C., … Bamidis, P. D. (2010). Are

females more responsive to emotional stimuli? A neurophysiological study across arousal and valence dimensions. *Brain Topography*, *23*(1), 27–40. https://doi.org/10.1007/s10548-009-0130-5

MacNamara, A., Vergés, A., Kujawa, A., Fitzgerald, K. D., Monk, C. S., & Phan, K. L. (2016). Age-related changes in emotional face processing across childhood and into young adulthood: Evidence from event-related potentials. *Developmental Psychobiology*, *58*(1), 27–38. https://doi.org/10.1002/dev.21341

Massidda, D. (2013). Retimes: Reaction time analysis. R package (Version 0.1–2) [Computer software]. Retrieved from https://rdrr.io/cran/retimes/

Mella, N., Studer, J., Gilet, A. L., & Labouvie-Vief, G. (2012). Empathy for pain from adolescence through adulthood: An event-related brain potential study. *Frontiers in Psychology*, *3*, 501. https://doi.org/10.3389/fpsyg.2012.00501

Meyer, A., Lerner, M. D., De Los Reyes, A., Laird, R. D., & Hajcak, G. (2017). Considering ERP difference scores as individual difference measures: Issues with subtraction and alternative approaches. *Psychophysiology*, *54*(1), 114–122. https://doi.org/10.1111/psyp.12664

Mitchell, D. G., Richell, R. A., Leonard, A., & Blair, R. J. R. (2006). Emotion at the expense of cognition: Psychopathic individuals outperform controls on an operant response task. *Journal of Abnormal Psychology*, *115*(3), 559–666. https://doi.org/10.1037/0021-843X.115.3.559

Nelson, B. D., Perlman, G., Hajcak, G., Klein, D. N., & Kotov, R. (2015). Familial risk for distress and fear disorders and emotional reactivity in adolescence: An event-related potential investigation. *Psychological Medicine*, *45*(12), 2545–2556. https://doi.org/10.1017/S0033291715000471

Nunnally, J. C., Bernstein, I. H., & Berge, J. M. T. (1967). *Psychometric theory* (Vol. *226*). New York, NY: McGraw-Hill.

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology*, *77*(3), 247–265. https://doi.org/10.1016/j.biopsycho.2007.11.006

Olvet, D. M., & Hajcak, G. (2009). Reliability of error-related brain activity. *Brain Research*, *1284*, 89–99. https://doi.org/10.1016/j.brainres.2009.05.079

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology*, *41*(2), 103–146. https://doi.org/10.1016/0301-0511(95)05130–9

Schmitz, A., Grillon, C., Avenevoli, S., Cui, L., & Merikangas, K. R. (2014). Developmental investigation of fear-potentiated startle across puberty. *Biological Psychology*, *97*, 15–21. https://doi.org/10.1016/j.biopsycho.2013.12.002

Schupp, H. T., Cuthbert, B. N., Bradley, M. M., Cacioppo, J. T., Ito, T., & Lang, P. J. (2000). Affective picture processing: The late positive potential is modulated by motivational relevance. *Psychophysiology*, *37*(2), 257–261.

Sheehan, D. V., Sheehan, K. H., Shytle, R. D., Janavs, J., Bannon, Y., Rogers, J. E., … Wilkinson, B. (2010). Reliability and validity of the Mini International Neuropsychiatric Interview for children and adolescents (MINI-KID). *Journal of Clinical Psychiatry*, *71*(3), 313–326. https://doi.org/10.4088/JCP.09m05305whi.

Speed, B. C., Nelson, B. D., Auerbach, R. P., Klein, D. N., & Hajcak, G. (2016). Depression risk and electrocortical reactivity during self-referential emotional processing in 8 to 14 year-old girls. *Journal of Abnormal Psychology*, *125*(1) 26–39. https://doi.org/10.1037/abn0000173

Speed, B. C., Nelson, B. D., Perlman, G., Klein, D. N., Kotov, R., & Hajcak, G. (2015). Personality and emotional processing: A relationship between extraversion and the late positive potential in adolescence. *Psychophysiology*, *52*(8), 1039–1047. https://doi.org/10.1111/psyp.12436

Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(2), 461–479.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science*, *150*(3700), 1187–1188. https://doi.org/10.1126/science.150.3700.1187

Tenke, C. E., Kayser, J., Pechtel, P., Webb, C. A., Dillon, D. G., Goer, F., … Parsey, R. (2017). Demonstrating test-retest reliability of electrophysiological measures for healthy adults in a multisite study of biomarkers of antidepressant treatment response. *Psychophysiology*, *54*(1), 34–50. https://doi.org/10.1111/psyp.12758

van den Bulk, B. G., Koolschijn, P. C. M., Meens, P. H., van Lang, N. D., van der Wee, N. J., Rombouts, S. A., … Crone, E. A. (2013). How stable is activation in the amygdala and prefrontal cortex in adolescence? A study of emotional face processing across three measurements. *Developmental Cognitive Neuroscience*, *4*, 65–76. https://doi.org/10.1016/j.dcn.2012.09.005

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, *7*(3), 424–465.

Weinberg, A., & Hajcak, G. (2010). Beyond good and evil: The time-course of neural activity elicited by specific picture content. *Emotion*, *10*(6), 767–782. https://doi.org/10.1037/a0020242.

Weinberg, A., & Hajcak, G. (2011a). The late positive potential predicts subsequent interference with target processing. *Journal of Cognitive Neuroscience*, *23*(10), 2994–3007. https://doi.org/10.1162/jocn.2011.21630

Weinberg, A., & Hajcak, G. (2011b). Longer term test–retest reliability of error-related brain activity. *Psychophysiology*, *48*(10), 1420–1425. https://doi.org/10.1111/j.1469-8986.2011.01206.x

Whelan, R. (2008). Effective analysis of reaction time data. *Psychological Record*, *58*(3), 475–482.

How to cite this article: Bondy E, Stewart JG, Hajcak G, et al. Emotion processing in female youth: Testing the stability of the late positive potential. *Psychophysiology*. 2018;55:e12977. https://doi.org/10.1111/psyp.12977