Reliability of Reward- and Error-Related Brain Activity in Early Childhood

Aline K. Szenczy, B.S.[1], **Amanda R. Levinson, Ph.D.**[1], **Greg Hajcak, Ph.D.**[2], **Kristin Bernard, Ph.D.**[1], **Brady D. Nelson, Ph.D.**[1]

[1]Department of Psychology, Stony Brook University

[2]Department of Psychology, Florida State University

Abstract

Event-related potential (ERP) measures of reward- and error-related brain activity have emerged as potential biomarkers of risk for the development of psychopathology. However, the psychometric properties of reward- and error-related brain activity have been primarily investigated in adolescents and adults. It is critical to also establish the reliability of ERPs in younger children, particularly if they are used as individual difference measures of risk during key developmental periods. The present study examined the reliability of the reward positivity (RewP) and error-related negativity (ERN) among 80 children (M_{age} = 6.9 years-old; 50% female). Participants completed the doors, flanker, and go/no-go tasks twice, separated by approximately 8 months, while electroencephalography (EEG) was recorded. Results indicated that the RewP demonstrated strong internal consistency and test-retest reliability. The ERN also demonstrated strong internal consistency, but test-retest reliability was only significant for the ERN measured during the flanker task and not the go/no-go task. These results are largely consistent with reported psychometric properties of reward- and error related ERPs in adolescents and adults, suggesting that the ERN and RewP may be appropriate biomarkers of individual differences in populations ranging from early childhood to adulthood.

Keywords

Aberrations in neural systems involved in reward processing and error detection have been linked to various forms of psychopathology. For example, a decreased neural response to rewards has been associated with depression (Keren et al., 2018), and an increased neural response to errors has been associated with anxiety and obsessive-compulsive and related disorders (Moser et al., 2013; Riesel, 2019). Many forms of psychopathology have

Corresponding Author: Aline K. Szenczy, aline.szenczy@stonybrook.edu, Phone: (631) 632-6000, Address: 100 Nicolls Rd, Stony Brook, NY 11794.

their initial onset during childhood and adolescence (Beesdo et al., 2009; Merikangas et al., 2009). Therefore, it is critical to establish whether neural risk factors can be reliably measured during early childhood.

Event-related potentials (ERPs) have often been employed to examine reward- and error-related neural processing in youth. The reward positivity (RewP) is a positive-going ERP component that is elicited approximately 300ms following the presentation of reward feedback that is reduced or absent following loss feedback (Foti et al., 2011; Proudfit, 2015). The RewP has been associated with neural activation in several brain regions, including the anterior cingulate cortex, medial prefrontal cortex, and ventral striatum (Becker et al., 2014; Carlson et al., 2011). The feedback negativity (FN) is a negative-going component that is elicited within a similar time window but is larger following loss feedback. To better isolate reward-related brain activity, the RewP is often examined as the difference between the RewP and FN (i.e., RewP-FN). A number of studies have indicated that a more blunted RewP is associated with depression (for a summary, see Kujawa & Burkhouse, 2017), including in preschool-aged children (Belden et al., 2016). There is also growing evidence that the RewP indexes individual differences in *risk* for the development of depression. Specifically, in children a more blunted RewP has been associated with maternal history of depression (Kujawa et al., 2014), and in adolescents a more blunted RewP has been shown to prospectively predict the onset of major depressive disorder and depressive symptoms two years later (Bress et al., 2013; Nelson et al., 2016).

The error-related negativity (ERN) is a negative-going ERP component that occurs approximately 50ms following the commission of an error (Hajcak, 2012). The correct response negativity (CRN) is a similar negative-going component that occurs during the same time window as the ERN and is elicited following correct responses. To better isolate error-related brain activity, the ERN if often examined as the difference between the ERN and CRN (i.e., ERN-CRN). A larger (i.e., more negative) ERN has been associated with multiple forms of psychopathology, including generalized anxiety disorder and obsessive-compulsive disorder (Carrasco et al., 2013; Weinberg et al., 2010). Moreover, a larger ERN has been associated with trait-like risk factors for anxiety disorders, such as anxious apprehension and behavioral inhibition (McDermott et al., 2009; Moser et al., 2013). In both children and adolescents, a larger ERN has been shown to prospectively predict the development of anxiety disorders (Meyer et al., 2015; Meyer, Nelson, et al., 2018). However, evidence also suggests that the relationship between the ERN and anxiety symptoms across development may differ among clinical and non-clinical populations. For example, among clinically anxious children, the relationship between anxiety and the ERN does not differ as a function of age (Meyer, 2017). Conversely, among non-clinically anxious children, the relationship between the ERN and anxiety differs across development, such that in young children anxiety is related to a more blunted ERN but in older children, anxiety is related to a larger (i.e., more negative) ERN (Meyer, 2017). Altogether, this research highlights the potential clinical utility of the RewP and ERN as neural biomarkers of risk that may help identify individuals who are likely to develop anxiety, depressive, and obsessive-compulsive and related disorders.

If the RewP and ERN are to be used as individual difference measures of risk, it is critical to first establish that they have adequate psychometric properties (Hajcak et al., 2019). In adults, the RewP and FN have demonstrated excellent internal consistency and strong, short-term test-retest reliability (Ethridge & Weinberg, 2018; Levinson et al., 2017; Marco-Pallares et al., 2011; Segalowitz et al., 2010). In addition, in adults the ERN has demonstrated acceptable internal consistently after as few as 7 errors (Meyer et al., 2013; Olvet & Hajcak, 2009; Pontifex et al., 2010), and a recent meta-analysis indicated that the overall internal consistency of the ERN for 8 trials is slightly below acceptability standards (Clayson, 2020). However, these studies also indicated that task differences may account for some of the variability present in ERN reliability estimates (Clayson, 2020; Meyer et al., 2013). The ERN has also demonstrated moderate to strong test-retest reliability across weeks (Larson et al., 2010) and up to 2.5 years (Weinberg & Hajcak, 2011).

Despite critical developmental changes in reward circuitry from childhood to adolescence, initial evidence suggests that the psychometric properties of the RewP and FN in child and adolescent samples are largely comparable to those of adults. For example, one investigation measured the RewP in children, adolescents, and adults and found that it was similar across groups (Lukie et al., 2014). Additionally, in study of 8 to 13 year-old children, results showed that two-year test-retest reliability of the RewP and FN was moderate (Bress et al., 2015). Evidence also suggests that internal consistency of the RewP and FN is not moderated by age (Ethridge & Weinberg, 2018; Luking et al., 2017). In a separate investigation spanning late childhood to middle adolescence, the RewP and FN demonstrated more robust test-retest reliability during early to middle adolescence than late childhood to early adolescence (Kujawa et al., 2018). Burani and colleagues (2019) also conducted a longitudinal investigation of children and adolescents and found that the RewP increased from late childhood to early adolescence, but not late adolescence, suggesting that the RewP exhibits greater temporal stability during late adolescence than in childhood.

There have only been a limited number of studies that have examined the psychometric properties of the ERN in children. For example, one investigation of 8 to 11 year-old children found that the ERN demonstrated excellent internal consistency in as few as 6 trials using the flanker task (Pontifex et al., 2010). Similarly, a separate investigation of 8 to 13 year-old children found that the ERN achieved an acceptable internal consistency in 8 trials using a flanker task, but internal consistency never reached the acceptable range using a go/no-go task (Meyer et al., 2014). Additionally, the ERN and CRN demonstrated moderate two-year test-retest reliability using the flanker task (Meyer et al., 2014). Although studies suggest the ERN can be reliably elicited in young children, some evidence suggests that the ERN may change across development (Meyer, Carlton, et al., 2018), such that it increases following pubertal onset (Davies et al., 2004; Meyer et al., 2012).

Several studies have also examined the psychometric properties of the ΔRewP and ΔERN. It is important to note that there should be different expectations for assessing the reliability of difference scores. Prior studies have demonstrated the ΔRewP often has lower reliability than its constituent measures in both children and adults (Bress et al., 2015; Levinson et al., 2017; Marco-Pallares et al., 2011). The literature assessing the reliability of the ΔERN is mixed, with some studies indicating weaker reliability, particularly in children

(e.g., Meyer et al., 2014), and others showing more robust reliability (Weinberg & Hajcak, 2011). The reliability of a difference score, such as the RewP and ERN, is dependent upon the reliability, variability, and correlation between its components. Thus, it is expected that if the individual measures that make up a difference score are highly correlated, have poor reliability, or have different variances, which is often the case, the reliability of the difference score may be negatively affected (Furr & Bacharach, 2013; Meyer et al., 2017). Indeed, the lower reliability of a difference score places a limit on how much reliable variance can relate to another individual difference measure. Nonetheless, it is possible that the reliable portion of variance in the difference score might be largely related to another measure (Patrick et al., 2019). This scenario would be similar to one in which a highly reliable measure only has a small portion of variance that is related to another measure. Therefore, despite the limitations of difference score reliability, the RewP and ERN may still be promising candidates in the search for clinically meaningful biomarkers.

Notwithstanding initial evidence that reward and error-related neural markers may be reliable in children, the child ERP psychometric literature is limited. Because most forms of depression and anxiety commonly have their onset in childhood and adolescence (Galvan, 2010; Tamnes et al., 2013), more research is needed to understand the internal and test-retest reliability of neural biomarkers in key developmental periods, particularly early childhood. Moreover, the extant literature examining the psychometrics of the ERN and RewP has largely utilized adolescent and adult samples, and it remains unclear the extent to which these properties extend to young children.

The present study examined the psychometric properties of the RewP and ERN in early childhood. Eighty 6 to 8 year-old children were assessed using the doors, flankers, and go/no-go tasks at two time points, separated by 8 months, while electroencephalography (EEG) was recorded. Internal consistency and test-retest reliability were assessed with the aims of 1) characterizing the psychometric properties of the RewP and ERN in an early childhood sample and 2) comparing the psychometric properties of the ERN across tasks (flankers and go/no-go). We aimed to assess whether these ERPs achieve acceptable psychometric properties in early childhood, a necessary prerequisite for determining if these measures may serve as clinically meaningful neural markers of risk for psychopathology.

Method

Participants

Participants included 80 6 to 8 year-old children (M_{age} = 6.9 years-old; SD_{age} = 0.6) and their mothers (M_{age} = 37.5; SD_{age} = 5.7) as a part of a randomized controlled trial that compared a parenting intervention to a waitlist control. Participants were recruited using commercial mailing lists, locally posted flyers, online and social media postings, advertisements created by a data-driven clinical trial technology company (e.g., Trialspark), word of mouth, and a psychology clinic. Dyads were invited to participate in the study if they had a child between the ages of 6 and 8 with no history of developmental delays and an English-speaking mother who identified as the primary caregiver of the child participant. Additionally, mothers had to endorse a history of depression and/or anxiety disorder based on screening modules from the Mini International Neuropsychiatric Interview (Lecrubier et

al., 1997) or low authoritative and/or high authoritarian parenting based on subscales of the Parenting Styles and Dimensions Questionnaire (Olivari et al., 2013). Forty children (50%) were female. Sixty-one (76.3%) children were White, eight (10.0%) were Biracial, three (3.8%) were African American, and three (3.8%) were Asian, and five (6.3%) identified as "Other." Additionally, ten children (12.5%) were Hispanic.

Procedure

The baseline assessment (Time 1) was conducted before random assignment to treatment group, and the follow-up assessment (Time 2) was conducted upon the completion of the parenting intervention (active intervention group) or approximately 8 months (mean time between visits= 7.7 months; SD = 2.0 months) following the initial baseline assessment (the waitlist group). Forty participants were assigned to each treatment condition. Procedures at both assessments were identical and task order was counterbalanced between the doors and go/no-go task. The flanker task was completed last since some evidence suggests that this task may be more challenging for young child participants (Davies et al., 2004; Torpey et al., 2012). All tasks were presented on a 21-inch computer monitor using Presentation software (Neurobehavioral Systems, Inc.). Study materials are available through the Open Science Framework (Szenczy et al., 2021).

Doors task.—The doors task is a forced-choice guessing task that has been shown to elicit the RewP (Proudfit, 2015). On each trial, participants were presented with two doors and told to choose the door they think has a prize behind it using the left and right mouse buttons. Once the choice was made, a fixation cross was presented for an intertrial interview of 1000ms. Following this fixation cross, either a green up arrow was displayed indicating a gain of $0.50 or a red down arrow was presented indicating a loss of $0.25 for 2000ms. A fixation cross was again presented on the screen for 1,500ms followed by the words "Click for next round" until the participant clicked the left or right mouse button. Participants were told that they could win up to $15 for the completion of the task. Prior to the experimental trials, children were presented with practice trials to ensure that they understood the task directions. All participants received 50% gain feedback (30 trials) and 50% loss feedback (30 trials) for a total of 60 trials. Participants received $8 for the completion of this task.

Flanker task.—An arrowhead version of the flanker task was used to elicit the ERN. Participants were shown five horizontal arrowheads for 200ms, followed by a fixation cross that was presented between 2,300 and 2,800ms. Half of the trials displayed compatible arrows (">>>>>" or "<<<<<") and half displayed incompatible arrows ("<<><<" or ">><>>"). Participants were instructed to press the right mouse button if the center arrow was facing right and to press the left mouse button if the center arrow was facing left. Participants were also told to respond as accurately and as fast as possible. The first 30 trials were practice trials where the experimenter assessed for task understanding. The actual task consisted of 11 participant-initiated blocks of 30 trials (330 trials total). Between blocks, participants received performance-based feedback on the computer monitor. If performance was 75% correct or lower, participants received the feedback "Please try to be more accurate" on the monitor; if performance was above 90% correct, participants received

the feedback "Please try to respond faster." Otherwise, participants received the feedback "You're doing a great job." The task was programmed to end after the child made 20 errors.

Go/no-go task.—A go/no-go task was also used to elicit the ERN. Participants were presented with a green triangle and instructed to press the left mouse button when they saw a vertically aligned upward facing triangle on the monitor (go stimulus) but to inhibit any response when the triangle was presented in any other orientation (no-go stimuli). The stimuli were presented in four possible orientations and were shown on the monitor for a duration of 1,200ms across 4 blocks of 60 trials (240 trials total). Each block contained: 60% vertically aligned and upward facing triangles, 20% vertically aligned downward facing triangles, 10% left tilted triangles, and 10% right tilted triangles. Following each stimulus, a fixation cross was presented on the monitor between 300 and 800ms. Three practice blocks were presented prior to the experimental task: the first block to explain the task and present the stimuli (8 trials), the second block with thumbs up and thumbs down feedback following the participant's response (20 trials), and the third block without feedback to ensure task understanding (30 trials). The experimenter reminded participants of the importance of response speed and directions between blocks.

Psychophysiological recording and data reduction.—Continuous electroencephalogram (EEG) recordings were collected using a 34-electrode elastic cap configured according to the 10/20 system, using the ActiveTwo BioSemi System (BioSemi, Amsterdam, The Netherlands). Two electrodes were also placed on the right and left mastoids and four facial electrodes were placed around the eyes: two placed approximately 1cm away from the outer edge of the right and left eyes and two electrodes placed 1cm above and below the right eye. Facial electrodes were used to measure eyes movements and eye blinks. The EEG signal was pre-amplified at the electrode site to improve the signal-to-noise ratio with a gain of one. The data were digitized at 24-bit resolution with a sampling rate of 1024 Hz using a low-pass fifth-order sinc filter with a half-power cutoff of 204 Hz. Each active electrode was measured online with respect to a common mode sense active electrode generating a monopolar channel. BrainVision Analyzer 2 (Brain Products, Munich, Germany) was used for all offline analysis. All data were referenced to the average of the left and right mastoid channels and bandpass filtered with high and low cutoffs of 0.1 and 30 Hz, respectively, were applied. Eyeblink and ocular corrections were conducted using the Gratton and Coles method (Gratton et al., 1983). A semiautomatic procedure was used for all data to correct for artifacts. Data from individual channels were automatically rejected if a voltage step of more than 50.0μV between sample points or a voltage difference of 300.0 μV within a trial existed. In addition, data were identified as artifacts if a voltage difference of less than 0.50 μV within 100-ms intervals was present. Finally, visual inspection of the data was conducted to detect and reject any remaining artifacts.

EEG data from the doors task was segmented for each trial beginning 200ms before feedback onset and continuing for 800ms (i.e., 1000 ms in total). A 200ms window from −200 to 0 ms prior to response onset served as the baseline that was subtracted from all data points. The FN and RewP were evaluated as the average activity to gain and loss feedback from 325–425ms at electrode FCz because the RewP was maximal at this time window at

Time 1 and we wanted to keep the time window consistent across time points. The RewP was quantified as the difference between the ERP response gains and losses (i.e., gain-loss). The flanker and go/no-go data were segmented from 500ms seconds before the response to 1000ms after stimulus onset and 500ms before stimulus onset to 800ms after stimulus onset, respectively. A 200ms window from −500 to-30ms prior to response onset served as the baseline for both tasks. The ERN and CRN were evaluated as the average activity on error and correct trials, respectively, from −50 to 100ms after response at electrode Cz. Behavioral measures for both tasks include accuracy and reaction time (RTs) for error and correct trials. The ERN was quantified as the difference between the ERP response error and correct (i.e., error-correct).

Data analysis.—At Time 1, 5 participants were excluded from analyses due to poor accuracy (i.e., less than 55%) and 6 for poor data quality. Ten participants did not complete one of the three tasks and 1 participant did not complete any of the tasks. At Time 2, 5 participants were excluded because they did not make enough errors (i.e., fewer than 4) and 4 for poor data quality. Seven participants did not complete the Time 2 visit. One individual with a seizure disorder was excluded from all Time 1 and 2 task analyses. In total, 77, 69, and 67 individuals completed the Time 1 doors, flankers, and go/no-go tasks, respectively, and 71, 70, and 65 completed the T2 doors, flankers, and go/no-go tasks, respectively.

Statistical analyses were conducted using IBM SPSS Statistics, Version 26.0 (Armonk, NY, USA). A 2 (trial type) x 2 (time) repeated measure analysis of variance (ANOVA) was used to assess for differences in mean ERP responses between gain and loss trials (for the doors task) and error and correct trials (for the flanker and go/no-go tasks), time, or their potential interaction. Split-half reliability was examined by calculating the correlation between averages based on odd- and even-numbered trials, corrected using the Spearman-Brown prophecy formula (Nunnally et al., 1967). In addition, the overall internal reliability of the RewP and ERN difference scores was estimated using an adjusted formula (Furr & Bacharach, 2013). Test-retest reliability was evaluated using Pearson's *r*. All participants had a minimum of 14 gain and 18 loss trials in the doors task and 4 errors and 39 correct trials in the flankers and go/no-go tasks. All psychometric analyses utilized all trials (30 gain and 30 loss trials) for the doors task and the first 50 error and correct trials for the flankers and go/no-go tasks. Given that research indicates that a reliable ERN can be achieved in as little as 6–8 error trials in children (Meyer et al., 2013), 50 trials was determined as a conservative cutoff for the current analyses.

Results

Preliminary Analyses

The intervention did not impact error- or reward-related neural activity (see Supplemental Materials). Therefore, all analyses were collapsed across both groups.

Task Effects

Table 1 shows the ERP and behavioral data for the flanker and go/nogo tasks at Time 1 and Time 2. Figure 1 depicts the doors, flankers, and go/no-go task waveforms and

scalp distributions. The neural response to gain feedback was more positive than the neural response to loss feedback, $F(1,68) = 16.43$, $p < .001$. Although the effect of outcome did not vary as a function of time, $F(1,68) = .08$, $p = .773$, there was a significant main effect of time, $F(1,68) = 13.50$, $p < .001$, such that both the FN and RewP were more positive at Time 2, $t(68) = -3.15$, $p = .002$, $t(68) = -2.85$, $p = .006$, respectively. In the flanker task, the neural response to errors was more negative than the neural response to correct responses, $F(1,60) = 10.50$, $p = .002$. The effect of outcome did not vary as a function of time, $F(1,60) = .03$, $p = .870$, and there was no significant main effect of time, $F(1,60) = .84$, $p = .364$. In the go/no-go task, the neural response to errors was more negative than the neural response to correct responses, $F(1,53) = 69.83$, $p < .001$. The effect of outcome did not vary as a function of time, $F(1,53) = 1.76$, $p = .191$, and there was no significant main effect of time, $F(1,53) = .57$, $p = .455$.

Internal Consistency

The RewP and FN achieved acceptable internal consistency at both Time 1 and Time 2 (Table 2; *r*'s ranged from .71–.80). As expected, the   RewP did not reach acceptable internal consistency (*r*'s were .27 and .21 at Time 1 and 2, respectively).

Using the flanker task, the CRN achieved acceptable internal consistency at both Time 1 and Time 2 (Table 2; *r*'s were .82 and .77 at Time 1 and 2, respectively). However, the ERN only reached acceptable internal consistency at Time 1 but not Time 2 (Table 2; *r*'s were .71 and .62 at Time 1 and 2, respectively). The   ERN did not reach acceptable internal consistency (*r*'s were .36 and .35 at Time 1 and 2, respectively).

Using the go/no-go task, the CRN achieved acceptable internal consistency at Time 1 and Time 2 (Table 2; *r*'s were both .79). In contrast, the ERN did not reach acceptable internal consistency at Time 1 or Time 2 (Table 2; *r*'s were .67 and .36, respectfully). The   ERN did not meet acceptable internal consistency (*r*'s were .48 and .38 at Time 1 and 2, respectively).

Test-Retest Reliability

As shown in Table 2, the FN and RewP were significantly correlated from Time 1 to Time 2, with large effect sizes (*r*'s = .56 and .59, respectively; Cohen, 1992). The   RewP was not significantly correlated from Time 1 to Time 2 (*r* = .16).

Using the flanker task, only the ERN was significantly correlated from Time 1 to Time 2, with a medium effect size (*r* = .48). The CRN and   ERN were not significantly correlated from Time 1 to Time 2 (*r*'s = .25 and .12, respectively).

Table 2 shows that using the go/no-go task, only the CRN was significantly correlated from Time 1 to Time 2, with a medium effect size (*r* = .32). However, the ERN and   ERN were not significantly correlated from Time 1 to Time 2 (*r* = −.02 and −.18, respectively).

Discussion

The present study examined the internal consistency and test-retest reliability of reward and error-related ERPs in early childhood. Consistent with the extant child and adult

psychometric literature (Bress et al., 2015; Levinson et al., 2019; Marco-Pallares et al., 2011), the RewP and FN exhibited acceptable internal consistency. The RewP and FN also showed moderate test-retest reliability across an 8-month time period. The internal consistency of the ERN and CRN were also largely within the acceptable range, with the exception of the Time 2 ERN using the go/no-go task.

In terms of stability, the test-retest reliability of the ERN and CRN was task-dependent, with the ERN showing significant, moderate reliability from Time 1 to Time 2 using the flanker task and nonsignificant, low reliability from Time 1 to Time 2 using the go/no-go task. Conversely, the CRN showed nearly significant, low reliability and from Time 1 to Time 2 using the flanker task and a significant, low reliability from Time 1 to Time 2 using the go/no-go task.

Altogether, the psychometric properties were stronger using the flanker task than the go/no-go task. These differences in psychometric properties are consistent with prior studies showing that internal consistency of the ERN is moderated by task, with some results indicating that the flanker task may more reliably elicit the ERN than the go/no-go task (Clayson, 2020; Meyer et al., 2014). As suggested by Meyer and colleagues (2014), these discrepancies in reliability of the ERN are likely due to differences in task designs and cognitive demands, given that, unlike the flanker task, the go/no-go task requires participants to inhibit their responses to no-go stimuli. Consistent with other studies showing that the ERN can achieve acceptable reliability in adults (Clayson, 2020) and in older children (Meyer et al., 2014; Pontifex et al., 2010), the present study results indicate that the ERN may be reliable in early child samples. Therefore, the present study results provide novel information about the reliability of the ERN, with results suggesting that the ERN may be suitable for samples ranging from early childhood to adulthood.

Similar to previously reported results, the psychometric properties of reward and error-related ERP difference scores (i.e., ΔRewP and ΔERN) fell short of typical metrics of acceptability and temporal stability across all three tasks (Bress et al., 2015; Ethridge & Weinberg, 2018; Levinson et al., 2017). This finding was expected given that the reliability of difference scores is dependent upon the reliability and variability of their individual measures. Thus, if the individual measures do not show acceptable reliability and/or have unequal variances, the reliability of the difference score will suffer. Additionally, the psychometric properties of the ΔRewP and ΔERN may have also been negatively impacted by strong intercorrelations between the RewP and FN, and between the ERN and CRN, respectively (Levinson et al., 2017). For these reasons, difference scores used to isolate the neural activity of interest generally have been found to have lower reliability, suggesting that ERP difference scores will likely not meet typical standards for robust internal and test-retest reliability. Yet, despite these limitations, difference score measures may still be clinically meaningful if the majority of the reliable variance relates to other measures of individual differences (Patrick et al., 2019). Further, if the ΔRewP and ΔERN relate better to clinical measures than other neural measures of error or reward processing in children, then ERP difference scores may be clinically useful indicators of risk for psychopathology.

The present empirical results should be considered in light of some limitations. First, although our results indicate that, overall, the parenting intervention had no effect on the ERN and RewP in young children, it is possible that the intervention did have some influence on the stability of these neural markers and/or that the effect of the intervention on these neural markers was not detected in the time window assessed. Thus, more research is needed aimed at examining the internal consistency and test-retest reliability of the RewP and ERN in young children to see whether the present study results can be replicated in other young child samples. Additionally, more research is needed investigating whether psychosocial or pharmacological interventions can modify these neural markers in children, particularly children who may be at risk for the development of psychopathology. Second, given that many children had a mother who endorsed a history of anxiety and/or depression, it is possible that psychopathological risk may have also influenced the present study results. Third, the present study examined the reliability of the ERN and CRN using 50 trials, but it is possible that including a greater number of trials in our analyses could have improved the reliability of these ERPs. Therefore, future studies are encouraged to examine the reliability of these neural markers using a greater number of error and correct trials. Finally, our sample was size was limited and future studies are encouraged to examine child ERPs in larger clinical and non-clinical samples.

Over the past few decades, researchers have made great strides in better understanding neural correlates of psychopathology with the aim of identifying clinically useful biomarkers that may aid in the prevention, assessment, and treatment of mental disorders (Insel et al., 2010). The RewP and ERN have shown much promise as potential biomarkers of risk for depression and anxiety disorders, respectively (Meyer, 2017; Proudfit, 2015). However, if these neural markers are to be clinically useful, it is critical to first establish that they have adequate psychometric properties as well as how their properties may differ among various samples. Childhood is a critical period for developmental changes in brain function and the emergence of particular forms of internalizing psychopathology thus, examining the reliability of the RewP and ERN in child samples is especially important for establishing these ERPs as clinically useful biomarkers (Galvan, 2010; Tamnes et al., 2013). The present study suggests that among early childhood samples, reward-and error-related ERPs are reliable measures of individual differences, but that the clinical utility of error-related ERPs may be somewhat dependent upon the task employed.

Future studies should regularly report the psychometric properties of ERPs in their samples, as others have advocated (Clayson, 2020; Clayson & Miller, 2017), since the reliability of ERPs may differ slightly in each sample. Tasks designed to examine error-related neural markers are challenging in that researchers cannot ensure that a particular number of errors are elicited by all participants. Thus, designing tasks to be adaptive to participants' skill level and developmentally appropriate are important considerations for study design. Importantly, this is one of the only studies to examine the reliability of error and reward-related neural markers in an early childhood sample, therefore, more research is needed to further replicate these findings in larger samples and examine the reliability of these neural markers across development.

Acknowledgments

References

Becker MPI, Nitsch AM, Miltner WHR, & Straube T (2014). A single-trial estimation of the feedback-related negativity and its relation to BOLD responses in a time-estimation task. Journal of Neuroscience, 19(34), 8. 10.1523/JNEUROSCI.3684-13.2014

Beesdo K, Knappe S, & Pine DS (2009). Anxiety and Anxiety Disorders in Children and Adolescents: Developmental Issues and Implications for DSM-V. Psychiatric Clinics of North America, 32(3), 483–524. 10.1016/j.psc.2009.06.002

Belden AC, Irvin K, Hajcak G, Kappenman ES, Kelly D, Karlow S, Luby JL, & Barch DM (2016). Neural Correlates of Reward Processing in Depressed and Healthy Preschool-Age Children. Journal of the American Academy of Child and Adolescent Psychiatry, 55(12), 1081–1089. 10.1016/j.jaac.2016.09.503 [PubMed: 27871643]

Bress JN, Foti D, Kotov R, Klein DN, & Hajcak G (2013). Blunted neural response to rewards prospectively predicts depression in adolescent girls. Psychophysiology, 50(1), 74–81. 10.1111/j.1469-8986.2012.01485.x [PubMed: 23252717]

Bress JN, Meyer A, & Proudfit GH (2015). The stability of the feedback negativity and its relationship with depression during childhood and adolescence. Development and Psychopathology, 27(4 Pt 1), 1285–1294. 10.1017/s0954579414001400 [PubMed: 26439074]

Burani K, Mulligan EM, Klawohn J, Luking KR, Nelson BD, & Hajcak G (2019). Longitudinal increases in reward-related neural activity in early adolescence: Evidence from event-related potentials (ERPs). Developmental Cognitive Neuroscience, 36. 10.1016/j.dcn.2019.100620

Carlson JM, Foti D, Mujica-Parodi LR, Harmon-Jones E, & Hajcak G (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: A combined ERP and fMRI study. NeuroImage, 57(4), 1608–1616. 10.1016/j.neuroimage.2011.05.037 [PubMed: 21624476]

Carrasco M, Harbin SM, Nienhuis JK, Fitzgerald KD, Gehring WJ, & Hanna GL (2013). Increased error-related brain activity in youth with obsessive-compulsive disorder and unaffected siblings. Depression and Anxiety, 30(1), 39–46. 10.1002/da.22035 [PubMed: 23225541]

Clayson PE (2020). Moderators of the internal consistency of error-related negativity scores: A meta-analysis of internal consistency estimates. Psychophysiology, 57(8). 10.1111/psyp.13583

Clayson PE, & Miller GA (2017). Psychometric considerations in the measurement of event-related brain potentials: Guidelines for measurement and reporting. International Journal of Psychophysiology, 111, 57–67. 10.1016/j.ijpsycho.2016.09.005 [PubMed: 27619493]

Davies PL, Segalowitz SJ, & Gavin WJ (2004). Development of response-monitoring ERPs in 7- to 25-year-olds. Developmental Neuropsychology, 25(3), 355–376. 10.1207/s15326942dn2503_6 [PubMed: 15148003]

Ethridge P, & Weinberg A (2018). Psychometric properties of neural responses to monetary and social rewards across development. International Journal of Psychophysiology, 132, 311–322. 10.1016/j.ijpsycho.2018.01.011 [PubMed: 29402529]

Foti D, Weinberg A, Dien J, & Hajcak G (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. Human Brain Mapping, 32(12), 2207–2216. 10.1002/hbm.21182 [PubMed: 21305664]

Furr RM, & Bacharach VR (2013). Psychometrics: An Introduction (2nd ed.). Sage PublicationsSage CA.

Galvan A (2010). Adolescent development of the reward system. Frontiers in Human Neuroscience, 4, 6. 10.3389/neuro.09.006.2010 [PubMed: 20179786]

Gratton G, Coles MGH, & Donchin E (1983). A new method for off-line removal of ocular artifact. Electroencephalography and Clinical Neurophysiology, 55(4), 468–484. 10.1016/0013-4694(83)90135-9 [PubMed: 6187540]

Hajcak G (2012). What We've Learned From Mistakes: Insights From Error-Related Brain Activity. Current Directions in Psychological Science, 21(2), 101–106. 10.1177/0963721412436809

Hajcak G, Klawohn J, & Meyer A (2019). The Utility of Event-Related Potentials in Clinical Psychology. Annual Review of Clinical Psychology, 15, 71–95. 10.1146/annurev-clinpsy-050718-095457

Insel T, Cuthbert B, Garvey M, Heinssen R, Pine D, Quinn K, Sanislow C, & Wang P (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. American Journal of Psychiatry Online, 167(7), 748–751.

Keren H, O'Callaghan G, Vidal-Ribas P, Buzzell GA, Brotman MA, Leibenluft E, Pan PM, Meffert L, Kaiser A, Wolke S, Pine DS, & Stringaris A (2018). Reward processing in depression: A conceptual and meta-analytic review across fMRI and EEG studies. American Journal of Psychiatry, 175(11), 1111–1120. 10.1176/appi.ajp.2018.17101124

Kujawa A, & Burkhouse KL (2017). Vulnerability to Depression in Youth: Advances From Affective Neuroscience. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2(1), 28–37. 10.1016/j.bpsc.2016.09.006 [PubMed: 28497126]

Kujawa A, Carroll A, Mumper E, Mukherjee D, Kessel EM, Olino T, Hajcak G, & Klein DN (2018). A longitudinal examination of event-related potentials sensitive to monetary reward and loss feedback from late childhood to middle adolescence. International Journal of Psychophysiology, 132(Pt B), 323–330. 10.1016/j.ijpsycho.2017.11.001 [PubMed: 29113953]

Kujawa A, Proudfit GH, & Klein DN (2014). Neural reactivity to rewards and losses in offspring of mothers and fathers with histories of depressive and anxiety disorders. Journal of Abnormal Psychology, 123(2), 287–297. 10.1037/a0036285 [PubMed: 24886003]

Larson MJ, Baldwin SA, Good DA, & Fair JE (2010). Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): The role of number of trials. Psychophysiology, 47(6), 1167–1171. 10.1111/j.1469-8986.2010.01022.x [PubMed: 20477982]

Lecrubier Y, Sheehan DV, Weiller E, Amorim P, Bonora I, Sheehan KH, Janavs J, & Dunbar GC (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. European Psychiatry, 12(5), 224–231. 10.1016/S0924-9338(97)83296-8

Levinson AR, Speed BC, & Hajcak G (2019). Neural Response to Pleasant Pictures Moderates Prospective Relationship Between Stress and Depressive Symptoms in Adolescent Girls. Journal of Clinical Child and Adolescent Psychology, 48(4), 643–655. 10.1080/15374416.2018.1426004 [PubMed: 29412004]

Levinson AR, Speed BC, Infantolino ZP, & Hajcak G (2017). Reliability of the electrocortical response to gains and losses in the doors task. Psychophysiology, 54(4), 601–607. 10.1111/psyp.12813 [PubMed: 28072462]

Lukie CN, Montazer-Hojat S, & Holroyd CB (2014). Developmental changes in the reward positivity: An electrophysiological trajectory of reward processing. Developmental Cognitive Neuroscience, 9, 191–199. 10.1016/j.dcn.2014.04.003 [PubMed: 24879113]

Luking KR, Nelson BD, Infantolino ZP, Sauder CL, & Hajcak G (2017). Internal Consistency of Functional Magnetic Resonance Imaging and Electroencephalography Measures of Reward in Late Childhood and Early Adolescence. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2(3), 289–297. 10.1016/j.bpsc.2016.12.004 [PubMed: 29057369]

Marco-Pallares J, Cucurell D, Münte TF, Strien N, & Rodriguez-Fornells A (2011). On the number of trials needed for a stable feedback-related negativity. Psychophysiology, 48(6), 852–860. 10.1111/j.1469-8986.2010.01152.x [PubMed: 21091959]

McDermott JM, Perez-Edgar K, Henderson HA, Chronis-Tuscano A, Pine DS, & Fox NA (2009). A History of Childhood Behavioral Inhibition and Enhanced Response Monitoring in Adolescence Are Linked to Clinical Anxiety. Biological Psychiatry, 65(5), 445–448. 10.1016/j.biopsych.2008.10.043 [PubMed: 19108817]

Merikangas KR, Nakamura EF, & Kessler RC (2009). Epidemiology of mental disorders in children and adolescents. Dialogues in Clinical Neuroscience, 11(1), 7–20. 10.31887/dcns.2009.11.1/krmerikangas [PubMed: 19432384]

Meyer A (2017). A biomarker of anxiety in children and adolescents: A review focusing on the error-related negativity (ERN) and anxiety across development. Developmental Cognitive Neuroscience, 27, 58–68. 10.1016/j.dcn.2017.08.001 [PubMed: 28818707]

Meyer A, Bress JN, & Proudfit GH (2014). Psychometric properties of the error-related negativity in children and adolescents. Psychophysiology, 51(7), 602–610. 10.1111/psyp.12208 [PubMed: 24646380]

Meyer A, Carlton C, Crisler S, & Kallen A (2018). The development of the error-related negativity in large sample of adolescent females: Associations with anxiety symptoms. Biological Psychology, 138, 96–103. 10.1016/j.biopsycho.2018.09.003 [PubMed: 30201401]

Meyer A, Hajcak G, Torpey-Newman DC, Kujawa A, & Klein DN (2015). Enhanced error-related brain activity in children predicts the onset of anxiety disorders between the ages of 6 and 9. Journal of Abnormal Psychology, 124(2), 266–274. 10.1037/abn0000044 [PubMed: 25643204]

Meyer A, Lerner MD, De Los Reyes A, Laird RD, & Hajcak G (2017). Considering ERP difference scores as individual difference measures: Issues with subtraction and alternative approaches. Psychophysiology, 54(1), 114–122. 10.1111/psyp.12664 [PubMed: 28000251]

Meyer A, Nelson B, Perlman G, Klein DN, & Kotov R (2018). A neural biomarker, the error-related negativity, predicts the first onset of generalized anxiety disorder in a large sample of adolescent females. Journal of Child Psychology and Psychiatry and Allied Disciplines, 59(11), 1162–1170. 10.1111/jcpp.12922

Meyer A, Riesel A, & Proudfit GH (2013). Reliability of the ERN across multiple tasks as a function of increasing errors. Psychophysiology, 50(12), 1220–1225. 10.1111/psyp.12132 [PubMed: 24730035]

Meyer A, Weinberg A, Klein DN, & Hajcak G (2012). The development of the error-related negativity (ERN) and its relationship with anxiety: Evidence from 8 to 13 year-olds. Developmental Cognitive Neuroscience, 2(1), 152–161. 10.1016/j.dcn.2011.09.005 [PubMed: 22308177]

Moser JS, Moran TP, Schroder HS, Donnellan MB, & Yeung N (2013). On the relationship between anxiety and error monitoring: A meta-analysis and conceptual framework. Frontiers in Human Neuroscience, 7, 466. 10.3389/fnhum.2013.00466 [PubMed: 23966928]

Nelson BD, Perlman G, Klein DN, Kotov R, & Hajcak G (2016). Blunted neural response to rewards as a prospective predictor of the development of depression in adolescent girls. American Journal of Psychiatry, 50(1), 74–81. 10.1176/appi.ajp.2016.15121524

Nunnally JC, Bernstein IH, & Berge JMT (1967). The Assessment of Reliability. In Psychometric Theory McGraw Hill.

Olivari MG, Tagliabue S, & Confalonieri E (2013). Parenting Style and Dimensions Questionnaire: A Review of Reliability and Validity. Marriage and Family Review, 49(6), 465–490. 10.1080/01494929.2013.770812

Olvet DM, & Hajcak G (2009). The stability of error-related brain activity with increasing trials. Psychophysiology, 46(5), 957–961. 10.1111/j.1469-8986.2009.00848.x [PubMed: 19558398]

Patrick CJ, Iacono WG, & Venables NC (2019). Incorporating neurophysiological measures into clinical assessments: Fundamental challenges and a strategy for addressing them. Psychological Assessment, 31(12), 1512–1529. 10.1037/pas0000713 [PubMed: 30896211]

Pontifex MB, Scudder MR, Brown ML, O'Leary KC, Wu CT, Themanson JR, & Hillman CH (2010). On the number of trials necessary for stabilization of error-related brain activity across the life span. Psychophysiology, 47(4), 767–773. 10.1111/j.1469-8986.2010.00974.x [PubMed: 20230502]

Proudfit GH (2015). The reward positivity: From basic research on reward to a biomarker for depression. Psychophysiology, 52(4), 449–459. 10.1111/psyp.12370 [PubMed: 25327938]

Riesel A (2019). The erring brain: Error-related negativity as an endophenotype for OCD—A review and meta-analysis. Psychophysiology, 56(4), e13348. 10.1111/psyp.13348 [PubMed: 30838682]

Segalowitz SJ, Santesso DL, Murphy TI, Homan D, Chantziantoniou DK, & Khan S (2010). Retest reliability of medial frontal negativities during performance monitoring. Psychophysiology, 47(2), 260–270. 10.1111/j.1469-8986.2009.00942.x [PubMed: 20030755]

Szenczy A, Levinson A, Hajcak G, Bernard K, & Nelson B (2021, 6 25). Reliability of Reward-and Error-Related Brain Activity in Early Childhood. Open Science Framework https://osf.io/tfpjc/

Tamnes CK, Walhovd KB, Torstveit M, Sells VT, & Fjell AM (2013). Performance monitoring in children and adolescents: A review of developmental changes in the error-related negativity and brain maturation. Developmental Cognitive Neuroscience, 6, 1–13. 10.1016/j.dcn.2013.05.001 [PubMed: 23777674]

Torpey DC, Hajcak G, Kim J, Kujawa A, & Klein DN (2012). Electrocortical and behavioral measures of response monitoring in young children during a Go/No-Go task. Developmental Psychobiology, 54(2), 139–150. 10.1002/dev.20590 [PubMed: 21815136]

Weinberg A, & Hajcak G (2011). Longer term test-retest reliability of error-related brain activity. Psychophysiology, 48(10), 1420–1425. 10.1111/j.1469-8986.2011.01206.x [PubMed: 21496055]

Weinberg A, Olvet DM, & Hajcak G (2010). Increased error-related brain activity in generalized anxiety disorder. Biological Psychology, 30(1), 39–46. 10.1016/j.biopsycho.2010.09.011
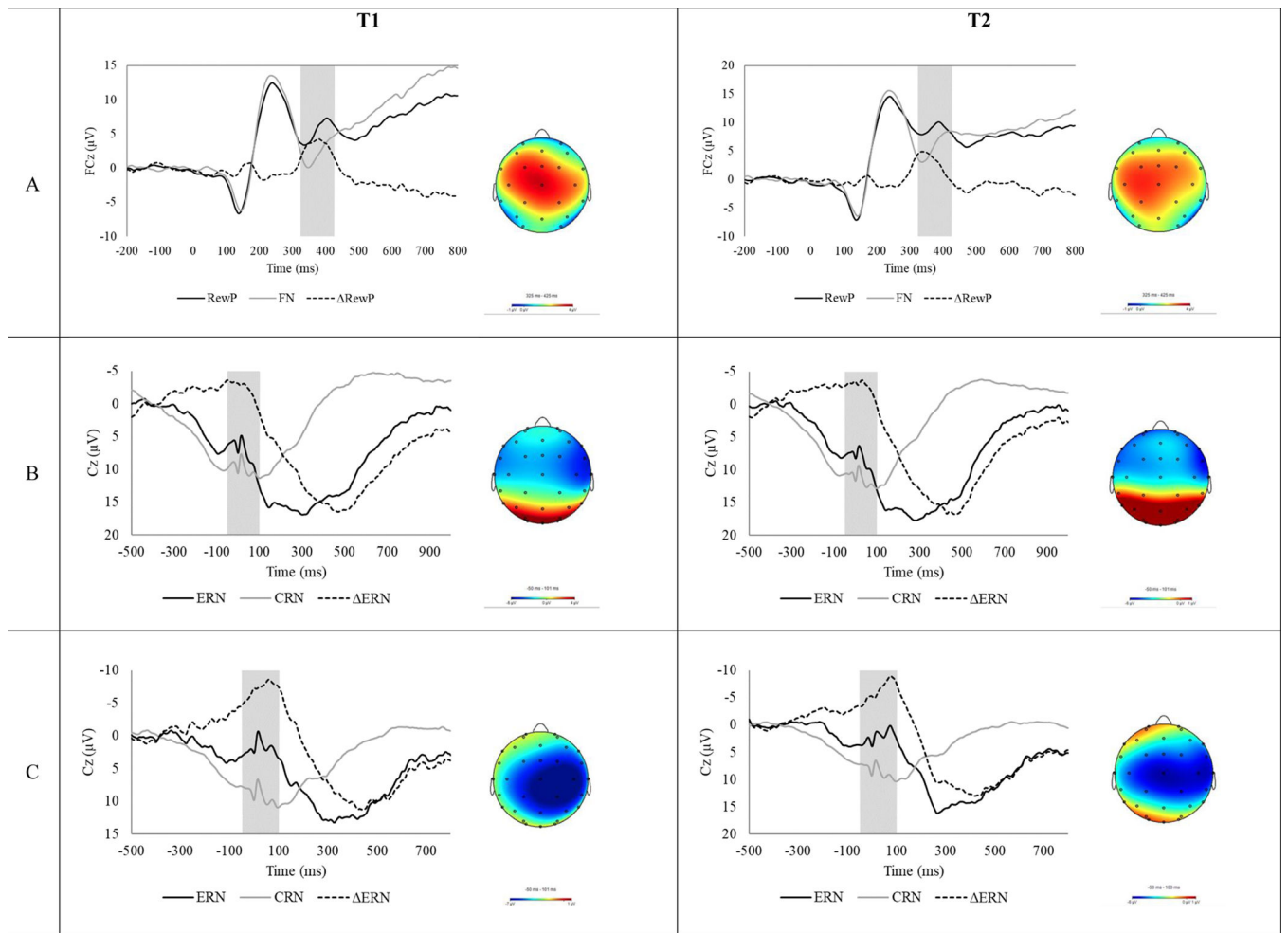
Figure 1.
Event-related potential (ERP) waveforms (left) and scalp distributions (right) for the doors (panel A), flanker (panel B), and go/no-go (panel C) tasks.

Table 1.

ERP and behavioral data for the flanker and go/no-go tasks at Time 1 and Time 2

	Flanker Time 1 **Mean (*SD*)**	**Flanker Time 2** **Mean (*SD*)**	**Go/No-Go Time 1** **Mean (*SD*)**	**Go/No-Go Time 2** **Mean (*SD*)**
Accuracy (% correct)	77.44 (10.55)	83.39 (7.88)	86.06 (7.99)	89.09 (6.30)
Correct RT (ms)	654 (150)	637 (131)	656 (67)	624 (62)
Error RT (ms)	433 (91)	424 (95)	549 (93)	514 (100)
ERN (µV)	7.56 (9.83)	7.40 (10.09)	−0.40 (9.37)	2.20 (15.79)
CRN (µV)	10.26 (8.65)	10.25 (8.17)	9.66 (7.83)	10.19 (9.20)
ERN (µV)	−2.70 (7.99)	−2.85 (8.37)	−9.71 (11.61)	−7.99 (18.64)

Note. All ERP data are based on participants' first 50 trials. RT= reaction time; SD= standard deviation.

Table 2.

Split-half reliability and test-retest reliability for ERP measures from the doors, flankers, and go/no-go tasks at time 1 and time 2

| Task | | Split-Half Reliability | | Test-Retest Reliability |
		Time 1	Time 2	Pearson's *r*
Doors	**RewP**	.73	.70	.59 ***
	FN	.80	.71	.56 ***
	RewP	.27	.21	.16
Flanker	**ERN**	.71	.62	.48 ***
	CRN	.82	.77	.25 +
	ERN	.36	.35	.12
Go/No-Go	**ERN**	.67	.36	−.02
	CRN	.79	.79	.32 *
	ERN	.48	.38	−.18

Note.

+ *p* < .10.

* *p* < .05.

** *p* < .01.

*** *p* < .001

Author Manuscript