Psychophysiology, •• (2014), ••–••. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/psyp.12208

Psychometric properties of the error-related negativity in children and adolescents

ALEXANDRIA MEYER, JENNIFER N. BRESS, AND GREG HAJCAK PROUDFIT
Department of Psychology, Stony Brook University, Stony Brook, New York, USA

Abstract

Error processing is frequently examined using the error-related negativity (ERN), a negative-going event-related potential occurring after the commission of an error at frontal-central sites, and has been suggested as a neural biomarker that may be useful in characterizing trajectories of risk for anxiety. While the ERN has been shown to have excellent psychometric properties in adults, few studies have examined psychometric properties of the ERN in children and adolescents. The current study examined the 2-year test-retest reliability of the ERN in a sample of children and adolescents, and the convergent validity of the ERN using a flanker and go/no-go task. Results suggest that the ERN is both reliable and stable across 2 years and across tasks. However, results also indicate that the internal consistency obtained using the flanker task is greater than the internal consistency obtained using the go/no-go task.

Descriptors: Error-related negativity, Psychometrics, Children, Development, Error processing, Risk markers

In a changing environment, a flexible system for detecting errors is necessary for the mobilization of appropriate responses (Holroyd & Coles, 2002). Over the past 20 years, event-related potential (ERP) studies of action monitoring have utilized the error-related negativity (ERN) to study error detection in humans (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). The ERN, a negative deflection appearing at frontocentral sites approximately 50 ms after the commission of an error, is thought to reflect the activity of a general error detection system that becomes active across a range of stimulus and response modalities (Falkenstein et al., 1991; Gehring et al., 1993; Holroyd, Dien, & Coles, 1998; Holroyd & Coles, 2002).

Numerous studies have examined the ERN in relation to individual difference measures. For instance, an increased ERN has been observed in individuals with clinical anxiety disorders, especially obsessive-compulsive disorder (Endrass, Klawohn, Schuster, & Kathmann, 2008; Gehring, Himle, & Nisenson, 2000; Xiao et al., 2011) and generalized anxiety disorder (Weinberg, Klein, & Hajcak, 2012; Weinberg, Olvet, & Hajcak, 2010). Additionally, individuals characterized by worry (Hajcak, McDonald, & Simons, 2003), behavioral inhibition (Amodio, Master, Yee, & Taylor, 2008), high negative affect (Hajcak, McDonald, & Simons, 2004), and punishment sensitivity (Boksem, Tops, Wester, Meijman, & Lorist, 2006) display an increased ERN. In light of these findings, the ERN has been suggested as a neural biomarker of trait anxiety, possibly related to anxious apprehension in particular (Weinberg, Riesel, & Hajcak, 2012).

The validity of using the ERN as an individual difference measure depends on its reliability (Cronbach & Meehl, 1955). Reliability refers to the ability of an instrument to consistently measure a characteristic. The ERN is derived by averaging many error trials, and if the trial-to-trial waveforms are unreliable (i.e., internally inconsistent) within a testing session, then the average will also be unreliable (Simons & Miles, 1990). Internal consistency can be measured in terms of split-half reliability (e.g., correlation between odd and even trials), or Cronbach's alpha (i.e., the average of all possible split-halves). Convergent validity can be examined using different tasks to elicit the ERN in the same individuals, measuring the extent to which the ERN reflects common error-related brain activity across tasks. Furthermore, utilizing multiple tasks provides an opportunity to compare internal reliability between tasks. Finally, if the ERN is traitlike, then it should also be relatively stable across testing sessions (i.e., high test-retest reliability).

Existing research in adults suggests that the ERN has good internal consistency (Larson, Baldwin, Good, & Fair, 2010; Olvet & Hajcak, 2009b; Riesel, Weinberg, Endrass, Meyer, & Hajcak, 2013), good convergent validity (Riesel et al., 2013), and high test-retest reliability over periods of weeks (Olvet & Hajcak, 2009a), and even up to 2 years (Weinberg & Hajcak, 2011). Taken together, these findings suggest that in adults the ERN is a neural measure with traitlike properties (i.e., a neurobehavioral trait; Hajcak, 2012).

Nonetheless, recent work suggests important task-related differences in both the psychometric properties of the ERN and its relationship with individual differences. For instance, although the ERN is highly correlated across tasks, internal reliability differs across tasks (Meyer, Riesel, & Proudfit, 2013). Results from this study indicated that the magnitude of the ERN in a go/no-go task was highly dependent on the number of error trials included in the averages, and that the internal consistency of the ERN elicited using a Stroop task remained low—even after 20 errors were

Address correspondence to: Alexandria Meyer, Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500. E-mail: Ammeyer3@gmail.com

included in averages. On the other hand, the ERN elicited using a flanker task achieved good internal reliability after approximately 10 errors were committed. Collectively, these data suggest that the ERN elicited during the flanker task might have more desirable psychometric properties for research on individual differences. Additionally, a recent study from our group highlighted the fact that the relationship between ERN and clinical outcomes depends on task-specific psychometric properties of the ERN (Foti, Kotov, & Hajcak, 2013).

Although the ERN appears to be both reliable and stable in adulthood, less is known about its psychometric properties in childhood and adolescence. Although the ERN can be elicited in children as young as 4–7 years of age (Brooker, Buss, & Dennis, 2011; Torpey, Hajcak, & Klein, 2009; Wiersema, van der Meere, & Roeyers, 2007) and generally has a morphology and scalp distribution similar to that of adults (Arbel & Donchin, 2011), the magnitude of the ERN has been shown to increase across development (Davies, Segalowitz, & Gavin, 2004; Tamnes, Walhovd, Torstveit, Sells, & Fjell, 2013a)—although there also appears to be a transient decrease in the ERN during puberty (Davies et al., 2004; Meyer, Bress, & Proudfit, Submitted). Considering such developmental changes, the ERN might be less traitlike, stable, and reliable in younger participants.

Consistent with work in adults, an increased ERN has been observed in a heterogeneous group of clinically anxious children (Ladouceur, Dahl, Birmaher, Axelson, & Ryan, 2006), children with obsessive-compulsive disorder (Carrasco et al., 2013; Hajcak, Franklin, Foa, & Simons, 2008; Hanna et al., 2012), and children with clinical anxiety as young as 6 years old (Meyer, Hajcak et al., 2013). Additionally, children with nonclinical symptoms of obsessive-compulsive disorder (Santesso, Segalowitz, & Schmidt, 2006) and adolescents with nonclinical anxiety (Meyer, Weinberg, Klein, & Hajcak, 2012) have also been shown to have an increased ERN. Thus, despite developmental changes in the ERN, it seems to reliably relate to anxiety in youth—and, therefore, might relate to developmental trajectories of risk (Hajcak, 2012; Meyer et al., 2012).

Considering the ERN's potential utility as an early neurobehavioral risk maker, it is surprising that few studies have investigated the reliability of the ERN in child and adolescent populations. One previous study examined the internal reliability of the ERN during a flanker task in a group of preadolescent children (i.e., 8–11 years old) and young adults (i.e., 18–25 years old), finding moderate internal consistency after 6 trials in both groups (Pontifex et al., 2010). Likewise, good test-retest reliability over a 3-week period (*r*s between .40 and .60) as well as good convergent validity across two tasks ($r = .54$) in a group of 15-year-old boys was reported in another study (Segalowitz et al., 2010). In light of neurodevelopmental changes that take place from childhood through adolescence, it is important to determine to what extent the ERN is stable, traitlike, and reliable in younger participants over a longer period of time. In addition, convergent validity of the ERN across multiple tasks requires further investigation, especially considering developmental changes in perceptual, motor, and cognitive abilities.

Moreover, it is important to compare internal reliability across tasks as a function of error number. Tasks may have comparable psychometric properties when all error trials are included, but the ERN from some tasks may achieve better psychometric properties with relatively fewer error trials (Meyer, Riesel, & Proudfit, 2013). This is particularly important when investigators wish to examine biomarkers using relatively brief tasks.

To address these issues, the current study examined the ERN elicited by the flanker task in the same children/adolescents at two testing sessions separated by 2 years. In addition, a go/no-go task was included at the second testing session to evaluate convergent validity of the ERN. Internal consistency of the ERN in all three tasks (flanker at Time 1 and 2, go/no-go at Time 2) was examined as a function of increasing error trials. Together, these data should shed light on the stability of the ERN over a relatively long test-retest interval across late childhood and early adolescence, the convergent validity across tasks in a developmental sample, and potential task-related differences in psychometric properties of the ERN. Based on previous studies (Segalowitz et al., 2010; Weinberg & Hajcak, 2011), we expected to find moderate test-retest reliability of the ERN across 2 years, although this hypothesis was tentative given work suggesting substantial developmental change in the ERN (Davies et al., 2004; Meyer et al., Submitted; Tamnes, Walhovd, Torstveit, Sells, & Fjell, 2013b). Additionally, we hypothesized that, whereas convergent validity between the flanker and go/no-go task would be moderate (Riesel et al., 2013; Segalowitz et al., 2010), important psychometric properties could differ, such that the flanker task might attain superior psychometric properties after fewer error trials.

Method

Participants

Overall, 44 participants had adequate data for the current study. Originally, participants were recruited via a commercial mailing list that targeted families with children in Stony Brook and the surrounding community. Letters, followed by phone calls, went out to approximately 800 families. The original sample consisted of 70 participants (30 female) between the ages of 8 and 13 ($M = 10.95$, $SD = 1.48$) who completed laboratory tasks at Time 1. From the baseline assessment, 4 participants were excluded due to poor quality recordings and 11 for making too many (i.e., 85 or more errors, less than 75% accuracy) or too few (i.e., fewer than 6) errors (Olvet & Hajcak, 2009b); the final Time 1 sample consisted of 55 participants (24 female). Two years later, 47 (29 male) of these participants, now between the ages of 10 and 15 ($M = 12.74$, $SD = 1.5$), returned to the laboratory to complete a second laboratory visit. At Time 2, three participants were excluded due to poor quality recordings and no one was excluded for making too few or too many errors, leaving a total sample of 44 participants (15 female) who had adequate data for the current study; 89% of the sample was Caucasian, 2% African-American, 8% identified as Other.

Tasks and Procedure

During both laboratory visits, participants performed multiple tasks while electroencephalogram (EEG) data were collected, and the order of the tasks was counterbalanced across subjects. To elicit an ERN at Time 1, participants completed an arrowhead version of the flanker task; at Time 2, participants completed the same version of the flanker task and a go/no-go task.

During each lab visit, after consent was obtained and a description of the experiment was provided, the EEG electrodes were attached. On each trial of the flanker task, horizontally aligned arrowheads were presented for 200 ms, followed by an intertrial interval (ITI) varying randomly between 2,300 and 2,800 ms. Half of the trials were compatible (">>>>>" or "<<<<<") and half were incompatible (“<<><<” or “>><>>”); the order of trials was randomly determined. Each set of arrowheads occupied approximately 1.3° of visual angle vertically and 9.2° horizontally. Participants were told to press the right mouse button if the center arrow was facing to the right and to press the left mouse button if the center arrow was facing to the left. After a practice block of 30 trials, participants completed 11 blocks of 30 trials (330 trials total) with each block initiated by the participant. Participants received feedback based on their performance at the end of each block. If performance was 75% correct or lower, the message “Please try to be more accurate” was displayed; if performance was above 90% correct, the message “Please try to respond faster” was displayed; otherwise the message “You’re doing a great job” was displayed.

During the go/no-go task, after a practice block of 20 trials, there were a total of 420 trials, consisting of 7 blocks of 60 trials each. The stimuli were green equilateral triangles in three different orientations; 20% of the triangles were slightly tilted to the right or left (“no-go” stimuli) and 80% of the triangles were vertically aligned and pointed up (“go” stimuli). All stimuli were presented for 200 ms, followed by an ITI that varied randomly between 600 to 1,000 ms. Children were instructed to respond to upward-pointing triangles by pressing a button, and to withhold responses to slightly tilted triangles. Participants received the same type of feedback based on their performance at the end of each block that was provided during the flanker task.

Psychophysiological Recording, Data Reduction, and Analysis

Continuous EEG recordings were collected using an elastic cap and the ActiveTwo BioSemi system (BioSemi, Amsterdam, The Netherlands). Thirty-four electrode sites were used, based on the 10/20 system, in addition to two electrodes on the right and left mastoids. Eye blinks and eye movements (electrooculogram, EOG) were recorded using four facial electrodes: vertical eye movements and blinks were measured via two electrodes place approximately 1 cm above and below the right eye, and horizontal eye movements were measured via two electrodes located approximately 1 cm outside the outer edge of the right and left eyes. The EEG signal was preamplified at each electrode to improve the signal-to-noise ratio and amplified with a gain of 1x by a BioSemi ActiveTwo system. 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 204.8 Hz. Each active electrode was measured online with respect to a common mode sense (CMS) active electrode producing a monopolar (nondifferential) channel. Offline, the data were referenced to the average of the right and left mastoids, and band-pass filtered with low and high cutoffs of 0.1 and 30 Hz, respectively. To detect and reject artifacts, we used a semiautomatic procedure for all segmented data, with a criteria of a voltage step of more than 50.0 μV between sample points, a maximum voltage difference of less than .50 μV within any 100-ms interval, and a voltage difference of 300.0 μV within a trial. These intervals were rejected from individual channels within a trial. Afterwards, the data were visually inspected to detect and reject any remaining artifacts. Eyeblink and ocular corrections were conducted per Gratton, Coles, and Donchin (1983).

For both the go/no-go and the flanker task, the EEG was segmented beginning 500 ms before the response and continuing for 600 ms after the response; a 200-ms window from −500 to −300 ms before the response onset served as the baseline. Correct and error trials were averaged separately. For each subject, the ERN was quantified as the average activity from 0–100 ms after the response at Cz (where it was maximal during the go/no-go task) and Fz (where it was maximal during the flanker task at both time points).[1] Although the primary focus was on the ERN, we also evaluated the correct response negativity (CRN) in the same time window at the same electrodes, after correct responses; the ΔERN was defined as the ERN minus the CRN. Behavioral measures for the flanker and go/no-go task include the number of error trials for each subject and accuracy expressed as a percentage of all trials. Reaction time (RTs) on error and correct trials were calculated separately.

Data Analysis

Statistical analyses were completed with SPSS (Version 17.0) using general linear model software, with Greenhouse-Geisser correction applied to p values associated with multiple degrees of freedom and repeated measures comparisons when necessitated by violation of the assumption of sphericity. Paired samples t tests were performed for follow-up post hoc comparisons.

For the flanker task, RTs were evaluated with a 2 (Trial Type: correct vs. error) × 2 (Time Point: Time 1 vs. Time 2) repeated measures analysis of variance (ANOVA). Accuracy (percentage of correct trials) was compared between Times 1 and 2 using a paired samples t test. A 2 (Trial Type: error vs. correct) × 2 (Time Point: Time 1 vs. Time 2) ANOVA was conducted to assess differences in ERP responses between error and correct trials, testing session, and their potential interaction. Intersubject stability from Time 1 to Time 2 was examined with Pearson’s r correlations.

Additionally, to compare RTs between the flanker and go/no-go task at the second testing session, a 2 (Trial Type: error vs. correct) × 2 (Task: flanker vs. go/no-go) ANOVA was conducted, and accuracy was compared between the two tasks using paired samples t tests. A 2 (Trial Type: error vs. correct) × 2 (Task: flanker vs. go/no-go) ANOVA was conducted to assess differences in ERP responses between error and correct trials where the ERN was maximal for each task (i.e., Cz for the go/no-go task and Fz for the flanker task), and their potential interaction. Convergent validity between the flanker and go/no-go task was examined with Pearson’s r correlations at all electrodes.

To examine internal consistency in the flanker task at both assessments and the ERN in the go/no-go task at Time 2, we calculated split-half reliability by taking the correlation between even and odd error trials and then adjusting with the Spearman-Brown prediction formula. This approach is advantageous in that it utilizes all ERP data from each participant to estimate the stability of the ERN across the entire task. Next, we computed the ERN as a function of increasing errors, deriving the ERN based on the first 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 error trials. We then performed a 3 (Task/Time) × 10 (ERP average) repeated measures ANOVA to determine if the ERN varied significantly within each task or as a function of increasing error number. To measure the degree to which the ERN based on a subset of error trials relates to the grand average ERN, we correlated these averages (i.e., ERN based on the first 2, 4, 6 . . . 20 trials) with the grand average ERN using Pearson’s correlation coefficient. Finally, we calculated Cronbach’s alpha (i.e., the average of all possible split-half reliabilities) as a function of increasing error trials to examine how

1. Results from paired samples t tests suggested that the ERN elicited in the flanker task at Time 1 and 2 was more negative at Fz than at Cz, $t(43) = -11.14$, $p < .001$, and $t(43) = -8.49$, $p < .001$. For the go/no-go task, the ERN was more negative at Cz than at Fz, $t(43) = -8.69$, $p < .001$. The pattern of results was the same for the ΔERN.

Table 1. *ERP and Behavioral Data for the Flanker Task at Time 1 and 2*

	Flanker: Time 1 Mean (*SD*)	Flanker: Time 2 Mean (*SD*)	Go/No-Go: Time 2 Mean (*SD*)	Flanker: Time 1 to Time 2 *r*	Flanker and Go/No-Go *r*
Accuracy (% correct)	84.2 (9.9)	84.8 (8.8)	87.4 (3.4)	.25	.55**
Correct RT (ms)	505 (89)	417 (54)	349 (55)	.58**	.61**
Incorrect RT (ms)	375 (76)	330 (42)	293 (35)	.40**	.38**
ERN (μV) at Fz	–4.14 (9.30)	–.68 (9.96)	–3.13 (6.03)	.34*	.51**
CRN (μV) at Fz	1.76 (5.97)	6.41 (10.60)	1.11 (5.43)	.14	.61**
ΔERN (μV) at Fz	–5.89 (7.60)	–7.10 (6.69)	–4.25 (5.32)	.18	.45**
ERN (μV) at Cz	7.23 (12.28)	11.86 (11.64)	3.05 (7.26)	.63**	.70**
CRN (μV) at Cz	11.17 (7.12)	15.47 (10.89)	8.26 (6.77)	.54**	.71**
ΔERN (μV) at Cz	–3.92 (9.20)	–3.61 (8.49)	–5.21 (5.09)	.20	.40**
Split-half reliability (ERN at Fz)	.63**	.71**	.38**	–	–
Split-half reliability (ERN at Cz)	.88**	.81**	.50**	–	–

Note. On the right, Pearson correlations for behavioral and ERP data between the flanker task at Time 1 and 2, and between the flanker and go/no-go task. *$p < .05$, **$p < .01$.

the number of errors impacts internal consistency. Due to the fact that the number of error trials varied across participants, the full sample was only available when calculating alpha using the first six error trials; after six errors, the number of participants decreased differentially in each task. For the go/no-go task: 18 errors ($N = 42$), 20 errors ($N = 42$), for the flanker task Time 1: 6 errors ($N = 43$), 8 errors ($N = 43$), 10 errors ($N = 43$), 12 errors ($N = 41$), 14 errors ($N = 41$), 16 errors ($N = 40$), 18 errors ($N = 39$), 20 errors ($N = 38$), and flanker task at Time 2: 14 errors ($N = 43$), 16 errors, ($N = 43$), 18 errors ($N = 43$), 20 errors ($N = 42$).

Results

Test-Retest Reliability: Flanker From Time 1 to Time 2

Behavioral data from Time 1 and 2 are presented in Table 1. Overall, RTs during error trials were faster than RTs during correct trials, $F(1,43) = 204.80$, $p < .001$, $\eta_p^2 = .83$; and faster at the second testing session, $F(1,43) = 46.61$, $p < .001$, $\eta_p^2 = .52$. The effect of time point was qualified by response type, $F(1,43) = 21.13$, $p < .001$, $\eta_p^2 = .33$, such that a larger reduction in RT over time was observed for correct trials than for error trials, $t(43) = 4.60$, $p < .001$. Accuracy did not differ from Time 1 to Time 2, $t(43) = .32$, $p = .75$.

Average ERP values from Time 1 and 2 are presented in Table 1, and Figure 1 presents the grand average response-locked ERPs for error and correct trials at Time 1 and Time 2. Confirming the impression from Figure 1, the neural response to errors was more negative than the neural response during correct trials at Fz, $F(1,43) = 61.43$, $p < .001$, $\eta_p^2 = .59$. Although the effect of trial type did not vary according to time point at Fz, $F(1,43) = .76$, $p = .39$, there was a significant main effect of time point, $F(1,43) = 6.87$, $p < .01$, $\eta_p^2 = .14$, such that both the CRN and ERN were more negative at Time 1, $t(43) = 2.71$, $p < .01$, and $t(43) = 2.07$, $p < .05$, respectively.

Test-retest reliability indices for behavioral and ERP measures are included in Table 1. Correlations from Time 1 to Time 2 were larger for the ERN than for the ΔERN, and larger at Cz than at Fz. Figure 2 presents Pearson's r correlations at each electrode after error commission (0–100 ms) from Time 1 to Time 2. As can be seen in the figure, error-related brain activity was most stable (rs = .6) at central sites along the midline (Cz and FCz). The stability, as indicated by intersubject stability (r), was moderate to high for the ERN, $r = .63$, $p < .01$, whereas the CRN was moderately stable, $r = .54$, $p < .01$.[2]

Convergent Validity: Comparing Flanker and Go/No-Go at Time 2

Go/no-go and flanker behavioral data from Time 2 is presented in Table 1. Overall, RTs during error trials were faster than RTs during correct trials, $F(1,43) = 235.08$, $p < .001$, $\eta_p^2 = .85$; and faster during the go/no-go task, $F(1,43) = 69.71$, $p < .001$, $\eta_p^2 = .62$. The effect of response type was qualified by task, $F(1,43) = 31.24$, $p < .001$, $\eta_p^2 = .42$, such that the difference between RTs on error and correct trials was larger during the flanker task, $t(43) = 5.59$, $p < .001$. Overall, accuracy was greater during the go/no-go task than during the flanker task, $t(43) = 2.28$, $p < .05$.

Average ERP values from the go/no-go and flanker task at Time 2 are also presented in Table 1, and Figure 1 presents the grand average response-locked ERPs for error and correct trials during the go/no-go and flanker task at Time 2. The neural response to errors was more negative than the neural response during correct trials at Cz, $F(1,43) = 25.83$, $p < .001$, $\eta_p^2 = .38$. Although the effect of trial type did not vary according to task, $F(1,43) = 1.76$, $p = .19$, there was a significant main effect of task at Cz, $F(1,43) = 57.61$, $p < .001$, $\eta_p^2 = .57$, such that both the CRN and ERN were more negative during the go/no-go task, $t(43) = 6.98$, $p < .001$, and $t(43) = 6.18$, $p < .001$, respectively.

Convergent validity between Time 2 flanker and go/no-go ERPs and behavioral data are presented in Table 1. Correlations across the tasks were larger for the ERN than for the ΔERN, and larger at Cz than at Fz. Figure 2 presents Pearson's r correlations at each electrode after error commission (0–100 ms) from the Time 2 ERNs elicited by the flanker and go/no-go tasks. Error-related brain activity was most consistent (rs = .7) at central sites along the midline (Cz and FCz). The convergent validity, as indicated by

2. Follow-up analyses suggest that when the sample is split by age at Time 1 (i.e., 8–10 versus 11–13 years old), the stability of error-related brain activity from Time 1 to Time 2 (at Cz) is somewhat larger, though not statistically so, among younger children ($r = .71$, $p < .001$, $n = 19$) relative to older children ($r = .45$, $p < .05$, $n = 25$; $z = 1.22$, $p = .11$).

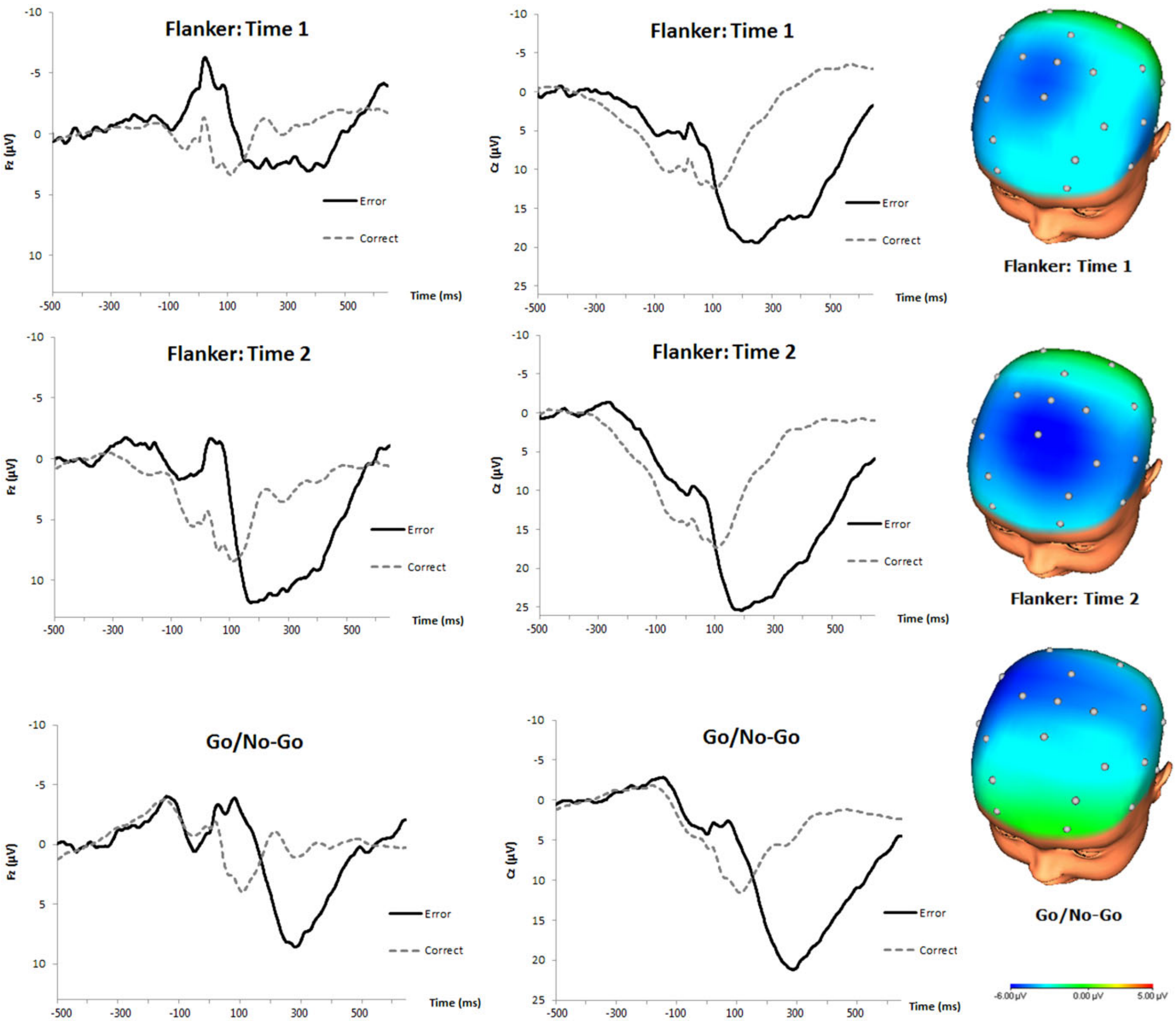

Figure 1. Response-locked ERP waveforms at Fz and Cz, during the flanker task at Time 1 and 2, as well as during the go/no-go task. Right: topographic maps depicting differences (in µV) between error and correct responses in the time range of the ERN (0–100 ms) for the flanker task at Time 1 and 2, as well as during the go/no-go task.

Pearson's correlation (*r*), was moderate to high for both ERN, $r = .70$, $p < .01$, and CRN, $r = .71$, $p < .01$.

Internal Consistency

At both time points, split-half correlations for the ERN at Fz and Cz were higher for the flanker task than for the go/no-go task (Table 1). Figure 3 depicts ERN as a function of trial number for each task.[3] A 3 (Task/Time) × 10 (ERP average: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 error trials) repeated measures ANOVA revealed a significant main effect of task at Cz, $F(2,630) = 8.39$, $p < .001$, $\eta_p^2 = .19$ but no main effect of ERP average, $F(9,630) = .93$, $p = .42$, nor a significant interaction between task and ERP average, $F(18,630) = .39$, $p = .79$, suggesting that the impact of increasing errors did not differentially impact the ERN as a function of task. At Fz, there was no main effect of task, ERP average, nor a significant interaction between task and ERP average, all *p*s > .40.

Figure 4 presents correlation coefficients between the grand average ERN and the ERN based on averages of fewer trials. All correlations are moderate to high, increasing as more trials were added (all correlations were significant at $p < .001$). For the flanker task at Time 1 and 2 and go/no-go, 6 error trials were required for the ERN to correlate > .70 with the grand average ERN at Fz and Cz.

Figure 5 presents Cronbach's alpha for the ERN for each task as more error trials were examined. A Cronbach's alpha exceeding .90 suggests excellent reliability, between .70 and .90 suggests high internal reliability, between .50 and .70 indicates moderate internal reliability, and below .50 suggests low reliability. At Cz, high internal reliability (.70) was achieved with the flanker task at both

3. Because the ERN is a relative negative deflection that overlaps with a broad positivity (i.e., the P3/Pe components), average values for the ERN are positive.

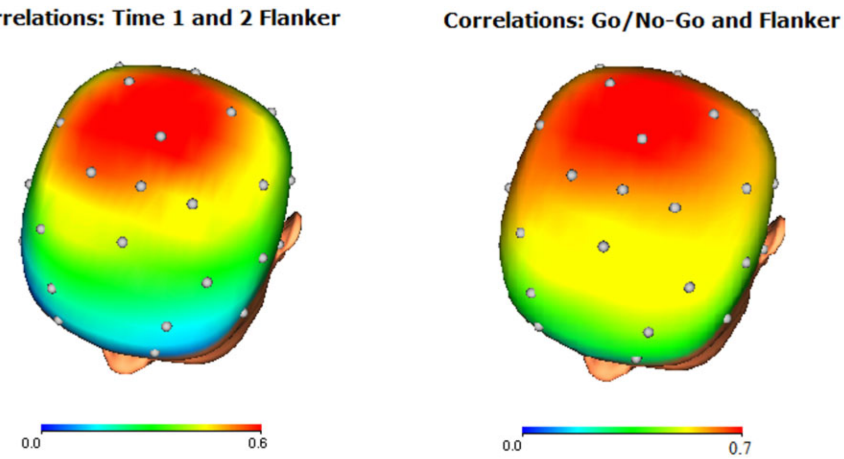

Figure 2. Left: head map of the correlation between ERN elicited by the flanker task at Time 1 and 2. Right: head map of the correlation between the ERN elicited by the flanker and go/no-go task at Time 2.

Time 1 and Time 2 after eight errors were committed; however, the reliability of the go/no-go task only reached moderate reliability (.50), even after 20 errors per subject were included (Figure 5). At Fz, high internal reliability was only achieved with the flanker task at Time 2 after 12 errors.

Discussion

The current study examined the 2-year test-retest reliability of the ERN in a sample of children and adolescents, as well as the convergent validity of the ERN elicited during flanker and go/no-go tasks. Additionally, we examined the internal consistency of the ERN across these tasks, as a function of error number. These results suggested that the ERN is reliable and stable across time and tasks—that the ERN is traitlike in children and adolescents. However, results also indicated that better internal consistency was obtained using the flanker task than the go/no-go task, which should be considered when examining the ERN in the context of individual differences.

Overall, children and adolescents were faster during the flanker task at Time 2 than at Time 1. Although RTs decreased, accuracy did not, suggesting that children and adolescents improved in performance with age or repetition of the task. This is consistent with previous findings linking age to increased performance on the flanker task (Davies et al., 2004; Meyer et al., 2012). Similar to findings in adults (Riesel et al., 2013), children and adolescents were faster and more accurate during the go/no-go task than the flanker task, suggesting this task was less demanding.

Considering that testing sessions were separated by 2 years and spanned a period of development marked by dramatic changes in brain structure and function, including changes in the ERN (Davies et al., 2004; Meyer et al., Submitted), test-retest reliability was

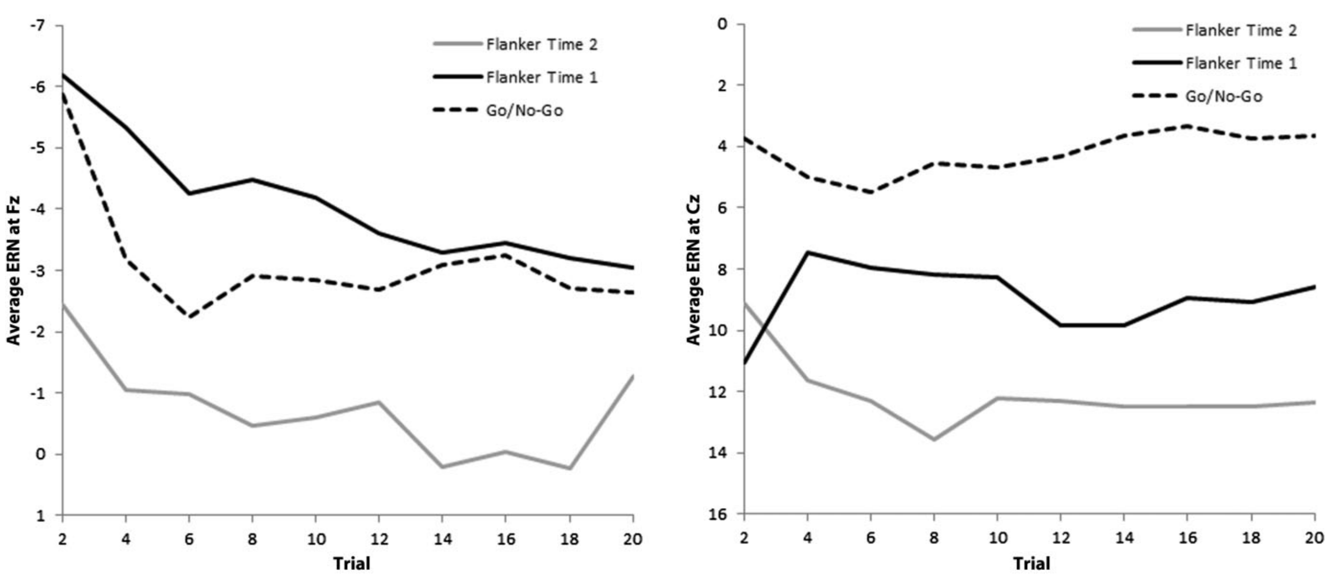

Figure 3. Magnitude of the ERN at Fz (left) and Cz (right) as a function of error number for the flanker task at Time 1 and 2, and the go/no-go task at Time 2.

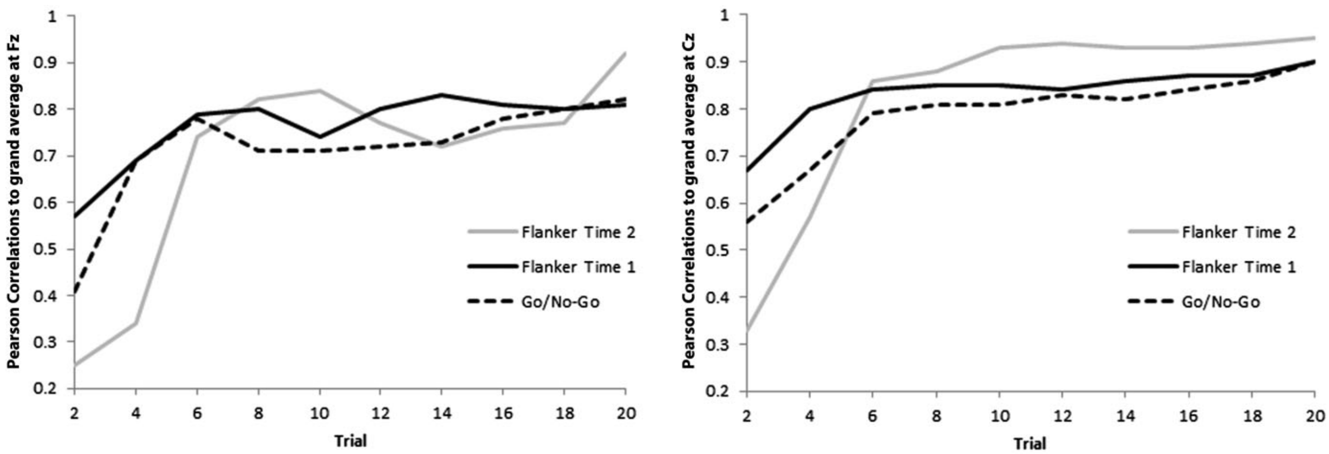

Figure 4. Correlation coefficients between the grand average ERN and ERN based on the average of fewer trials at Fz (left) and Cz (right).

remarkably high. Moreover, correlations in error-related brain activity from Time 1 to Time 2 during the flanker task were large (> .50) and maximal at central sites along the midline. This pattern of correlations is consistent with the typical scalp distribution of the ERN observed in children and adolescents (Arbel & Donchin, 2011; Torpey et al., 2013). These data indicate that central sites best reflect stable individual differences in error-related brain activity across development. These findings are consistent with a previous study in adults in which the reliability of the ERN across 2 years was approximately .67 (Weinberg & Hajcak, 2011). Follow-up analyses suggested the stability of the ERN across time may be larger among younger children than older children. Therefore, it may be important for future studies to determine whether relatively lower test-retest reliability in early adolescence relates to specific developmental changes (e.g., puberty) during this period. Overall, the results from the current study extend previous findings in adults and suggest that the ERN may be considered a stable, traitlike marker in children and adolescents.

Although the long-term test-retest reliability of the ERN was markedly similar to what has previously been reported in adults, the ΔERN (i.e., the ERN minus the CRN) was less stable in the current study. In adults, the ΔERN was stable across 2 years ($r = .69$; Weinberg & Hajcak, 2011); however, the ΔERN was not significantly correlated in the current study from Time 1 to Time 2. Recent evidence suggests that the magnitude of the ΔERN, but not the ERN alone, may be related to pubertal development (Meyer et al., Submitted), which could contribute to reduced stability in the ΔERN observed in this study. Additionally, difference scores tend to be less reliable in general (Chiou & Spreng, 1996; Edwards, 2001; Johns, 1981; Peter, Churchill, & Brown, 1993), which may be a contributing factor to the decreased reliability of the ΔERN. Furthermore, the reliability of a difference score is proportional to the average reliability of each individual score (i.e., the reliability of the ERN and CRN) minus the correlation between the scores (i.e., the correlation between the ERN and CRN; Overall & Woodward, 1975; Spreng, 1994). It is possible that the correlation between the CRN and ERN is higher in children than in adults, thereby reducing the reliability of the ΔERN. The increased stability of the ERN relative to the stability of the ΔERN in children and adolescents may suggest that the former measure is more traitlike across development.

Similar to the high test-retest reliabilities found for the ERN, high convergent validity was found measuring the ERN using the flanker and go/no-go tasks. The correlations of error-related brain activity were highest (> .60) at central sites along the midline—a pattern remarkably similar to correlations found for the flanker task between both time points. Taken together, these results suggest that the ERN measured at midcentral sites best reflects error-related

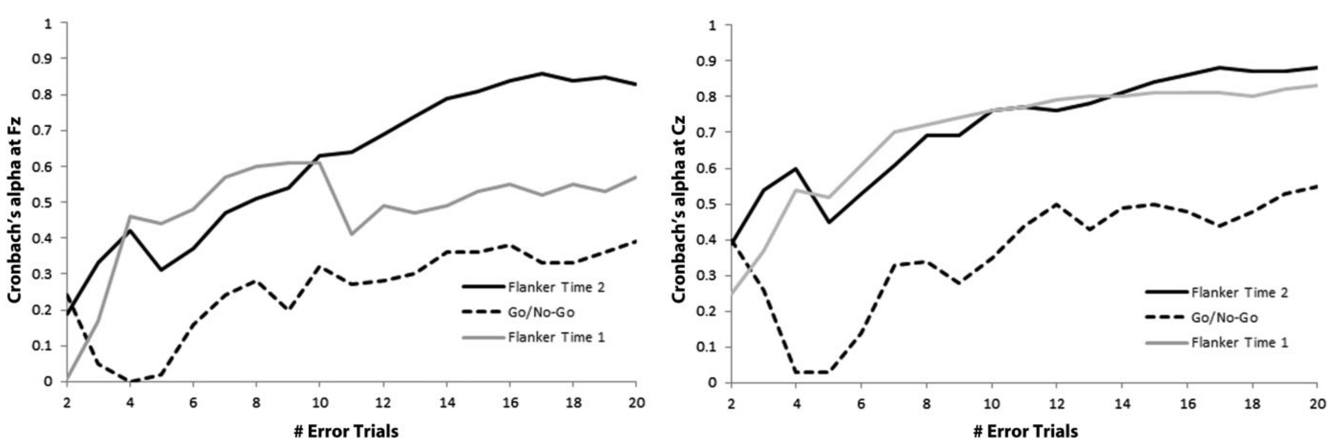

Figure 5. Cronbach's alpha of the ERN at Fz (left) and Cz (right) as a function of the number of error trials for each task.

brain activity common across the two tasks examined. Additionally, the high correlations of the ERN across tasks suggest that both the flanker and go/no-go task activate a common neural network in a sample of children and adolescents. These results are consistent with previous studies (Riesel et al., 2013; Segalowitz et al., 2010), which found correlations between .50 and .65 across a flanker and go/no-go task.

Although the ERN elicited by the flanker and go/no-go task were highly correlated, an examination of the internal consistency of the tasks revealed important psychometric differences. The split-half reliability of the ERN at Cz elicited by the flanker task at both time points ($r = .88$ and $r = .81$) was substantially higher than the ERN elicited by the go/no-go task ($r = .50$). Additionally, while Cronbach's alpha for the flanker task at Cz at both time points reached high internal reliability (.70) after 8 errors were committed, the go/no-go task achieved only moderate reliability (.50) even after 20 errors were examined, suggesting a substantial amount of trial-to-trial variation of the ERN within this task. Consistent with previous studies, the flanker task reached high internal reliability quickly (Larson et al., 2010; Meyer, Riesel, & Proudfit, 2013; Olvet & Hajcak, 2009b; Pontifex et al., 2010) even in a sample of children and adolescents, and might therefore be prioritized in studies aiming to examine the ERN in relation to traitlike individual differences.

The relatively weaker psychometric properties of the go/no-go task in the current study contrast with results from older subjects. A previous study in adults indicated that the magnitude of the ERN became larger as the go/no-go task progressed, and high split-half reliability and Cronbach's alpha suggested that the pattern was consistent across participants (Meyer, Riesel, & Proudfit, 2013). However, in the current study, the number of errors included did not impact the ERN amplitude during the go/no-go task but relatively low split-half correlations and Cronbach's alpha (even after 20 errors) suggested sizeable intraindividual variation. The differences in reliability of the go/no-go task observed in children and adolescents relative to adults may be due in part to differences in cognitive control strategies employed at different developmental stages. The go/no-go task may differ in complexity from the flanker task insofar as participants are required to inhibit responses on no-go trials, whereas participants make a response on every trial for the flanker task.

Differences also emerged in test-retest reliability, convergent validity, and internal reliability between the ERN measured at different electrode sites (i.e., Fz and Cz). Correlations for error-related brain activity across time and task were larger at Cz than at Fz. Additionally, split-half reliability and Cronbach's alpha were also higher at Cz. This is notable, considering that the ERN was maximal at Fz in the flanker task at both time points. Comparable results were obtained in adults (Riesel et al., 2013); therefore, future work should examine the ERN where it is maximal and where psychometric properties are best in children and adults to better understand this issue.

Overall, the results of the current study suggest good test-retest reliability of the ERN across 2 years in a sample of youth who spanned a relatively wide age range. Additionally, evidence indicated the convergent validity of the ERN across the two tasks was moderate. Taken together, these findings suggest the ERN is traitlike—even in children and adolescents, and therefore may be a useful biomarker of individual differences. However, differences observed in internal consistency suggest the flanker task might be prioritized when eliciting the ERN in children and adolescents. In light of the fact that the validity of an individual difference measure is dependent upon its reliability, it is important that future developmental research explore whether the differential reliability of the ERN elicited by different tasks influences its relationship with anxiety, and therefore our ability to predict risk trajectories. Additionally, future studies should explore the impact of pubertal changes on reliability of the ERN to determine to what extent the ERN may be utilized as a risk marker across development.

References

Amodio, D. M., Master, S. L., Yee, C. M., & Taylor, S. E. (2008). Neurocognitive components of the behavioral inhibition and activation systems: Implications for theories of self-regulation. *Psychophysiology*, *45*, 11–19. doi: 10.1111/j.1469-8986.2007.00609.x

Arbel, Y., & Donchin, E. (2011). When a child errs: The ERN and the Pe complex in children. *Psychophysiology*, *48*, 55–63. doi: 10.1111/j.1469-8986.2010.01042.x

Boksem, M. A. S., Tops, M., Wester, A. E., Meijman, T. F., & Lorist, M. M. (2006). Error-related ERP components and individual differences in punishment and reward sensitivity. *Brain Research*, *1101*, 92–101. doi: 10.1016/j.brainres.2006.05.004

Brooker, R. J., Buss, K. A., & Dennis, T. A. (2011). Error-monitoring brain activity is associated with affective behaviors in young children. *Developmental Cognitive Neuroscience*, *1*, 141–152. doi: 10.1016/j.dcn.2010.12.002

Carrasco, M., Harbin, S. M., Nienhuis, J. K., Fitzgerald, K. D., Gehring, W. J., & Hanna, G. L. (2013). Increased error-related brain activity in youth with obsessive-compulsive disorder and unaffected siblings. *Depression and Anxiety*, *30*, 39–46.

Chiou, J.-S., & Spreng, R. (1996). The reliability of difference scores: A re-examination. *Journal of Consumer Satisfaction Dissatisfaction and Complaining Behavior*, *9*, 158–167.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281–302. doi: 10.1037/h0040957

Davies, P. L., Segalowitz, S. J., & Gavin, W. J. (2004). Development of response-monitoring ERPs in 7- to 25-year-olds. *Developmental Neuropsychology*, *25*, 355–376.

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

Endrass, T., Klawohn, J., Schuster, F., & Kathmann, N. (2008). Overactive performance monitoring in obsessive-compulsive disorder: ERP evidence from correct and erroneous reactions. *Neuropsychologia*, *46*, 1877–1887. doi: 10.1016/j.neuropsychologia.2007.12.001

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology*, *78*, 447–455. doi: 10.1016/0013-4694(91)90062-9

Foti, D., Kotov, R., & Hajcak, G. (2013). Psychometric considerations in using error-related brain activity as a biomarker in psychotic disorders [Supplement]. *Journal of Abnormal Psychology*, *122*, 520–531. doi: 10.1037/a003261810.1037/a0032618.supp

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science*, *4*, 385–390.

Gehring, W. J., Himle, J., & Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychological Science*, *11*, 1–6. doi: 10.1111/1467-9280.00206

Gratton, G., Coles, M. G. H., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484. doi: 10.1016/0013-4694(83)90135-9

Hajcak, G. (2012). What we've learned from mistakes. *Current Directions in Psychological Science*, *21*, 101–106. doi: 10.1177/0963721412436809

Hajcak, G., Franklin, M. E., Foa, E. B., & Simons, R. F. (2008). Increased error-related brain activity in pediatric obsessive-compulsive disorder before and after treatment. *American Journal of Psychiatry*, *165*, 116–123. doi: 10.1176/appi.ajp.2007.07010143

Hajcak, G., McDonald, N., & Simons, R.F. (2003). Anxiety and error-related brain activity. *Biological Psychology*, *64*, 77–90. doi: 10.1016/s0301-0511(03)00103-0

Hajcak, G., McDonald, N., & Simons, R.F. (2004). Error-related psychophysiology and negative affect. *Brain and Cognition*, *56*, 189–197. doi: 10.1016/j.bandc.2003.11.001

Hanna, G. L., Carrasco, M., Harbin, S. M., Nienhuis, J. K., LaRosa, C. E., Chen, P., . . . Gehring, W. J. (2012). Error-related negativity and tic history in pediatric obsessive-compulsive disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*, 902–910.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*, 679–709. doi: 10.1037/0033-295x.109.4.679

Holroyd, C. B., Dien, J., & Coles, M. G. H. (1998). Error-related scalp potentials elicited by hand and foot movements: Evidence for an output-independent error-processing system in humans. *Neuroscience Letters*, *242*, 65–68. doi: 10.1016/s0304-3940(98)00035-4

Johns, G. (1981). Difference score measures of organizational behavior variables: A critique. *Organizational Behavior and Human Performance*, *27*, 443–463.

Ladouceur, C. D., Dahl, R. E., Birmaher, B., Axelson, D. A., & Ryan, N. D (2006). Increased error-related negativity (ERN) in childhood anxiety disorders: ERP and source localization. *Journal of Child Psychology and Psychiatry*, *47*, 1073–1082.

Larson, M. J., Baldwin, S. A., Good, D. A., & Fair, J. E. (2010). Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): The role of number of trials. *Psychophysiology*, *47*, 1167–1171. doi: 10.1111/j.1469-8986.2010.01022.x

Meyer, A., Bress, J., & Proudfit, G. H. Adolescence and pubertal development: Associations with the error-related negativity and feedback negativity. Manuscript submitted for publication.

Meyer, A., Hajcak, G., Torpey, D. C., Kujawa, A., Kim, J., Bufferd, S., . . . Klein, D. N. (2013). Increased error-related brain activity in six-year-old children with clinical anxiety. *Journal of Abnormal Child Psychology*, *41*, 1257–1266.

Meyer, A., Riesel, A., & Proudfit, G. H. (2013). Reliability of the ERN across multiple tasks as a function of increasing errors. *Psychophysiology*, *50*, 1220–1225.

Meyer, A., Weinberg, A., Klein, D. N., & 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*, 152–161. doi: 10.1016/j.dcn.2011.09.005

Olvet, D. M., & Hajcak, G. (2009a). Reliability of error-related brain activity. *Brain Research*, *1284*, 89–99.

Olvet, D. M., & Hajcak, G. (2009b). The stability of error-related brain activity with increasing trials. *Psychophysiology*, *46*, 957–961.

Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin*, *82*, 85–86.

Peter, J. P., Churchill, G. A., Jr., & Brown, T. J. (1993). Caution in the use of difference scores in consumer research. *Journal of Consumer Research*, *19*, 655–662.

Pontifex, M. B., Scudder, M. R., Brown, M. L., O'Leary, K. C., Wu, C.-T., Themanson, J. R., & Hillman, C. H. (2010). On the number of trials necessary for stabilization of error-related brain activity across the life span. *Psychophysiology*, *47*, 767–773.

Riesel, A., Weinberg, A., Endrass, T., Meyer, A., & Hajcak, G. (2013). The ERN is the ERN is the ERN? Convergent validity of error-related brain activity across different tasks. *Biological Psychology*, *93*, 377–385.

Santesso, D. L., Segalowitz, S. J., & Schmidt, L. A. (2006). Error-related electrocortical responses are enhanced in children with obsessive-compulsive behaviors. *Developmental Neuropsychology*, *29*, 431–445.

Segalowitz, S. J., Santesso, D. L., Murphy, T. I., Homan, D., Chantziantoniou, D. K., & Khan, S. (2010). Retest reliability of medial frontal negativities during performance monitoring. *Psychophysiology*, *47*, 260–270.

Simons, R. F., & Miles, M. A. (1990). Nonfamilial strategies for the identification of subjects at risk for severe psychopathology: Issues of reliability in the assessment of event-related potential and other marker variables *Event-related brain potentials: Basic issues and applications* (pp. 343–363). New York, NY: Oxford University Press.

Spreng, R. A. (1994). A reassessment of the reliability of difference scores in the measurement of disconfirmation of expectations. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, *7*, 114–118.

Tamnes, C. K., Walhovd, K. B., Torstveit, M., Sells, V. T., & Fjell, A. M. (2013a). 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. doi: 10.1016/j.dcn.2013.05.001

Tamnes, C. K., Walhovd, K. B., Torstveit, M., Sells, V. T., & Fjell, A. M. (2013b). 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. doi: 10.1016/j.dcn.2013.05.001

Torpey, D. C., Hajcak, G., Kim, J., Kujawa, A. J., Dyson, M. W., Olino, T. M., & Klein, D. N. (2013). Error-related brain activity in young children: Associations with parental anxiety and child temperamental negative emotionality. *Journal of Child Psychology and Psychiatry*, *54*, 854–862. doi: 10.1111/jcpp.12041

Torpey, D. C., Hajcak, G., & Klein, D. N. (2009). An examination of error-related brain activity and its modulation by error value in young children. *Developmental Neuropsychology*, *34*, 749–761.

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

Weinberg, A., Klein, D. N., & Hajcak, G. (2012). Increased error-related brain activity distinguishes generalized anxiety disorder with and without comorbid major depressive disorder. *Journal of Abnormal Psychology*, *221*, 885–896. doi: 10.1037/a0028270

Weinberg, A., Olvet, D. M., & Hajcak, G. (2010). Increased error-related brain activity in generalized anxiety disorder. *Biological Psychology*, *85*, 472–480. doi: 10.1016/j.biopsycho.2010.09.011

Weinberg, A., Riesel, A., & Hajcak, G. (2012). Integrating multiple perspectives on error-related brain activity: The ERN as a neural indicator of trait defensive reactivity. *Motivation and Emotion*, *36*, 84–100.

Wiersema, J. R., van der Meere, J. J., & Roeyers, H. (2007). Developmental changes in error monitoring: An event-related potential study. *Neuropsychologia*, *45*, 1649–1657. doi: 10.1016/j.neuropsychologia.2007.01.004

Xiao, Z., Wang, J., Zhang, M., Li, H., Tang, Y., Wang, Y., . . . Fromson, J. A. (2011). Error-related negativity abnormalities in generalized anxiety disorder and obsessive–compulsive disorder. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, *35*, 265–272. doi: 10.1016/j.pnpbp.2010.11.022

(Received September 20, 2013; Accepted February 6, 2014)