Journal of Abnormal Psychology
2013, Vol. 122, No. 2, 520–531

0021-843X/13/$12.00 DOI: 10.1037/a0032618

Psychometric Considerations in Using Error-Related Brain Activity as a Biomarker in Psychotic Disorders

Dan Foti, Roman Kotov, and Greg Hajcak
Stony Brook University

Psychotic disorders are characterized by profoundly blunted neural responses to errors, as indicated by reductions in two event-related potential (ERP) components: the error-related negativity (ERN) and error positivity (Pe). The potential utility of the ERN and Pe as biomarkers for psychotic disorders is currently limited, however, by an incomplete understanding of their psychometric properties. To address this gap in the literature, we considered the reliability and validity of these measures in both healthy individuals ($n = 52$) and patients with psychotic illness ($n = 84$) across two experimental paradigms that have been used in previous studies in schizophrenia: a flankers task and a picture/word matching task. Internal consistency reliability was higher on the flankers compared to the picture/word task overall. On the flankers task, fair internal consistency was achieved among patients with relatively few trials (ERN = five trials, Pe = 12 trials). The number of available error trials influenced reliability among patients more than among healthy individuals, and on the picture/word task more than the flankers task. Moderate convergent validity for the ERN and Pe was observed across tasks in both the patient and healthy groups. ERPs on the flankers task exhibited external validity, and were related to several clinical characteristics, including diagnosis, negative symptom severity, rehospitalization, employment, and neuroticism; associations with the picture/word task were generally weaker. These data indicate that task differences can strongly affect psychometric properties of error-related neural activity indices in healthy and patient populations.

Keywords: error-related negativity, schizophrenia, psychosis, reliability, validity

There is considerable interest in identifying biomarkers of cognitive dysfunction in psychotic disorders (Allen, Griss, Folley, Hawkins, & Pearlson, 2009; Luck et al., 2011). In fact, the National Institute of Mental Health recently launched the Research Domain Criteria (RDoC) project, which seeks to reclassify psychiatric disorders in terms of dysfunctional neural circuitry (Insel et al., 2010). Implicit in the broader RDoC framework is the assumption that neurobiological measures derived from laboratory paradigms can meaningfully capture variation across individuals. Indeed, a biological marker is only useful if it can effectively differentiate people with the target condition from everyone else. Neural measures are typically defined by *within-subjects* comparisons of experimental manipulations, which alone do not imply the existence of reliable *between-subjects* differences. For a biomarker to be an informative measure of cognitive dysfunction in psychotic disorders, the psychometric properties of that measure ought to first be established.

One of the RDoC constructs of interest is cognitive control, of which a primary constituent process is action monitoring as indicated by error-related neural activity in the anterior cingulate cortex (ACC; Ridderinkhof, van den Wildenberg, Segalowitz, & Carter, 2004)—an established neurocognitive deficit in psychotic disorders (Minzenberg, Laird, Thelen, Carter, & Glahn, 2009). To further examine impaired action monitoring in psychosis, a number of studies have leveraged the temporal resolution of event-related potentials (ERPs) and focused on two components elicited by errors on speeded reaction time tasks: the error-related negativity (ERN), which peaks within 100 ms following error commission and is maximal at frontocentral sites; and the error positivity (Pe), which peaks at approximately 200–400 ms and is maximal at centroparietal sites (Falkenstein, Hoormann, Christ, & Hohnsbein, 2000). The ERN and Pe are thought to reflect distinct facets of action monitoring, with the ERN related to automatic error detection and the Pe to conscious error recognition and response adjustment (Hajcak, McDonald, & Simons, 2003; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001). Converging ERP and neuroimaging evidence has indicated that the ERN is generated within the dorsal ACC (Taylor, Stern, & Gehring, 2007), whereas the Pe has been attributed to a nearby source in the ventral ACC (Herrmann, Römmler, Ehlis, Heidrich, & Fallgatter, 2004). Bringing this basic research to bear on clinical deficits in action monitoring related to psychosis, the ERN has consistently been shown to be blunted among individuals with schizophrenia (Alain, McNeely, He, Christensen, & West, 2002; Bates, Kiehl, Laurens, & Liddle, 2002; Bates, Liddle, Kiehl, & Ngan, 2004; Kim et al., 2006; Kopp & Rist, 1999; Mathalon, Fedor, et al., 2002; Morris, Yee, & Nuechterlein, 2006). Furthermore, recent evidence has suggested that although the ERN is similarly impaired in the prodromal stage and in other psychotic disorders, Pe impairment

Dan Foti, Department of Psychology, Stony Brook University; Roman Kotov, Department of Psychiatry and Behavioral Sciences, Stony Brook University; and Greg Hajcak, Department of Psychology, Stony Brook University.

Correspondence concerning this article should be addressed to Greg Hajcak, Department of Psychology, Stony Brook University, Stony Brook, NY 11794-2500. E-mail: greg.hajcak@stonybrook.edu

may be relatively specific to schizophrenia (Foti, Kotov, Bromet, & Hajcak, 2012; Perez et al., 2012).

These results indicate that the ERN and Pe could be useful biomarkers of cognitive dysfunction in psychosis, but several important questions about the psychometric properties of these components remain unanswered. To ensure that the ERN and Pe measure individual differences relevant to psychopathology (e.g., impaired action monitoring in psychotic disorders) and have clinical applications, a more complete understanding of their reliability and construct validity is required.

Reliability refers to the precision of a measure, reflecting the amount of true variance versus error variance in the scores. The three primary types of reliability are: (a) internal consistency, reflecting the homogeneity of items within a test; (b) test−retest, reflecting the consistency of scores over time; and (c) alternate forms, indicating agreement between scores on different versions of a test (Anastasi & Urbina, 1997; Schmidt, Le, & Ilies, 2003). With regard to ERPs, internal consistency represents the homogeneity of a component across trials, or the consistency of responses comprising the averaged ERP waveform. As noted by Simons and Miles (1990), averaging many trials to score an ERP does not ensure a reliable score—the average of unrelated data will not be reliable. Internal consistency can be quantified using split-half correlations as well as Cronbach's alpha, which represents the mean of all possible split-half correlations (Anastasi & Urbina, 1997; Cronbach & Warrington, 1951; Schmidt et al., 2003). Test–retest reliability is the correlation of an ERP with itself across time, when a single paradigm is administered at multiple assessments. Finally, alternate forms reliability is the correlation of an ERP component derived from parallel forms of a laboratory task administered to the same sample.

Of interest for the ERN and Pe is not only the reliability of each component on error trials, but also the reliability of the difference score for each (i.e., error minus correct). As noted above, ERP components are defined by within-subjects comparisons, and the ERN/Pe are understood to specifically reflect error-related neural activity because the amplitude of each component differs significantly between error and correct trials. The reliability of a difference score is proportional to the average of the reliabilities of each individual score minus the correlation between scores (Overall & Woodward, 1975; Spreng, 1994). Insofar as the number of correct trials is often large, exceeding the number of error trials by an order of magnitude or more, the reliability of the response on correct trials is excellent (Olvet & Hajcak, 2009a). The limiting factors for the reliability of the difference scores, therefore, will be twofold: unreliability of the ERN/Pe, or high correlation between neural responses on error and correct trials.

Whereas reliability is necessary for the ERN/Pe to be useful individual difference measures, the paramount psychometric consideration is whether they are also *valid* measures of action monitoring (Clark & Watson, 1995; Cronbach & Meehl, 1955). Construct validity can be established by demonstrating that an ERP component is correlated with other known measures of the same construct (i.e., convergent validity), is not correlated with measures of unrelated constructs (i.e., discriminant validity), and relates to external measures known to be linked to the construct (i.e., external validity). Convergent and discriminant validity can be evaluated using a multitrait-multimethod matrix (Campbell & Fiske, 1959), consisting of correlations between multiple ERP components elicited within multiple tasks or paradigms. Two unrelated ERP components will share only method-specific variance across subjects (e.g., skull thickness), whereas two ERP measures of the same construct ought to be more closely related due to shared trait variance. Specifically, evidence that the ERN−ERN and Pe−Pe correlations are stronger than the ERN−Pe correlations across tasks would support the notion that each component is a valid measure of a specific facet of action monitoring. To draw strong inferences about the validity of ERN or Pe amplitude as a biomarker for impaired action monitoring in psychosis, it is insufficient to simply demonstrate that either component is abnormal within a patient sample—a blunted ERP could reflect a global reduction in neural activity that is nonspecific to action monitoring. Instead, reduced ERP amplitude should also show specificity to the target psychological process (i.e., correlate with other measures of it and be distinct from indicators of other processes).

A first step in this process is to examine the psychometric properties of the ERN and Pe in nonclinical samples, and several recent studies on this topic have yielded favorable results. For both components, internal consistency is fair to excellent, with Cronbach's alpha exceeding .70 with just 10 error trials (Olvet & Hajcak, 2009b; Pontifex et al., 2010). Attaining reliable error-related ERPs with relatively few trials is a key factor for acceptability of a measure in clinical settings, as well as minimizing patient burden and maximizing tolerability. Test–retest reliability is moderate to high, with 2- to 6-week estimates of .40−.82 (Olvet & Hajcak, 2009a; Segalowitz et al., 2010) and 2-year estimates of .56−.67 (Weinberg & Hajcak, 2011). There is some evidence of convergent validity as well: the ERN across flankers and go/no-go tasks was moderately correlated (.54) (Segalowitz et al., 2010). Thus, the ERN and Pe are reliable and potentially valid neural measures of action monitoring, making them promising candidates to examine individual differences in error processing relevant to psychopathology (Hajcak, 2012; Olvet & Hajcak, 2008).

Of import, the measurement characteristics of the ERN and Pe in patient populations remain unknown. This is a significant shortcoming because if reliabilities of these ERPs differ between healthy and clinical populations, this can lead to spurious findings or null effects (Chapman & Chapman, 1973). Furthermore, the ERN and Pe have been studied in psychosis using a wide range of laboratory paradigms, including flankers, Stroop, go/no-go, and picture/word matching tasks. If convergent validity among tasks is low in patients, for example, relations of ERPs with symptoms and functioning in this population may vary depending on the task employed. This would give the misleading impression of inconsistent findings, when in fact different tasks may be tapping into distinct neural processes.

To begin to address these issues, we examined the ERN and Pe in psychotic and healthy populations using both flankers and picture/word tasks. First, we calculated the internal consistency of the ERN and Pe. In light of previous data (Olvet & Hajcak, 2009b; Pontifex et al., 2010), we expected that the internal consistency of the flankers ERN and Pe would be high within the healthy sample, and we compared this reliability measure across tasks and groups. We also assessed convergent validity by testing the consistency of ERN and Pe amplitudes across tasks. Within the healthy sample, we expected the flankers ERN to correlate moderately with the picture/word ERN (Segalowitz et al., 2010), and we also evaluated

this association in the psychotic sample. Further, we assessed discriminant validity by comparing these convergent correlations to correlations between the ERN and Pe across tasks. We predicted that the flankers ERN would correlate more strongly with the picture/word ERN than with the picture/word Pe, and vice versa. With regard to external validity, in a prior report from this sample, we found that the ERN and Pe on the flankers task were impaired in psychotic disorders, and that a blunted ERN in particular related to negative symptom severity, history of rehospitalization, and unemployment, while increased ERN amplitude related to neuroticism (Foti et al., 2012). Of interest here was whether these associations with clinical characteristics would also be evident in the picture/word task—and how psychometric properties of ERPs from each task might influence these findings.

Method

Participants

The patient group consisted of 104 individuals with a history of psychosis; 48 had a longitudinal consensus diagnosis of a schizophrenia spectrum disorder (i.e., schizophrenia, schizoaffective), and 56 had other psychotic disorders (i.e., affective psychosis, substance induced, or not otherwise specified). The sample was drawn from the Suffolk County Mental Health Project (Bromet et al., 2011; Bromet et al., 1992), an epidemiologic longitudinal study of first-admission psychosis. Participants were recruited from 12 inpatient psychiatric facilities from 1989–1995; eligibility criteria included the presence of psychosis, age 15–60 years at admission, and ability to provide informed consent. On average, the present ERP assessment was conducted 15 years after first admission (range: 12.4–19.1 years). Participants were excluded from the present analysis for poor performance (<75% correct trials), poor quality ERP data (>50% trials with artifacts), for having fewer than two artifact-free error trials, or for declining to complete the task (see Figure 1). Seventy-six patients had available ERP data on the flankers task (sex = 67.1% male; age: M = 43.34 years, range: 28–67; race = 59 White, 17 other), and 84 had available data on the picture/word task (sex = 64.3% male; age: M = 43.75 years, range: 30–68; race = 64 White, 21 other); 70 had available data on both tasks (sex = 68.6% male; age: M = 42.97 years, range: 30–67; race = 55 White, 15 other).

The healthy group was comprised of 61 individuals with no history of Axis I diagnoses, no current psychiatric medication usage, and no history of traumatic brain injury or neurological illness. Healthy individuals were administered the Structured Clinical Interview for DSM–IV Disorders (SCID; First, Spitzer, Gibbon, & Williams, 2002) by master's-level clinicians. As shown in Figure 1, 52 healthy individuals had available ERP data on the flankers task (sex = 50.0% male; age: M = 39.00 years, range:

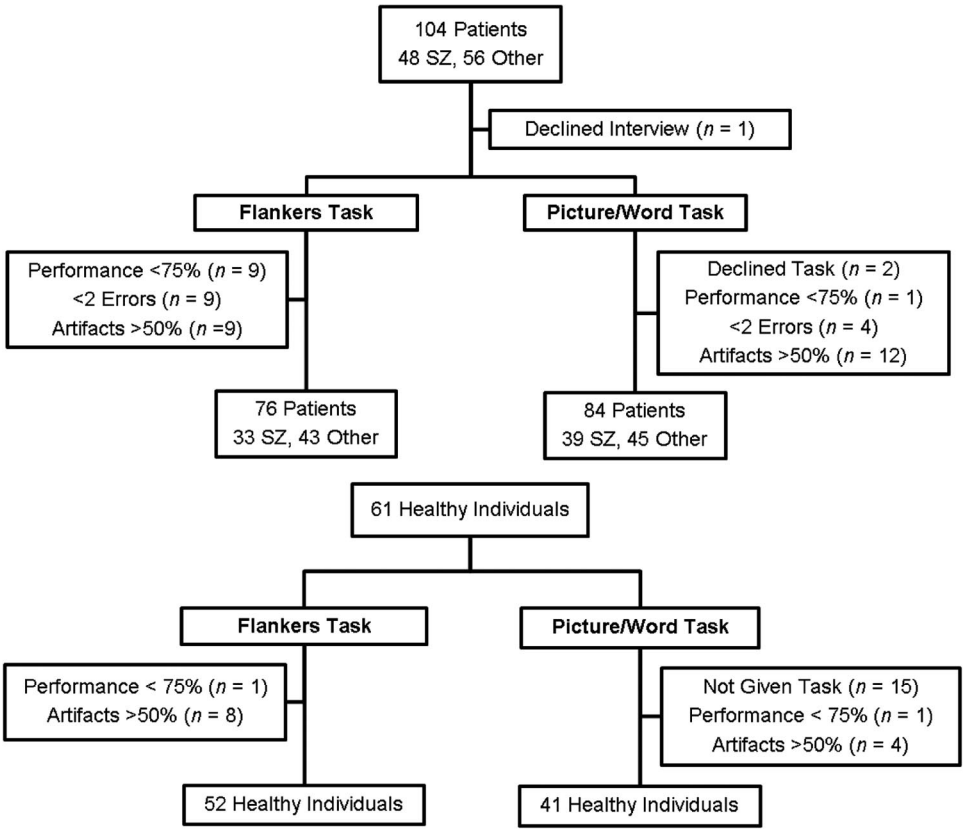

Figure 1. Patients with psychotic disorders (top) and healthy individuals (bottom) with available event-related potential data on the flankers and picture/word tasks. SZ = schizophrenia or schizoaffective disorder; Other = affective psychosis, substance induced, not otherwise specified.

18–65; race = 34 White, 18 other). A subgroup of 46 individuals also completed the picture/word task, and of these, 41 had available ERP data (sex = 39.0% male; age: M = 36.34 years, range: 18–63; race = 26 White, 15 other). Thirty-five had available data on both tasks (sex = 40.0% male; age: M = 35.31 years, range: 18–58; race = 22 White, 13 other). This research was approved by the institutional review board at Stony Brook University.

Task and Materials

Individual difference measures. We considered information on symptoms and functioning that were related to the flankers ERN in a prior report on this sample (Foti et al., 2012). Past-month symptoms of psychosis were rated using the Scale for the Assessment of Positive Symptoms (SAPS; Andreasen, 1983b) and the Scale for the Assessment of Negative Symptoms (SANS; Andreasen, 1983a). Ratings were made by two master's-level interviewers, and reliability was excellent (average intraclass rs = .83 for SANS and SAPS). Based on a prior factor analysis (Kotov, Guey, Bromet, & Schwartz, 2010), the SANS was scored as a single index and the SAPS as two symptom subscales: Psychotic (hallucinations, delusions) and Disorganized (bizarre behavior, thought disorder). Symptom information was obtained using the SCID (First et al., 2002). Personality traits were assessed with the 44-item Big Five Inventory (BFI), a measure of the five general personality dimensions (John & Srivastava, 1999); of interest here was the Neuroticism subscale.

Three indicators of real-world functioning were obtained from previous assessments of the sample: rehospitalizations during the early illness phase (within 4 years of first admission; coded as 0/1 vs. 2+), rehospitalizations during the later phase (between years 5–10; coded as 0/1 vs. 2+), and employment status (employed vs. not at the 10-year assessment). Associations among these individual difference measures were modest: a diagnosis of schizophrenia was associated with negative symptom severity (r = .53); all other correlations were less than .40 (see Appendix).

Flankers task. An arrow flankers task was used to elicit error-related ERPs (Eriksen & Eriksen, 1974). On each trial, five arrowheads were presented, with half of the trials being compatible ("< < < < <" or "> > > > >") and half incompatible ("< < > < <" or "> > < > >"). The arrows were presented in the center of a 19-in (48.3-cm) monitor and, at a viewing distance of approximately 24 in. (61 cm), occupied 1.3° of the visual field vertically and 9.2° horizontally. The arrows were presented for 200 ms, and were followed by an intertrial interval that varied randomly from 2,300–2,800 ms. Participants were instructed to press the left or right mouse button, corresponding to the direction of the center arrow on that trial, and to respond in such a way as to maximize both speed and accuracy. Participants first completed a practice block of 30 trials; the actual task consisted of 11 blocks of 30 trials (i.e., 330 trials total). At the end of each block, participants received feedback based on their performance: Performance greater than 75% correct was followed by "Please try to be more accurate"; performance greater than 90% by "Please try to respond faster"; and intermediate performance by "You're doing a great job."

Picture/word task. A picture/word matching task was also used to elicit error-related ERPs (Mathalon, Fedor, et al., 2002). The pictures consisted of 102 line drawings selected for nameability (Snodgrass & Vanderwart, 1980). On each trial, one picture was presented, followed by a word that could either match (50% of trials) or not match (50% of trials). For example, a picture of a shirt followed by the word "shirt" would be a match, whereas the word "sweater" or another unrelated word would be a nonmatch. All stimuli were presented in the center of the monitor; on average, pictures occupied 4.7° of the visual field vertically and horizontally, and words occupied 1.4° vertically and 5.6° horizontally. The picture was presented first for 250 ms, and the word was presented 75 ms later until a response was made. Participants were instructed to indicate a match or nonmatch using the left and right mouse buttons, with button designation counterbalanced across participants. As on the flankers task, participants were instructed to respond in such a way as to maximize both speed and accuracy. Participants first completed a practice block of 20 trials; in the actual task, each of the 102 pictures was presented once in each of four blocks (i.e., 408 trials total).

The picture/word task is also effective at eliciting the N400, an ERP index of semantic processing (Mathalon, Faustman, & Ford, 2002). One advantage of this task is that abnormalities in both error and semantic processing can be assessed in a single paradigm. N400 data from this sample will be presented in a separate report.

Procedure

At the beginning of the session, the study was described and written informed consent was obtained. Eligibility of healthy individuals was confirmed using the SCID. Patients completed interview measures and the BFI. Next, both groups participated in the electroencephalography (EEG) assessment. They performed multiple tasks during the experiment, and the order of the tasks was counterbalanced across subjects. Patients received $100 for their participation, and healthy individuals received either $80 or $95, depending on the length of the session.

EEG Recording, Processing, and Data Reduction

The continuous EEG was recorded using an elastic cap and the ActiveTwo system (BioSemi, Amsterdam, Netherlands). For the patient and healthy groups, the signal was digitized at 24-bit resolution with a least significant bit value of 31.25 nV and a sampling rate of 1024 Hz, using a low-pass fifth-order sinc filter with −3 dB cutoff point at 205 Hz; recordings were taken from 34 scalp electrodes based on the 10/20 system (including FCz and Iz). Two electrodes were placed on the left and right mastoids, and each electrode was measured with respect to a common mode sense active electrode that formed a monopolar channel. The electro-oculogram was recorded from four facial electrodes: two were placed approximately 1 cm above and below the right eye, one 1 cm to the left of the left eye, and one 1 cm to the right of the right eye.

Offline analysis was performed using BrainVision Analyzer software (Brain Products, Munich, Germany). Data were rereferenced to the mastoid average and bandpass filtered with cutoffs of 0.1 and 30 Hz. The EEG was segmented for each trial, spanning −400 to 800 ms relative to the response, and corrected for blinks and eye movements (Gratton, Coles, & Donchin, 1983). Channels were rejected in each trial using a semiautomated pro-

cedure, with artifacts defined as: a step of more than 50.0 μV between samples, a difference of 300 μV within a trial, or a maximum difference of less than .50 μV within 100-ms intervals. Additional artifacts were identified using visual inspection. Response-locked ERP averages were created for correct and incorrect responses, and activity from −400 to −200 ms served as the baseline. The ERN was scored as the mean activity from 0–100 ms at Cz on errors, and the Pe as the mean from 200–400 ms at Pz. To assess psychometric properties, single-trial ERPs were also scored on artifact-free error trials, using the same criteria.

Data Analysis

Internal consistency was estimated using two approaches. First, we calculated split-half reliability by taking the correlation between even and odd error trials, and then by adjusting with the Spearman−Brown prophecy formula (Anastasi & Urbina, 1997). The advantages of this approach are that it uses all ERP data from each participant and does not require all participants to have the same number of available trials. We then repeated this calculation for individuals with ≥ 5 and ≥ 20 errors to examine the influence of the number of errors. Second, we calculated Cronbach's alpha and evaluated it as a function of increasing number of error trials; this approach allowed us to estimate the minimum number of error trials necessary to establish a reliable neural measure. Alpha is considered the best measure of internal consistency because it is not dependent on a particular split (i.e., it is the average of all possible split-halves); however, its calculation requires the same number of analyzable trials for each participant. The full sample was available only when calculating alpha using the first two error trials; as more trials were entered into the calculation, the number of participants with sufficient available data decreased. The reliability of the difference scores was also calculated separately, as a function of the split-half reliabilities of the ERN/Pe, that of the analogous correct responses, and the correlations between the error and correct responses; equal variance was assumed (Overall & Woodward, 1975; Spreng, 1994):

$$\alpha_{diff} = \frac{0.5(r_x + r_y) - r_{xy}}{1 - r_{xy}}$$

In all reliability analyses, the clinical utility of a measure was considered to be unacceptable for values below .70, fair for values of .70−.79, good for values of .80−.89, and excellent for values of .90 and above (Cicchetti, 1994).

Convergent and discriminant validity was assessed by correlating ERP difference scores (i.e., error minus correct) across the flankers and picture/word tasks, calculated separately for patient and healthy groups; this approach yielded the full multitrait-multimethod matrix (Campbell & Fiske, 1959). Two characteristics—automatic error detection (indicated by ΔERN amplitude) and conscious error recognition (indicated by ΔPe amplitude)—were each assessed using two methods, the flankers and picture/word tasks. External validity was examined by relating ERP difference scores from each task to the symptom, personality, and functioning measures. For this latter set of analyses, we converted the ERN to a numerically positive number such that positive correlation coefficients indicated a direct association and vice versa (cf. Foti et al., 2012). Bonferroni correction was used within each domain containing multiple comparisons (symptoms: positive, negative, disorganized; functioning: early rehospitalization, late rehospitalization, employment status).

Results

Behavior

On the flankers task, errors were more common on incompatible compared to compatible trials among both patients (compatible: $M = 6.04$, $SD = 5.57$; incompatible: $M = 17.08$, $SD = 11.44$), $t(74) = 10.25$, $p < .001$, and healthy individuals (compatible: $M = 6.67$, $SD = 6.50$; incompatible: $M = 23.58$, $SD = 12.28$), $t(51) = 10.59$, $p < .001$, suggesting that the task functioned similarly within in each group. Patients made fewer errors than healthy individuals on incompatible, $t(126) = 3.14$, $p < .01$, but not compatible trials ($p = .56$).

On the picture/word task, errors were more common on match compared to nonmatch trials only among healthy individuals (match: $M = 16.90$, $SD = 12.38$; nonmatch: $M = 11.10$, $SD = 9.80$), $t(40) = 3.25$, $p < .01$; among patients, errors were equally common on match and nonmatch trials (match: $M = 9.83$, $SD = 10.34$; nonmatch: $M = 10.61$, $SD = 10.40$), $t(83) = .74$, $p = .46$. This pattern suggests that the context in which errors were elicited on the picture/word task differed across groups. Patients made fewer errors than healthy individuals on match, $t(123) = 3.36$, $p < .01$, but not on nonmatch trials ($p = .80$).

ERPs

Across both the flankers and picture/word tasks, the ERN was observed as a frontocentral negativity on error trials, peaking at approximately 50 ms (see Figure 2). The Pe was observed as a centroparietal positivity, maximal between 200 and 400 ms (see Figure 3). Although the ERN and Pe were noticeably smaller among the patients, the timing and scalp distributions were comparable to that of the healthy individuals.

Reliability

Internal consistency. Split-half reliability estimates are presented in Table 1. On the flankers task, reliability was good among healthy individuals for both the raw scores (ERN, Pe) and difference scores (ΔERN, ΔPe; $r > .80$). Among patients, reliability of the Pe was fair for the raw and difference score ($r > .70$), but reliability of the ERN was unacceptable. The reliabilities of both components were fair when analysis was restricted only to those patients with five or more errors ($r > .70$) and were good among patients with 20 or more errors ($r \geq .80$). For both patients and healthy individuals, reliability was invariably lower for the picture/word task than for the flankers task (average difference in reliabilities = .25). Among healthy individuals but not patients, difference score reliabilities were fair among participants with five or more errors ($r > .70$). When only those participants with 20 or more error trials were considered, the reliabilities of the raw scores on the picture/word task improved substantially and were fair to good ($r > .70$); the reliabilities of the difference scores remained unacceptable among patients. High correlations between the ERPs

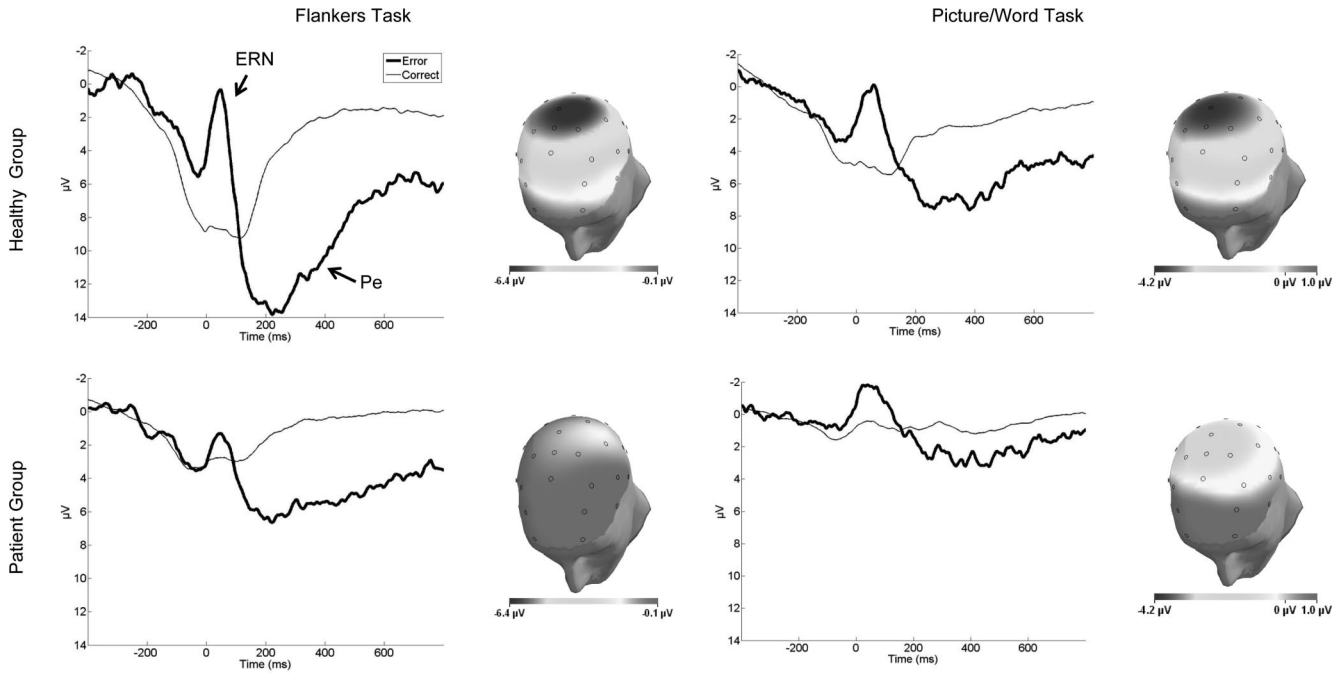

Figure 2. Error-related negativity (ERN) across tasks, presented for the healthy group (top) and patients with psychotic disorders (bottom). Waveforms represent activity at electrode Cz, and head maps represent the error minus correct difference from 0−100 ms, corresponding to how the ERN was scored. Pe = error positivity.

on error and correct trials in this group limited reliability of ΔERN/ΔPe.

Cronbach's alpha estimates as a function of the number of error trials entered into the ERP average are presented in Figures 4 and 5. Estimates for the ERN were higher for the flankers compared to the picture/word task among both groups (see Figure 4). Among patients, fair reliability ($\alpha > .70$) was achieved for the flankers ERN with only five trials and good reliability ($\alpha > .80$) with 12 trials. Among healthy individuals, fair reliability of the flankers ERN was achieved with eight trials, and good reliability required more than 20 trials. For the picture/word ERN, even 20 trials yielded estimates of reliability that only approached the fair range (patients: $\alpha = .68$, healthy: $\alpha = .65$). We note that these estimates are lower than the corresponding values in Table 1 in part because alpha is based only on the first 20 error trials, whereas the split-half estimates used all ERP data (20+ trials). The reliability of the Pe, on the other hand, was comparable across tasks (see Figure 5). Among patients, fair Pe reliability was achieved with 12 trials on the flankers task and with 19 trials on the picture/word task; among healthy individuals, the flankers task yielded fair reliability with 14 trials and good reliability with 19 trials, while the picture/word task yielded fair reliability with 13 trials.

Convergent and Discriminant Validity

The multitrait-multimethod matrix is presented in Table 2. Difference scores were used, and this analysis was restricted to participants with at least five artifact-free error trials on each task. Convergent validity was assessed by correlating the same ERP component across tasks (e.g., flankers ΔERN with picture/word ΔERN); discriminant validity was assessed by correlating different ERP components across tasks (e.g., flankers ΔERN with picture/word ΔPe). Unadjusted correlations are presented in the matrix below the diagonal. Within both groups, convergence between tasks was clear, with convergent correlations (average = .51) substantially higher than discriminant correlations (average = .22). Significant differences between convergent and discriminant correlations were observed for both components among patients ($\Delta ERN_{pw}-\Delta ERN_{f}$ vs. $\Delta ERN_{pw}-\Delta Pe_{f}$: $z = 2.36$, $p < .05$; $\Delta Pe_{f}-\Delta Pe_{pw}$ vs. $\Delta Pe_{f}-\Delta ERN_{pw}$: $z = 3.14$, $p < .01$) and controls ($\Delta ERN_{pw}-\Delta ERN_{f}$ vs. $\Delta ERN_{pw}-\Delta Pe_{f}$: $z = 2.16$, $p < .05$; $\Delta Pe_{f}-\Delta Pe_{pw}$ vs. $\Delta Pe_{f}-\Delta ERN_{pw}$: $z = 2.12$, $p < .05$); other comparisons were in the expected direction but not statistically significant (z range: 0.47–1.72, $ps > .20$). Convergence is limited by the reliability of the indices, because measurement error within each task will dilute the association among ERPs; disattenuated correlations are presented in the matrix above the diagonal. After correction for this attenuation, the convergent correlations were large (average = .75) and remained greater than discriminant correlations (.33).[1]

[1] Correction for attenuation was motivated by the observation of different reliabilities across tasks, which would otherwise have precluded meaningful task-wise comparisons of event-related potentials (Charles, 2005). This approach ought to be used with caution, and the results verified in larger samples (Nunnally & Bernstein, 1994). An alternative would have been to limit the analysis to participants with sufficient data to achieve higher internal consistency, although this would have further reduced the sample size. When only those patients with 20 or more errors on each task were considered ($n = 16$), an identical pattern of convergence was found: Convergent correlations of .62 (ΔERN) and .52 (ΔPe) were substantially higher than the discriminant correlations (.43 and .26, respectively).

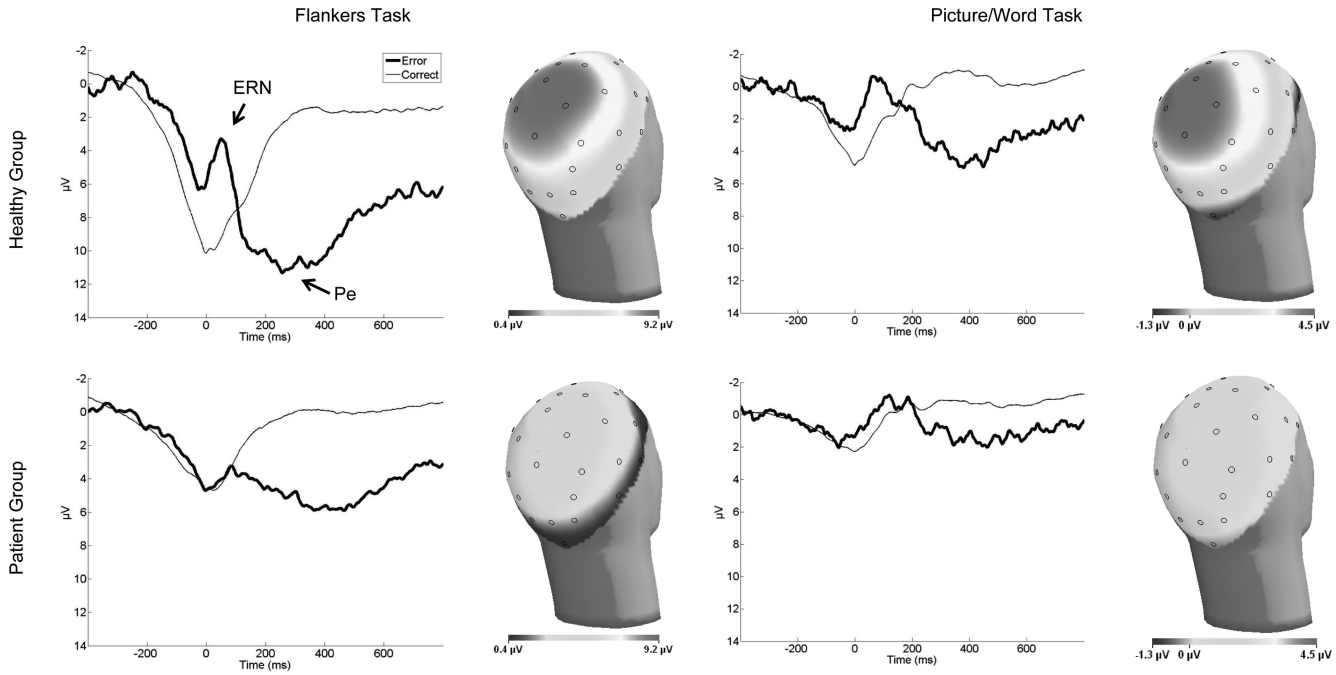

Figure 3. Error positivity (Pe) across tasks, presented for the healthy group (top) and patients with psychotic disorders (bottom). Waveforms represent activity at electrode Pz, and head maps represent the error minus correct difference from 200−400 ms, corresponding to how the Pe was scored. ERN = error-related negativity.

The unadjusted convergent correlations were comparable across groups. After correction for attenuation, however, convergence was lower in the healthy group compared to patients because reliability of picture/word task was higher in the healthy group. On the picture/word task, in healthy participants, the correlation between the ΔERN and ΔPe was much lower and in the opposite direction than in the psychosis group (−.07 and .37, respectively; $z = 2.06$, $p < .05$); this is in contrast to the flankers task, where it was comparable (.38 and .33, respectively, $p = .80$). This pattern suggests that group differences in convergence were driven by the picture/word task, which behaved very differently in the two populations, and is consistent with the lower observed internal consistency on the picture/word task overall.

External Validity

Associations between ERP variables and patient characteristics are presented in Table 3. Analyses were restricted to patients with a number of trials sufficient to attain fair internal consistency on the raw score from error trials (Cronbach's alpha > .70, taken from Figures 4–5) as well as the error minus correct difference (split-half $r > .70$). This was only achieved for the flankers task:

Table 1
Split-Half Reliability of the ERN and Pe Across Tasks

	Patient group			Healthy group		
Task	All	≥5 errors	≥20 errors	All	≥5 errors	≥20 errors
Flankers						
ERN	.63	.75	.89	.86	.86	.78
ΔERN	.48	.78	.85	.84	.84	.76
Pe	.75	.78	.80	.81	.81	.93
ΔPe	.73	.75	.81	.83	.83	.93
Picture/word						
ERN	.35	.51	.72	.41	.54	.83
ΔERN	.40	.48	.48	.69	.76	.92
Pe	.28	.45	.71	.66	.64	.76
ΔPe	.39	.46	.45	.79	.78	.71
No. of Participants						
Flankers	76	69	38	52	52	33
Picture/word	84	74	27	41	39	20

Note. Split-half reliability on correct trials was excellent in all cases ($r > .95$). ERN = error-related negativity; Pe = error positivity.

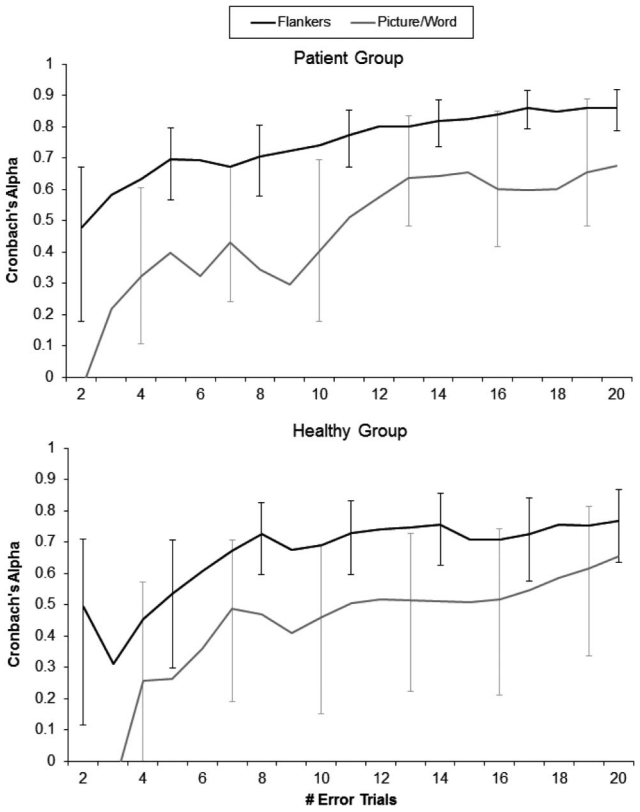

Figure 4. Cronbach's alpha of the error-related negativity for patients with psychotic disorders (top) and the healthy group (bottom), as a function of the number of trials. Error bars display representative 95% confidence intervals. Patients with available data ranged from 76 (two trials) to 38 (20 trials) on the flankers task, and from 84–25 on the picture/word task; the number of healthy individuals ranged from 52–33 on the flankers task and 41–21 on the picture/word task.

five trials were sufficient to achieve fair reliability for the ERN and ΔERN, while 12 trials were sufficient for the Pe and ΔPe. For the picture/word task, reliability was lower and a minimum of 20 error trials was used.

Consistent with a prior report from this sample (Foti et al., 2012), ERPs measured on the flankers task related to a number of clinical variables: A blunted ΔERN was associated with severity of negative symptoms, rehospitalization during the early phase of illness, and unemployment at the 10-year assessment; an enhanced ERN was associated with trait neuroticism. A blunted Pe was associated with a schizophrenia spectrum diagnosis (vs. other psychotic disorders), as well as rehospitalization during the later phase of illness.

On the picture/word task, statistical power was reduced due to the drastically restricted sample size; a significant association was found only between blunted ΔPe amplitude and severity of negative symptoms. Comparing effect sizes across tasks, a similar pattern was observed for the picture/word task as was with the flankers task for associations with two indicators: blunted ΔERN amplitude and negative symptom severity (flankers: $r = -.24$, picture/word: $r = -.25$), as well as blunted ΔPe amplitude and a schizophrenia spectrum diagnosis (−.29 vs. −.22). On the other hand, smaller effects of the picture/word task were observed between blunted ΔERN amplitude and early rehospitalization (−.26 vs. −.12), ΔERN amplitude and unemployment (−.39 vs. −.03), and ΔPe amplitude and later hospitalization (−.43 vs. −.08). The association between ΔERN amplitude and neuroticism was weaker and in opposite direction from that of the flankers task (.25 vs. −.11).

Discussion

Building on prior work examining the psychometric properties of the ERN and Pe in unselected samples (Olvet & Hajcak, 2009a, 2009b; Pontifex et al., 2010; Segalowitz et al., 2010; Weinberg & Hajcak, 2011), the current study indicates that these ERP components can be reliably measured in both healthy and patient populations. There are important differences across tasks, however, which influence both the reliability of the ERP responses and their associations with illness characteristics. On the flankers task, 12 error trials were sufficient to yield good reliability for the ERN and fair reliability for the Pe among patients, suggesting that this task is well-suited to be used as a brief assessment tool for error-related neural activity. In contrast, the reliability of the ERN and Pe

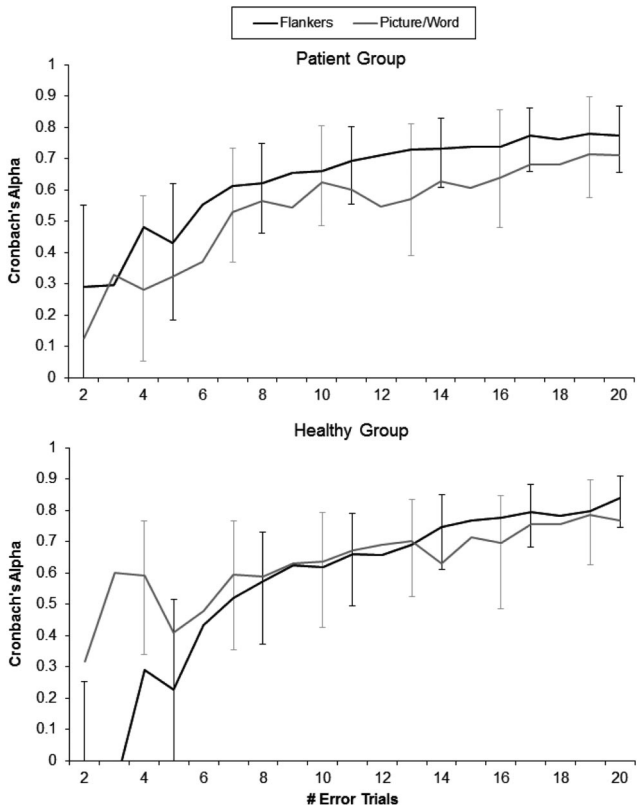

Figure 5. Cronbach's alpha of the error positivity for patients with psychotic disorders (top) and the healthy group (bottom), as a function of the number of trials. Error bars display representative 95% confidence intervals. Patients with available data ranged from 76 (two trials) to 38 (20 trials) on the flankers task, and from 84–27 on the picture/word task; the number of healthy individuals ranged from 52–33 on the flankers task and 40–20 on the picture/word task.

Table 2
Convergent and Discriminant Validity: Multitrait-Multimethod Matrix

	Flankers		Picture/word	
Task by Group	ΔERN	ΔPe	ΔERN	ΔPe
Patients (n = 58)				
Flankers				
ΔERN	(.78)	.50	**.80**	*.72*
ΔPe	.38	(.75)	*.28*	**.99**
Picture/word				
ΔERN	**.49**	*.17*	(.48)	.79
ΔPe	*.43*	**.58**	.37	(.46)
Healthy (n = 35)				
Flankers				
ΔERN	(.84)	.40	**.56**	*.27*
ΔPe	.33	(.83)	*.05*	**.66**
Picture/word				
ΔERN	**.45**	*.04*	(.76)	−.09
ΔPe	*.22*	**.53**	−.07	(.78)

Note. Participants with ≥ 5 errors on each task were considered. Diagonal denotes the split-half reliability (in parentheses). Below the diagonal denotes the unadjusted correlations. Above the diagonal denotes the disattenuated correlations. Boldface denotes convergent correlations, while italics denotes discriminant correlations. ERN = error-related negativity; Pe = error positivity.

elicited by the picture/word task was lower; fair internal consistency was achieved after only 20 error trials, and the duration of the task would need to be increased considerably to ensure that all participants commit at least 20 errors. The number of available error trials influenced the reliability of ERPs among patients more than healthy individuals, and on the picture/word task more than the flankers task. That is, restricting analysis to those participants with a large number of errors yielded a greater improvement in reliability estimates among patients and within the picture/word task. Although group differences in psychosis have been previously found using both tasks (Foti et al., 2012; Kopp & Rist, 1999; Mathalon, Fedor, et al., 2002; Morris et al., 2006; Perez et al., 2012), the flankers task appears to produce more replicable results than the picture/word task. The difference scores for each component on the flankers task (i.e., error minus correct) also yielded fair reliability with relatively few trials, indicating that the flankers task is effective for capturing neural activity specific to error processing among psychotic patients.

With the patient and healthy samples, we found evidence of moderate convergent validity for the ERN and the Pe. Also, convergent correlations were higher than discriminant correlations, indicating discriminant validity for both components. After correcting for unreliability of measurement, components were strongly related across the two paradigms. In other words, the relative amplitude of error-related brain activity among individuals was similar whether error-related brain activity was measured on the flankers or picture/word tasks, but it was reduced by unreliability in the components. Of note, disattenuated correlations were not perfect, and there were some true differences between components across tasks.

Indeed, external validation revealed that ERP responses across tasks differentially relate to features of psychotic illness. After selecting patients with a sufficient number of trials to yield reliable ERP measures, we confirmed the results of a prior report from this sample (Foti et al., 2012): On the flankers task, a blunted ERN related to negative symptom severity, rehospitalization during the early phase of illness, unemployment, and reduced neuroticism; a blunted Pe differentiated schizophrenia from other psychotic disorders.

Within the picture/word task, relatively few patients had a sufficient number of trials to yield reliable ERP data, and statistical power was limited. Although not statistically significant, observed effect sizes indicated possible overlap with the flankers task: a blunted ERN was linked to negative symptom severity, and a blunted Pe was linked to schizophrenia diagnosis. Associations with employment, rehospitalization, and neuroticism, however,

Table 3
External Validity: Associations With Clinical Features and Patient Characteristics

	Flankers		Picture/word	
Correlations among patients	ΔERN ≥5 errors	ΔPe ≥12 errors	ΔERN ≥20 errors	ΔPe ≥20 errors
Δ score split-half reliability (n)	.78 (69)	.71 (51)	.48 (27)	.45 (27)
Schizophrenia vs. other psychoses	−.20	−**.29***	−.06	−.22
Symptom severity				
Negative	−**.24***	−.14	−.25	−**.38***
Psychotic	−.16	−.04	.06	−.02
Disorganized	−.13	−.06	.00	−.21
Real-world functioning				
Rehospitalization, first 4 years	−**.26***	−.17	−.12	−.04
Rehospitalization, years 5–10	−.05	−**.43****	−.08	−.08
Unemployed, year 10	−**.39****	−.12	−.03	−.23
Neuroticism	**.25***	−.13	−.11	−.03

Note. Analyses were restricted to patients with a sufficient number of trials to yield fair internal consistency of event-related potential variables on raw and difference scores (Cronbach's alpha or split-half $r > .70$). Difference scores were used, and the ERN difference was converted to a positive number such that positive correlation coefficients reflected a direct association, and vice versa. ERN = error-related negativity; Pe = error positivity. Boldface denotes significant correlations.
* $p < .05$ (uncorrected). ** $p < .017$ (threshold following Bonferroni correction within each domain).

were negligible. There was a significant association between a blunted picture/word Pe and severity of negative symptoms, an effect that was not apparent on the flankers task. Together, these results indicate that, even though there was evidence of moderate convergent validity among patients, ERP indices of error processing elicited by flankers and picture/word tasks may relate to different characteristics of psychotic illness. In addition, the psychometric properties of the tasks—particularly the differences in internal consistency—may partially account for these differential associations with external measures.

It is noteworthy that, although moderate convergent validity was observed within each group, convergence was somewhat lower within the healthy population after disattenuation. This group difference appears to be driven by the picture/word task, where the ERN−Pe correlation was much lower than in patients. By contrast, the ERN and Pe were strongly related among both healthy individuals and patients within the flankers task. Group differences were also observed for the context in which errors were elicited on each task, with the flankers task functioning similarly among both groups but the picture/word task functioning differently. Thus, the behavior of response styles and ERPs in the flankers task was consistent across groups, whereas picture/word response styles and ERPs showed notable inconsistencies. Although the picture/word task effectively elicited error-related neural activity among healthy individuals—indeed, the timing and scalp distribution of the ERN/Pe were highly similar to that on flankers task—differences in reliability, convergent validity, and external validity across populations indicate that the two tasks are not interchangeable. This could be due in part to the fact that the flankers task is a relatively pure paradigm for eliciting error-related brain activity, whereas the picture/word task can be used to efficiently elicit neural responses associated with action monitoring as well as semantic processing (Mathalon, Faustman, et al., 2002).

Although the current data indicate favorable psychometric properties of error-related neural activity in psychotic disorders—indeed, the flankers ERN achieved good reliability among patients with only 12 trials—the data here are not definitive. The current sample size is large relative to typical ERP studies in psychosis, yet more precise estimates require sample sizes that exceed 300 (Charter, 1999). This would likely necessitate the coordinated efforts of multiple research sites, which would also allow for an examination of the generalizability of these reliability estimates. Another important future direction will be to assess test–retest reliability of the ERN and Pe in psychotic disorders, examining the consistency of these neural responses over time and how each may relate to fluctuations in symptomatology and level of functioning across multiple assessments.

Our results for the ERN and Pe are promising, yet there is substantial room for improvement on these measures. Reliability of .90 or higher is considered optimal for a clinical assessment tool (Cicchetti, 1994; Nunnally & Bernstein, 1994), but was not achieved even with 20 trials. Reliability of .70 is sufficient for research applications, but it comes at the cost of reduced convergent and external correlations, which we observed as well. The reliability of any ERP component will be negatively affected by a host of quality control factors not systematically considered here, including electrode impedance, electrostatic shielding, and room temperature. Practices vary between laboratories and projects, so it is important to evaluate and report reliability in each study. Furthermore, tighter quality controls offer another approach to improving reliability besides increasing the number of trials.

Although our comparison of the flankers and picture/word tasks is a novel contribution to the study of impaired error monitoring in psychosis, we did not consider other tasks that have previously been used in this patient population, such as Stroop (Kopp & Rist, 1999) and go/no-go tasks (Bates et al., 2002). It would be informative for future work to examine the convergent validity of the ERN and Pe across a comprehensive battery of paradigms. A limitation inherent to each of these tasks is subject attrition, which will be driven in part by differences in difficulty and demands across tasks. Here, 70 patients (67.3%) provided usable data on both tasks, and attrition was somewhat higher on the flankers than on the picture/word task. One contributing factor may have been the blockwise performance feedback that participants received on the flankers task, which was designed to foster a balance between response speed and accuracy and to ensure a sufficient number of error trials for analysis. Among healthy individuals this feedback was effective, such that all participants made at least two errors. The tradeoff between response speed and accuracy may have functioned differently among patients: nine (8.7%) made fewer than two errors, indicating a response style favoring accuracy at the expense of speed. This was somewhat less of a concern on the picture/word task, with only four patients (3.8%) excluded for near-perfect performance. Attrition is even higher when fair reliability of ERPs is a selection criterion, with 69 patients (66.3%) having an adequate number of trials for the flankers ERN, 51 (49.0%) for the flankers Pe, 25 (24.0%) for the picture/word ERN, and 27 (26.0%) for the picture/word Pe. Participant attrition is a general concern for studies of action monitoring and many other neural processes, although it has not been addressed systematically. Consistent tracking of attrition and evaluation of contributing factors would help to improve clinical utility, including modifications to task instructions and difficulty level to minimize subject attrition.

To effectively apply the ERN and Pe as biomarkers of impaired action monitoring in psychosis—that is, to translate findings from within-subjects experiments to between-subjects psychopathology research and clinical applications—the psychometric properties of those measures ought to first be established in both healthy and psychiatric populations. The current results indicate that task differences can have profound effects on both the reliability and the validity of the ERN/Pe, especially with regard to their ability to track illness characteristics in psychotic disorders. As illustrated here, two speeded response time tasks may elicit superficially similar error-related neural activity that actually shows little overlap across individuals, a pitfall that may lead to inconsistent findings in the literature. Error-related ERPs on the flankers task in particular are reliable and potentially valid tools for quantifying action monitoring deficits in psychotic disorders. The psychometric properties of the ERN and Pe can be directly assessed using well-established methods, and this information will be vital as these biological measures are further developed and made suitable for translational research and clinical applications.

References

Alain, C., McNeely, H. E., He, Y., Christensen, B. K., & West, R. (2002). Neurophysiological evidence of error-monitoring deficits in patients

with schizophrenia. *Cerebral Cortex, 12,* 840–846. doi:10.1093/cercor/12.8.840

Allen, A. J., Griss, M. E., Folley, B. S., Hawkins, K. A., & Pearlson, G. D. (2009). Endophenotypes in schizophrenia: A selective review. *Schizophrenia Research, 109,* 24–37. doi:10.1016/j.schres.2009.01.016

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.

Andreasen, N. C. (1983a). *Scale for the Assessment of Negative Symptoms (SANS)*. Iowa City, IA: University of Iowa.

Andreasen, N. C. (1983b). *Scale for the Assessment of Positive Symptoms (SAPS)*. Iowa City, IA: University of Iowa.

Bates, A. T., Kiehl, K. A., Laurens, K. R., & Liddle, P. F. (2002). Error-related negativity and correct response negativity in schizophrenia. *Clinical Neurophysiology, 113,* 1454–1463. doi:S1388245702001542

Bates, A. T., Liddle, P. F., Kiehl, K. A., & Ngan, E. T. (2004). State dependent changes in error monitoring in schizophrenia. *Journal of Psychiatric Research, 38,* 347–356. doi:10.1016/j.jpsychires.2003.11.002

Bromet, E. J., Kotov, R., Fochtmann, L. J., Carlson, G. A., Tanenberg-Karant, M., Ruggero, C. J., & Chang, S. W. (2011). Diagnostic stability revisited: Shifts during the decade following first-admission for psychosis. *The American Journal of Psychiatry, 168,* 1186–1194. doi:10.1176/appi.ajp.2011.11010048

Bromet, E. J., Schwartz, J. E., Fennig, S., Geller, L., Jandorf, L., Kovasznay, B., . . . Rich, C. (1992). The epidemiology of psychosis: The Suffolk County Mental Health Project. *Schizophrenia Bulletin, 18,* 243–255. doi:10.1093/schbul/18.2.243

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105. doi:10.1037/h0046016

Chapman, L. J., & Chapman, J. P. (1973). *Disordered thought in schizophrenia*. East Norwalk, CT: Appleton-Century-Crofts.

Charles, E. P. (2005). The correction for attenuation due to measurement error: Clarifying concepts and creating confidence sets. *Psychological Methods, 10,* 206–226. doi:10.1037/1082-989X.10.2.206

Charter, R. A. (1999). Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *Journal of Clinical and Experimental Neuropsychology, 21,* 559–566. doi:10.1076/jcen.21.4.559.889

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6,* 284–290. doi:10.1037/1040-3590.6.4.284

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7,* 309–319. doi:10.1037/1040-3590.7.3.309

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

Cronbach, L. J., & Warrington, W. G. (1951). Time-limit tests: Estimating their reliability and degree of speeding. *Psychometrika, 16,* 167–188. doi:10.1007/BF02289113

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics, 16,* 143–149. doi:10.3758/BF03203267

Falkenstein, M., Hoormann, J., Christ, S., & Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biological Psychology, 51,* 87–107.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured Clinical Interview for DSM–IV–TR Axis I disorders, research version, non-patient edition (SCID-I/NP)*. New York, NY: New York State Psychiatric Institute, Biometrics Research.

Foti, D., Kotov, R., Bromet, E. J., & Hajcak, G. (2012). Beyond the broken error-related negativity: Functional and diagnostic correlates of error processing in psychosis. *Biological Psychiatry, 71,* 864–872. doi:10.1016/j.biopsych.2012.01.007

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

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

Hajcak, G., McDonald, N., & Simons, R. F. (2003). To err is autonomic: Error-related brain potentials, ANS activity, and post-error compensatory behavior. *Psychophysiology, 40,* 895–903. doi:10.1111/1469-8986.00107

Herrmann, M. J., Römmler, J., Ehlis, A. C., Heidrich, A., & Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Brain Research. Cognitive Brain Research, 20,* 294–299. doi:10.1016/j.cogbrainres.2004.02.013

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., . . . Wang, P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. *The American Journal of Psychiatry, 167,* 748–751. doi:10.1176/appi.ajp.2010.09091379

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality* (2nd ed., pp. 102–138). New York, NY: Guilford Press.

Kim, M. S., Kang, S. S., Shin, K. S., Yoo, S. Y., Kim, Y. Y., & Kwon, J. S. (2006). Neuropsychological correlates of error negativity and positivity in schizophrenia patients. *Psychiatry and Clinical Neurosciences, 60,* 303–311. doi:10.1111/j.1440-1819.2006.01506.x

Kopp, B., & Rist, F. (1999). An event-related brain potential substrate of disturbed response monitoring in paranoid schizophrenic patients. *Journal of Abnormal Psychology, 108,* 337–346. doi:10.1037/0021-843X.108.2.337

Kotov, R., Guey, L. T., Bromet, E. J., & Schwartz, J. E. (2010). Smoking in schizophrenia: Diagnostic specificity, symptom correlates, and illness severity. *Schizophrenia Bulletin, 36,* 173–181. doi:10.1093/schbul/sbn066

Luck, S. J., Mathalon, D. H., O'Donnell, B. F., Hamalainen, M. S., Spencer, K. M., Javitt, D. C., & Uhlhaas, P. J. (2011). A roadmap for the development and validation of event-related potential biomarkers in schizophrenia research. *Biological Psychiatry, 70,* 28–34. doi:10.1016/j.biopsych.2010.09.021

Mathalon, D. H., Faustman, W. O., & Ford, J. M. (2002). N400 and automatic semantic processing abnormalities in patients with schizophrenia. *Archives of General Psychiatry, 59,* 641–648. doi:10.1001/archpsyc.59.7.641

Mathalon, D. H., Fedor, M., Faustman, W. O., Gray, M., Askari, N., & Ford, J. M. (2002). Response-monitoring dysfunction in schizophrenia: An event-related brain potential study. *Journal of Abnormal Psychology, 111,* 22–41. doi:10.1037/0021-843X.111.1.22

Minzenberg, M. J., Laird, A. R., Thelen, S., Carter, C. S., & Glahn, D. C. (2009). Meta-analysis of 41 functional neuroimaging studies of executive function in schizophrenia. *Archives of General Psychiatry, 66,* 811–822. doi:10.1001/archgenpsychiatry.2009.91

Morris, S. E., Yee, C. M., & Nuechterlein, K. H. (2006). Electrophysiological analysis of error monitoring in schizophrenia. *Journal of Abnormal Psychology, 115,* 239–250. doi:10.1037/0021-843X.115.2.239

Nieuwenhuis, S., Ridderinkhof, K. R., Blom, J., Band, G. P., & Kok, A. (2001). Error-related brain potentials are differentially related to awareness of response errors: Evidence from an antisaccade task. *Psychophysiology, 38,* 752–760. doi:10.1111/1469-8986.3850752

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

Olvet, D. M., & Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: Toward an endophenotype. *Clinical Psychology Review, 28,* 1343–1354. doi:10.1016/j.cpr.2008.07.003

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

Olvet, D. M., & Hajcak, G. (2009b). The stability of error-related brain activity with increasing trials. *Psychophysiology, 46,* 957–961. doi:10.1111/j.1469-8986.2009.00848.x

Overall, J. E., & Woodward, J. A. (1975). Unreliability of difference scores: A paradox for measurement of change. *Psychological Bulletin, 82,* 85–86. doi:10.1037/h0076158

Perez, V. B., Ford, J. M., Roach, B. J., Woods, S. W., McGlashan, T. H., Srihari, V. H., . . . Mathalon, D. H. (2012). Error monitoring dysfunction across the illness course of schizophrenia. *Journal of Abnormal Psychology, 121,* 372–387. doi:10.1037/a0025487

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. doi:10.1111/j.1469-8986.2010.00974.x

Ridderinkhof, K. R., van den Wildenberg, W. P., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition, 56,* 129–140. doi:10.1016/j.bandc.2004.09.016

Schmidt, F. L., Le, H., & Ilies, R. (2003). Beyond alpha: An empirical examination of the effects of different sources of measurement error on reliability estimates for measures of individual differences constructs. *Psychological Methods, 8,* 206–224. doi:10.1037/1082-989X.8.2.206

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. doi:10.1111/j.1469-8986.2009.00942.x

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. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 343–363). New York, NY: Oxford University Press.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215. doi:10.1037/0278-7393.6.2.174

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.

Taylor, S. F., Stern, E. R., & Gehring, W. J. (2007). Neural systems for error monitoring: Recent findings and theoretical perspectives. *Neuroscientist, 13,* 160–172. doi:10.1177/1073858406298184

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

Appendix

Bivariate Correlations Among Clinical Characteristics

Variables	Diagnosis	Negative symptoms	Psychotic symptoms	Disorganized symptoms	Rehospitalization, 0–4 years	Rehospitalization, 5–10 years	Unemployed	Neuroticism
Diagnosis	—							
Negative symptoms	−.53***	—						
Psychotic symptoms	−.22*	.25*	—					
Disorganized symptoms	−.19	.29**	.29**	—				
Rehospitalization, 0–4 years	−.03	.05	.08	.14	—			
Rehospitalization, 5–10 years	−.15	.28*	.23*	.06	.36***	—		
Unemployed	−.35***	.36**	.12	.26**	.31**	.39***	—	
Neuroticism	−.00	.18	.04	−.02	−.01	.16	.10	—

Note. Diagnosis indicates the comparison between schizophrenia and other psychoses.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Received March 28, 2012
Revision received February 25, 2013
Accepted March 6, 2013 ■

Correction to Foti, Kotov, and Hajcak (2013)

In the article "Psychometric considerations in using error-related brain activity as a biomarker in psychotic disorders" by Dan Foti, Roman Kotov, and Greg Hajcak (*Journal of Abnormal Psychology*, Vol. 122, No. 2, pp. 520–531. doi:10.1037/a0032618), the URL for supplemental material was missing.

Supplemental material for this article is available at: http://dx.doi.org/10.1037/a0032618.supp. The online version of this article has been corrected.

DOI: 10.1037/a0033666