2010

Temporal stability of the error-related negativity (ERN)and post-error positivity (Pe): The role of number of trials

Scott A. Baldwin
Brigham Young University - Provo

Michael J. Larson

Daniel A. Good

Joseph E. Fair

Psychophysiology, 47 (2010), 1167–1171. Wiley Periodicals, Inc. Printed in the USA.

DOI: 10.1111/j.1469-8986.2010.01022.x

BRIEF REPORT

Temporal stability of the error-related negativity (ERN) and post-error positivity (Pe): The role of number of trials

MICHAEL J. LARSON,[a,b] SCOTT A. BALDWIN,[a] DANIEL A. GOOD,[a] AND JOSEPH E. FAIR[a]
[a]Department of Psychology, Brigham Young University, Provo, Utah, USA
[b]Neuroscience Center, Brigham Young University, Provo, Utah, USA

Abstract

The error-related negativity (ERN) and post-error positivity (Pe) components of the event-related potential (ERP) are relatively stable over time. The current study further assessed the temporal reliability of ERN and Pe amplitudes for random samples of 2 to 14 trials per participant and the grand mean over a 2-week retest interval. In a replication of previous results, intraclass and zero-order correlations revealed moderate to good temporal stability for participants' ($N = 20$) grand mean ERN and Pe component amplitudes. Adding trials increased test–retest reliabilities; however, the temporal stability of ERN and Pe amplitudes with 14 or fewer trials were modest at best and considerably lower than that for the grand means. Overall, data support the temporal stability of grand-mean ERN and Pe amplitudes and suggest that more than 14 trials are needed to include in ERN and Pe averages for adequate test–retest reliability.

Descriptors: Error negativity (Ne), Anterior cingulate, Reliability, Test-retest, Temporal stability

The reliability of the error-related negativity (ERN) and post-error positivity (Pe) components of the scalp-recorded event-related potential (ERP) is currently the subject of considerable investigation. The ERN is a fronto-central negative-going deflection in the response-locked ERP that is larger following errors than correct trials and peaks within 100 ms after response (Falkenstein, Hohnsbein, Hoormann, & Banke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). Evidence suggests that the ERN reflects the activity of a performance- and action-monitoring system when there is a mismatch between intended and produced responses or when competing response options are simultaneously activated (Falkenstein et al., 1991; Gehring et al., 1993; Holroyd & Coles, 2002; Yeung, Cohen, & Botvinick, 2004).

The Pe is a positive deflection in the ERP that occurs between 100 and 400 ms following participant response and is more positive following error trials than correct trials (Falkenstein et al., 1991; Overbeek, Nieuwenhuis, & Ridderinkhof, 2005). Current theories suggest that the Pe is associated with signaling for post-error adjustments in behavior and the conscious recognition of errors, as Pe amplitudes are decreased when individuals are unaware of performance errors or neurologic deficits (Hajcak, McDonald, & Simons, 2003; Larson & Perlstein, 2009; Nieuwenhuis, Ridderinkhof, Blom, Band, & Kok, 2001).

Recent research indicates moderate to good reliability coefficients for the ERN and Pe components. For example, Segalowitz et al. (2010) observed adequate temporal stability of the ERN over short (20 min) and long (3 to 6 weeks) time periods. Olvet and Hajcak (2009a) report high internal consistency (split-half reliability) and test–retest reliability for the ERN and its correct-trial counterpart, the correct-trial negativity (CRN). Test–retest reliability estimates in this study were similar for individuals who committed a low number of errors ($M = 20$) and a high number of errors ($M = 36$). Olvet and Hajcak (2009b) also demonstrated moderate to high levels of internal consistency with as few as six and eight error trials for ERN and Pe component amplitudes, respectively. These findings were subsequently replicated across the life span (Pontifex et al., 2010).

Although six to eight trials may produce adequate internal consistency for the ERN and Pe, other forms of reliability, such as test–retest reliability, may require increased numbers of trials due to more possible sources of error variation (Kaplan & Saccuzzo, 2008). Thus, the primary purpose of this study was to examine the test–retest reliability of the ERN and Pe components with increasing numbers of error trials. We also sought to replicate previous findings of good test–retest reliability of the ERN and Pe components across a 2-week interval.

Method

Participants

Twenty-eight individuals between the ages of 19 and 29 initially enrolled in the study. Seven participants were excluded because they committed fewer than 14 errors (see Olvet & Hajcak, 2009b)

Address correspondence to: Michael J. Larson, Department of Psychology, Brigham Young University, 244 TLRB, Provo, UT 84602, USA. E-mail: michael_larson@byu.edu

PSYCHOPHYSIOLOGY

and 1 participant failed to return for the retest session. Thus, final enrollment included 20 healthy, right-handed individuals (eight female), with a mean (±*SD*) age of 22.35 (2.48) years. Exclusion criteria included history of psychiatric disorder, psychoactive medication use, substance abuse or dependence, neurological disorders, or uncorrected visual impairment. Participants were compensated for study participation. Study procedures were approved by the Institutional Review Board at Brigham Young University.

Experimental Task

Participants performed a modified version of the Eriksen Flanker Task (Eriksen & Eriksen, 1974) wherein congruent (e.g., <<<<<) and incongruent (e.g., <<><<) arrow stimuli were presented centered on a 17-in. computer monitor ~20 in. from the participant's head. Participants were instructed to respond as quickly and accurately as possible with an index-finger button press if the middle arrow pointed to the left and a middle-finger button press if the middle arrow pointed to the right. Flanker stimuli were presented for 100 ms prior to the onset of the target stimulus, which remained on the screen for 600 ms. To decrease expectancy effects, the intertrial interval (ITI) varied randomly between 800 and 1200 ms, with a mean ITI of 1000 ms. Three blocks of 300 trials (900 total trials) were presented; the distribution of congruent and incongruent trials was equal (450 trials each). Following task completion, a second session was scheduled for approximately 2 weeks later (average of 14.55 [±1.88] days between sessions).

Electrophysiological Data Recording and Reduction

Electroencephalogram (EEG) was recorded from 128 scalp sites using a geodesic sensor net and Electrical Geodesics, Inc. (EGI; Eugene, OR) amplifier system (20K nominal gain, band-pass = 0.10–100 Hz). During recording, EEG was referenced to the vertex electrode and digitized continuously at 250 Hz with a 24-bit analog-to-digital converter. Impedances were maintained below 50 kΩ. Data were average-re-referenced off-line and digitally low-pass filtered at 30 Hz. Eye movement and blink artifacts were corrected using the algorithm described by Gratton, Coles, and Donchin (1983).

Following Olvet and Hajcak (2009b), individual-subject response-locked averages were derived spanning 400 ms prior to response and 800 ms following response. Epochs were baseline corrected from −400 to −200 ms. Error-trial amplitudes for the ERN were extracted as the average activity from 0 to 100 ms at electrode site FCz. Latency measurements for the ERN were indexed at FCz as the peak negative-going amplitude within the 0–100-ms window. Amplitudes for the Pe were extracted as the average activity from 200 to 400 ms postresponse at electrode Pz. Given the tonic nature of the Pe, no latency times were calculated.

Statistical Analysis

Median response times (RT), mean error rates, and ERP component amplitude and latency data were analyzed using repeated measures analysis of variance (ANOVA) including the factors congruency (congruent, incongruent) and time (Time 1, Time 2) for RT and error-rate data and the factors accuracy (correct, error) and time for ERP data. Paired-samples *t* tests were used to decompose significant main effects and interactions. For reliability analyses we followed the procedures outlined by Lew, Gray, and Poole (2007) and Olvet and Hajcak (2009a, 2009b); test–retest reliability both of the grand mean and as a function of increasing number of trials was assessed using the single measure intraclass correlation (ICC) with a one-way random-effects model and zero-order correlations. With two time points, the ICC can range from −1.0 to 1.0. Acceptable values of the ICC vary with different authors. Anastasi (1998) indicated that values of the ICC at or above .60 are adequately reliable, whereas others indicated that ICCs < .40 are poor, ICCs between .41 and .59 are moderate, ICCs between .60 and .74 are good, and ICCs above .75 are excellent (Cicchetti, 2001; Cicchetti & Sparrow, 1981). For zero-order correlations, values above .50 are generally considered reliable for experimental research based on groups (Helmstadter, 1964; Segalowitz et al., 2010).

To explore the effect of the number of error trials on the reliability results, we compared the temporal reliability of error-trial ERN amplitude, error-trial ERN latency, and error-trial Pe amplitude separately for random samples of 2 to 14 trials per participant (i.e., we sampled two trials per participant, then three trials per participant, and so on). We used 14 as the maximum number of trials in order to provide a direct comparison with Olvet and Hajcak (2009b). To reduce the impact of sampling error, we replicated the random draw 2,500 times for each number of error trials and computed the mean reliability across the 2,500 replications.

Results

Response Times and Error Rates

Response time (in milliseconds) for congruent and incongruent trials at Time 1 were 363.20 (25.64) and 431.23 (26.87), respectively; RT data for congruent and incongruent trials at Time 2 were 348.48 (25.35) and 406.75 (28.47). A Congruency × Time ANOVA on RTs revealed the expected main effect of congruency, $F(1,19) = 507.12$, $p < .001$, $\eta^2 = .96$, with significantly longer RTs to incongruent relative to congruent trials at both Time 1, $t(19) = 23.11$, $p < .001$, and Time 2, $t(19) = 18.99$, $p < .001$. A significant main effect of time, $F(1,19) = 38.75$, $p < .001$, $\eta^2 = .67$, showed the effect of practice on performance, with decreased (i.e., faster) RTs from Time 1 to Time 2. There was also a significant Congruency × Time interaction, $F(1,19) = 20.22$, $p < .001$, $\eta^2 = .52$, with faster RTs from Time 1 to Time 2 for both incongruent, $t(19) = 7.79$, $p < .001$, and congruent trials, $t(19) = 4.19$, $p < .001$.

Error rates (percent errors) for congruent and incongruent trials at Time 1 were .03 (.02) and .12 (.06), respectively, and .02 (.03) and .09 (.05) for congruent and incongruent trials at Time 2. Analyses of error rates revealed a significant main effect of congruency, $F(1,19) = 49.62$, $p < .001$, $\eta^2 = .72$, indicating that participants made more errors to incongruent than congruent trials at both Time 1, $t(19) = 6.29$, $p < .001$, and Time 2, $t(19) = 6.60$, $p < .001$. The main effect of time and the Congruency × Time interactions were not statistically reliable, Fs < 2.9, ps > .11, indicating no overall differences in error rates from Time 1 to Time 2.

Event-Related Potential Data

Response-locked correct-trial and error-trial waveforms at Time 1 contained an average (±*SD*) of 728.35 (58.62) trials and 52.45 (29.85) trials, respectively. For Time 2, there was an average of 724.50 (102.23) correct trials and 42.85 (24.42) error trials contained in the averages. The number of error trials in the grand means ranged from 14 to 120. An Accuracy × Time ANOVA on the number of trials included showed no significant main

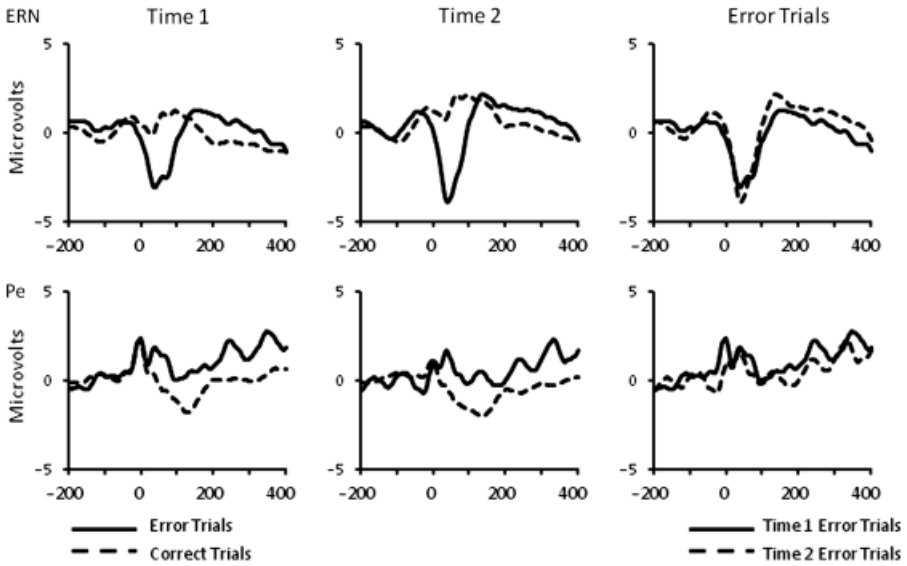

Figure 1. Grand mean ERP waveforms depicting response-locked correct- and error-related activity for the ERN at electrode FCz and the Pe at electrode Pz at Time 1 and Time 2 as well as a direct comparison of Time 1 and Time 2 error trials.

effect of time, $F(1,19) = 0.37$, $p > .55$, $\eta^2 = .02$, as well as no significant Accuracy × Time interaction, $F(1,19) = 0.06$, $p > .81$, $\eta^2 = .003$, indicating the number of trials in the grand means did not differ between time points. Average ERP waveforms for correct and error trials reflecting the ERN and Pe at Time 1 and Time 2 are shown in Figure 1. Supplemental analyses showed no relationship between number of error trials and ERP amplitude and latency data.

Amplitude measures. An Accuracy × Time ANOVA on response-locked ERPs for the ERN yielded a significant main effect of accuracy, $F(1,19) = 48.04$, $p < .001$, $\eta^2 = .72$, with a significant error relative to correct ERN at both Time 1, $t(19) = 5.21$, $p < .001$, and Time 2, $t(19) = 7.27$, $p < .001$. The main effect of time was not significant, $F(1,19) = 2.76$, $p > .12$, $\eta^2 = .13$, indicating that the overall magnitude of the ERP amplitudes did not significantly differ between time points. The Accuracy × Time interaction was also not significant, $F(1,19) = 3.06$, $p > .09$, $\eta^2 = .14$. Paired-samples t tests showed no significant differences between sessions for ERN amplitudes, $t(19) = -.24$, $p > .81$, but a significant difference between sessions for correct-trial amplitudes, $t(19) = -3.66$, $p < .002$. Results of an Accuracy × Time ANOVA on ERN latencies revealed no significant main effects or interactions, $Fs < .85$, $ps > .36$, indicating there were no differences in ERP latency between correct and error trials and across time points.

An Accuracy × Time ANOVA on Pe amplitudes was similar to that for the ERN. There was a main effect of accuracy, $F(1,19) = 7.62$, $p < .01$, $\eta^2 = .29$, with increased amplitude Pe for error trials relative to correct trials at Time 1, $t(19) = 2.65$, $p < .01$, and Time 2, $t(19) = 2.20$, $p < .04$. There was not a significant main effect of time or a significant Accuracy × Time interaction, $Fs < 1.23$, $ps > .28$, indicating the amplitude of the Pe generally did not differ between time points.

Temporal stability of the grand means. Most important to the current study is the temporal stability of error-related ERP components across time. Analyses indicated statistically reliable temporal stability for grand mean ERN amplitudes, ICC = .66, $p < .009$, and CRN amplitudes, ICC = .75, $p < .001$. Zero-order correlations on the grand means between time points supported these results, with significant correlations for the ERN, $r = .49$, $p < .03$, and the CRN, $r = .72$, $p < .001$.[1]

Grand mean latencies for the ERN showed low test–retest reliability for error trials, ICC = .33, $p > .18$, but good reliability for correct trials, ICC = .63, $p < .02$. Zero-order correlations were not significant between Time 1 and Time 2 for ERN latencies, $r = .25$, $p > .29$; the correlation for CRN latency was significant, $r = .46$, $p < .04$.

Analysis of the temporal stability of Pe amplitudes showed moderate retest reliability for error trials, ICC = .48, $p > .08$, and good reliability for correct trials, ICC = .68, $p < .006$. Zero-order correlations supported the results of the intraclass correlations, with modest reliability of the error-trial Pe amplitude across time points, $r = .32$, $p > .17$, but adequate reliability for correct trials, $r = .59$, $p < .007$.

Temporal stability with increasing error trials. Mean intraclass correlations and zero-order correlations with 80% confidence intervals across the 2,500 replications for increasing error trials are presented in Figure 2. The results for intraclass correlations and zero-order correlations for ERN amplitude, ERN latency, and Pe amplitude are similar. Not surprisingly, reliability increased with increasing number of error trials up to 14, but never reached the reliability of the grand mean. The benefit of each additional trial is small but does not level off within the scope of

[1]We recalculated the grand mean ICCs for fronto-central and centro-parietal regions of interest (ROIs) based on the scalp distributions of the current data. Amplitudes of the ERN were averaged across seven fronto-central electrode sites (5, 6, 7, 12, 13, 106, and 112; see Larson, Fair, Good, & Baldwin, 2010, for montage) and seven centro-parietal electrode sites for the Pe (6, 7, 13, 31, 80, 106, and 112). Grand mean amplitudes were more stable over time for the ROIs for the ERN, ICC = .77, $p < .001$, and CRN, ICC = .82, $p < .001$, relative to amplitudes at FCz for the ERN, ICC = .66, $p < .009$, and CRN, ICC = .75, $p < .001$. Similarly, ROI grand mean amplitudes were more stable for both error-trial Pe, ICC = .79, $p < .001$, and correct-trial Pe amplitudes, ICC = .91, $p < .001$, relative to amplitudes for the Pe at site Pz for error trials, ICC = .48, $p < .08$, and correct trials, ICC = .68, $p < .006$.

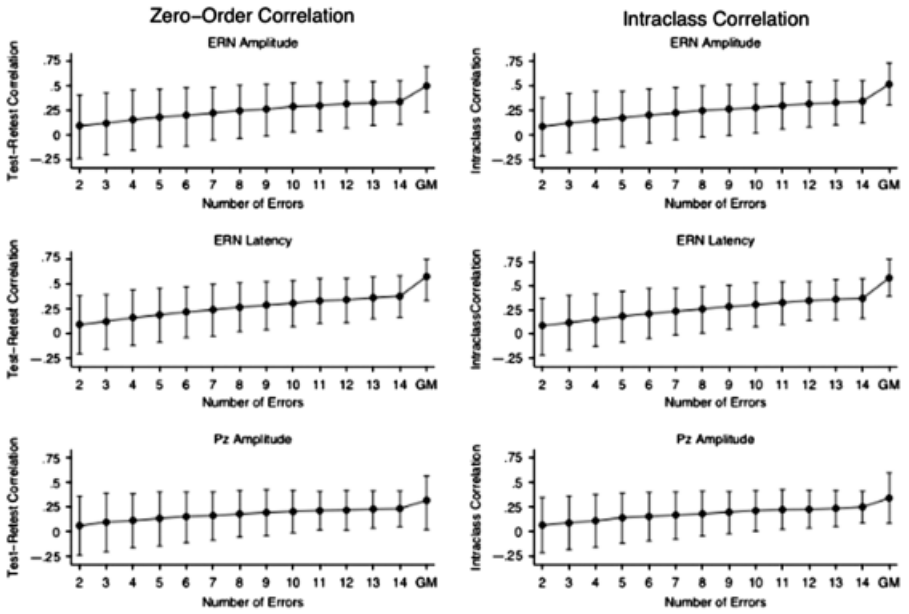

Figure 2. Average intraclass and zero-order correlations and 80% confidence intervals following 2,500 random samples for the ERN at electrode FCz and Pe at electrode Pz with increasing number of errors and the grand mean (GM).

our sampling scheme. Consequently, we suspect that reliability would continue to increase if we were able to increase the number of error trials beyond 14, although there would eventually be diminishing returns. Importantly, reliability levels for error-trial ERN amplitude, ERN latency, and Pe amplitude do not reach acceptable levels until all error trials are included in the grand means.

Discussion

Test–retest reliability for the grand means of both ERN and Pe amplitudes were reliable across time points despite considerable practice effects on the behavioral data leading to faster RTs and a trend toward decreased error rates over time. Changes in behavioral performance are likely due to the effect of practice on behavioral performance; however, error-trial amplitudes for the ERN and Pe did not significantly differ between Time 1 and Time 2, indicating that there was not an attenuating effect of practice on error-related amplitudes across multiple sessions (cf. Schrijvers et al., 2009). Correct-trial ERP amplitudes were more reliable than error-trial ERP amplitudes; this is likely due to the greater number of trials included in all correct-trial averages. We note, however, that overall CRN amplitudes differed between sessions, whereas those for the ERN did not. Measurements of ERN latency were not as reliable as those for ERN amplitudes; however, there was little difference in ERN latency values between sessions. Variability in ERN latency is consistent with previous research and may be due to inconsistency in participant response times and component processing times (see Olvet & Hajcak, 2009a).

We also examined the test–retest reliability of the ERN and Pe components of the ERP with increasing numbers of trials. Our findings indicate that the temporal stability of ERN and Pe amplitudes improved with each error trial included in the averages, but that adequate levels of temporal stability were not reached with up to 14 errors. In contrast, analysis of the grand means for ERN and Pe amplitudes showed moderate to good temporal stability over a 2-week test–retest interval. These results indicate that more than 14 trials are needed to achieve adequate levels of test–retest reliability, but that grand means with an average of 42 or more error trials are temporally stable.

Current results replicate previous findings of moderate to good temporal stability for grand mean ERN and Pe component amplitudes over time (Olvet & Hajcak, 2009a; Segalowitz et al., 2010). Our results also augment previous findings that six to eight error trials are enough to achieve adequate internal consistency for ERN and Pe amplitudes (Olvet & Hajcak, 2009b; Pontifex et al., 2010). That is, current data, using the mean of 2,500 samples for measures of test–retest reliability to reduce the effect of sampling error, show that more than 14 error trials are required for ERN and Pe amplitude averages to achieve adequate test–retest reliability. These data do not directly specify the number of trials necessary to achieve reliable temporal stability for these components; however, previous findings indicate moderate to good test–retest reliability for ERN and Pe amplitudes in a sample of individuals that made an average of 20 errors (Olvet & Hajcak, 2009a).

Some degree of caution should be taken in interpretation and comparison of findings regarding numbers of trials needed for reliability. Studies differ in EEG acquisition characteristics, task characteristics, and the nature of the sample. For example, data for the current study were collected using a high-impedance EGI system, whereas a previous study examining numbers of trials used BioSemi active electrodes (Olvet & Hajcak, 2009b). Similarly, participants may be more or less motivated to complete the task depending on levels of compensation or examiner characteristics. Thus, whereas temporal stability of the ERN is shown to be good in multiple studies, data should be interpreted within the context of specific laboratory and study procedures.

Test–retest amplitude and latency results are consistent with test–retest reliabilities reported in several studies of different ERP components in several different modalities. For example, Lew et al. (2007) showed ICCs ranging from .60 to .80 for the

amplitudes of the N1, mismatch negativity, P3, and N4 components of the auditory ERP in healthy individuals. Similarly, test–retest reliabilities are generally in the moderate to good ranges for neuropsychological measures, such as the Rey Auditory Verbal Learning Test, the Controlled Oral Word Association Test, and the Rey-Osterrieth Complex Figure Task, administered to healthy individuals (Strauss, Sherman, & Spreen, 2006). Thus, the temporal stability of electrophysiological measures of error processing is generally consistent with commonly used physiological and cognitive outcome measures.

Our results have at least four important implications for future research using ERN and Pe amplitudes. First, and most importantly, researchers should ensure that an adequate number of trials are included in grand means of the ERN to achieve adequate temporal stability. Second, the good temporal stability of ERN and Pe amplitudes allows for the possibility of a physiological measure of change that could provide insight into the neural mechanisms underlying treatment-related changes. Third, current treatment-related studies of the ERN show a wide range of changes pre- to posttreatment, with no changes in a study of pediatric obsessive-compulsive disorder (Hajcak, Franklin, Foa, & Simons, 2008), some indications for a link with symptom reduction in a study of individuals with depression (Schrijvers et al., 2009), and a clear relationship with 6 weeks of antipsychotic treatment in individuals with schizophrenia (Bates, Liddle, Kiehl, & Ngan, 2004). Findings of adequate temporal stability of the ERN indicate that the variation in treatment-related findings is not due solely to error and suggests the need for future studies. Fourth, findings may help elucidate the role of the ERN as an endophenotype for psychopathology by allowing confidence in multiple measures of ERN amplitude to determine if state-related changes are present in individuals with psychopathology (see Olvet & Hajcak, 2008), although the aforementioned variability in state-related ERN amplitudes associated with treatment indicate a need for considerable research in this regard.

REFERENCES

Anastasi, A. (1998). *Psychological testing* (6th ed). New York: Macmillan.

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.

Cicchetti, D. V. (2001). The precision of reliability and validity estimates re-visited: Distinguishing between clinical and statistical significance of sample size requirements. *Journal of Clinical and Experimental Neuropsychology*, *23*, 695–700.

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*, 127–137.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non-search task. *Perception & Psychophysics*, *16*, 143–149.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Banke, 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.

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.

Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography and Clinical Neurophysiology*, *55*, 468–484.

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.

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.

Helmstadter, G. C. (1964). *Principles of psychological measurement*. New York: Appleton-Century Crofts.

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.

Kaplan, R. M., & Saccuzzo, D. P. (2008). *Psychological testing: Principles, applications and issues* (7th ed). Belmont, CA: Wadsworth.

Larson, M. J., Fair, J. E., Good, D. A., & Baldwin, S. A. (2010). Empathy and error processing. *Psychophysiology*, *47*, 415–424.

Larson, M. J., & Perlstein, W. M. (2009). Awareness of deficits and error processing after traumatic brain injury. *NeuroReport*, *20*, 1486–1490.

Lew, H. L., Gray, M., & Poole, J. H. (2007). Temporal stability of auditory event-related potentials in healthy individuals and patients with traumatic brain injury. *Journal of Clinical Neurophysiology*, *24*, 392–397.

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. *Journal of Psychophysiology*, *19*, 319–329.

Olvet, D. M., & Hajcak, G. (2008). The error-related negativity (ERN) and psychopathology: Toward an endophenotype. *Clinical Psychology Review*, *28*, 1342–1354.

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.

Overbeek, T. J. M., Nieuwenhuis, S., & Ridderinkhof, K. R. (2005). Dissociable components of error processing: On the functional significance of the Pe vis-à-vis the ERN/Ne. *Journal of Psychophysiology*, *19*, 319–329.

Pontifex, M. B., Scudder, M. R., Brown, M. L., O'Leary, K. C., Wu, C. T., Themanson, J. R., et al. (2010). On the number of trials necessary for stabilization of error-related brain activity across the lifespan. *Psychophysiology* (in press).

Schrijvers, D., De Bruijn, E. R. A., Maas, Y. J., Vancoillie, P., Hulstijn, W., & Sabbe, B. G. C. (2009). Action monitoring and depressive symptom reduction in major depressive disorder. *International Journal of Psychophysiology*, *71*, 218–224.

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* (in press).

Strauss, S., Sherman, E. M. S., & Spreen, O. (Eds.). (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary* (3rd ed). New York: Oxford University Press.

Yeung, N., Cohen, J. D., & Botvinick, M. M. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, *111*, 931–959.

(Received May 15, 2009; Accepted November 25, 2009)