The Test-Retest Reliability of the Visually-Evoked Contingent Negative Variation (CNV) in Children and Adults

Brittany K. Taylor, **William J. Gavin**, and **Patricia L. Davies**
Colorado State University

Abstract

Establishing the reliability of event-related potentials (ERPs) is critical for future applications to biomarker development and clinical research. Few studies have examined the reliability of the contingent negative variation (CNV), and only in adults. The current study explored test-retest reliability of the visually-evoked CNV and its embedded components, the O-wave and the E-wave, in children (7-to-13 years) and young adults (19-to-28 years) during a visual Go-NoGo task over one-to-two weeks. Test-retest reliability of the components was moderate for children, and low-to-moderate for adults. These findings were in contrast to previous work with adults showing moderate-to-high reliability of the auditory-evoked CNV.

Researchers who use electroencephalography (EEG) to study brain activity during cognitive tasks often examine event-related potentials (ERPs) derived from multi-channel recordings. Surprisingly, despite the long history and ever-widening use of ERPs in the study of cognition and brain development, the psychometric properties of ERP measures have received little discussion in the literature. Researchers often are content to assume that if the outcome of a particular experiment conducted in their lab or reported in the literature produces statistically significant findings then the measures of interest must be valid and reliable. Although significant findings in a study lend support to the validity of the measures, generalizing the outcome to future research efforts may not be appropriate. The constructs of validity and reliability should be evaluated separately from the testing of the research questions, particularly when the stimulus parameters are modified such as switching from auditory to visual stimulus modalities, or changing the population studied such as young children instead of college-aged adults. Most importantly, merely establishing the validity of a measure does not necessarily establish the reliability of a measure.

In addition, recent research efforts are beginning to address the potential of ERP measures to serve as diagnostic as well as descriptive functions of brain processing of sensory stimuli or of cognitive processes specific to clinical populations; that is, to serve as biomarkers of a disorder. If such efforts are to be successful, even more rigor regarding reliability of brain activity measures will be required. Of the various types of reliability (temporal reliability, inter-examiner reliability, alternate forms reliability, split-half reliability, and intra-sample

reliability), temporal reliability (test-retest) is one of the more rigorous estimation procedures because it compares the results from the administration of the same assessment protocol in two independent testing sessions separated by a short time interval. The present study focuses on determining the temporal reliability of ERP measures derived from a visual Go-NoGo paradigm, a cognitive task commonly used to elicit the contingent negative variation (CNV). This type of task has been demonstrated to have developmental and, potentially, diagnostic associations indicative of sustained attention (Segalowitz & Davies, 2004; Walter, Winter, Cooper, McCallum, & Aldridge, 1964).

The CNV is an ERP measure that presents in Go trials as a slow, negative voltage drift during the latter period defined by two stimuli, a warning stimulus followed by an imperative stimulus signaling the initiation of a Go response. This negative drift found in Go trials is contrasted to an absence of a negative drift in a similar period of the NoGo trials. Most researchers measure the CNV component as a voltage averaged over a large portion of time (e.g., 800 ms) ending just prior to the onset of the imperative stimulus. In addition, studies have shown that the CNV has two embedded components: the O-wave, an early component, and the E-wave, a late component. The O-wave has been noted as an orienting response (Giard, Perrin, Pernier, & Bouchet, 1990; Rohrbaugh, Newlin, Varner, & Ellingson, 1984; Zimmer & Demmel, 2000), and the E-wave is hypothesized to represent expectancy and response preparation (Basile, Baldo, de Castro, & Gattaz, 2002; Bender, Resch, Weisbrod, & Oelkers-Ax, 2004; Knott et al., 1991). As a literature review on CNV paradigms shows, substantial experimental evidence exists for establishing both the face and concurrent validity of the resulting CNV measures in both adults and children (e.g., Filipovi , Jahanshahi, & Rothwell, 2001; Segalowitz & Davies, 2004; Smith, Johnstone, & Barry, 2007).

The generators of the CNV and its two embedded components have been localized mainly to frontal regions with studies showing the dorsolateral prefrontal cortex as the primary generator of the total CNV (Bareš, Rektor, Kaˇovský, & Streitová, 2000, 2003; Basile et al., 2003). Additional neural influences include the supplementary motor cortex, primary motor cortex, anterior cingulate cortex, basal ganglia, thalamus, orbitofrontal cortex, and even parietal areas of the brain for the O-wave and E-wave (Basile et al., 2002; Bender et al., 2004; Falkenstein, Hoormann, Hohnsbein, & Kleinsorge, 2003; Giard, Perrin, Pernier, & Bouchet, 1990; Knott et al., 1991; Zimmer & Demmel, 2000). Because the frontal lobes mature throughout childhood and adolescence, researchers have begun to explore the developmental trajectory of the CNV and its components (Hämmerer et al., 2010; Jonkman, 2006; Jonkman, Lansbergen, & Stauder, 2003; Segalowitz & Davies, 2004). Developmental research has found that children tend to have smaller, less negative CNV components compared to adults (Hämmerer et al., 2010; Jonkman, 2006; Jonkman et al., 2003). Additionally, data indicate that the CNV component amplitudes gradually become more negative throughout development into young adulthood (Segalowitz & Davies, 2004). Significant differences in amplitude have been found even between seven-year-old children and ten-year-old children (Jonkman et al., 2003).

A few known studies have examined the temporal reliability of the CNV and its embedded components, but only in adult samples, and each study measured the amplitude of the CNV

at different time periods within the component (Griesel & Bartel, 1975; Kropp, Kiewitt, Göbel, Vetter, & Gerber, 2000; Roth, Kopell, Tinklenberg, Hunstberger, & Kraemer, 1975). For example, Roth et al. (1975) had participants perform a visual-auditory Go-NoGo task, and the CNV was measured as the peak amplitude at three time points (600, 800 and 1000ms following the warning stimulus) for each of the two sessions scheduled seven days apart. The authors reported moderate reliability for each of the three measurements (r = .60 - .67). The O-wave and E-wave were not measured in this study. A later study by Kropp et al. (2000) had participants perform an auditory Go-NoGo paradigm. Unlike the previous study, Kropp et al. (2000) did evaluate all three CNV measures across two separate recording sessions scheduled ten days apart. The O-wave was measured as the averaged amplitude in a 200ms window surrounding each individual's maximum negative amplitude between 550 and 750ms. The E-wave was measured as the averaged amplitude within the last 200ms directly preceding the imperative stimulus and the total CNV was measured as the averaged amplitude within the total timeframe between the conditional and imperative stimuli (3000ms total). The authors reported moderate-to-high reliability for all three components (r = .63 - .86) where the O-wave was the most reliable, and the E-wave and total CNV were similarly less reliable than the O-wave.

Although the outcomes of these few reliability studies on adults are encouraging, to our knowledge, the temporal reliability of the CNV and its embedded components has not been studied in children. As developmental research continues to emerge, it becomes more and more clear that children tend to have more variability in their EEG and ERP tracings compared to adults. Specifically, children's ERPs tend to have a lower signal-to-noise ratio (SNR) compared to adults, in part due to the reduced number of segments included in the averaging process for children as a result of eye-blink and movement-related artifacts (Gavin & Davies, 2008). SNR decreases in a non-linear fashion as a function of the number segments included in the averaging process decreases (Junghöfer, Elbert, Tucker, & Rockstroh, 2000). In addition to effects that result from participants' cooperation during recording, studies have shown that children's ERP component amplitudes and latencies vary as a function of development as a result of differences in responsivity of neural generators, refractory rates of neural populations, and conductivity of brain and skull tissues though this is not an exhaustive list (for a complete review, see Kappenman & Luck, 2012). These factors highlight issues that potentially contribute to increased sources of between- and with-subject variances which may confound statistical testing to show the experimental effects of interest (see Gavin & Davies, 2008 for more discussion of these issues). Critical to this study, the possibility of increased measurement error is likely to influence reliability of a measure.

The purpose of the present study was to examine the temporal (i.e., test-retest) reliability of the CNV and its embedded components in children and adults. We first determined whether the three CNV measures exhibit face and construct validity by comparing averaged voltage measures for Go trials to NoGo trials in both adults and children across two testing sessions. We predicted that both groups would show significantly more negative amplitudes for Go trials compared to NoGo trials in the three CNV time periods. We also predicted that the children would have significantly smaller CNV measures than adults as described in the literature (Hämmerer et al., 2010; Jonkman, 2006; Jonkman et al., 2003; Segalowitz &

Davies, 2004). Additionally, in accordance with the literature, we predicted that among children, CNV amplitudes would become more negative as a function of increasing age (Segalowitz & Davies, 2004). Next, the test-retest reliability of the CNV data was determined. In accordance with previous reliability studies (Kropp et al., 2000), the O-wave was expected to be the most reliable component followed by the E-wave and finally the total CNV for both children and adults. We expected that adults would have significantly better reliability indices for all components compared to children. Due to the possibility that the CNV component may contain more measurement error especially for children, we also determined if the earlier stimulus-processing components, specifically the N1, P2, N2, and P3 to the conditional stimulus, were reliable across the two sessions. We expected reliability coefficients of the N1, P2, N2, and P3 amplitudes and latencies to be consistently high for both groups across the two sessions.

Methods

Participants

Participants for this study were 33 adults between the ages of 19 and 28 years, and 62 children between the ages of 7 and 13 years. All participants were screened for neurological or developmental disorders, use of psychopharmaceutical drugs (e.g., antidepressants), and history of head trauma via a self-report by adult participants and parent-report for child participants. One adult and three children were excluded from this study with a diagnosis of attention-deficit hyperactivity disorder. One child was excluded due to a reported speech delay. Seven children were excluded from analyses due to inadequate performance, as is described in detail in the results. The final sample consisted of 32 young adult college students ($M = 23.28$ years, $SD = 2.31$; 22 males) and 51 children ($M = 10.30$ years, $SD = 1.56$; 26 males). All participants had normal or corrected-to-normal vision. Before data collection procedures began, all adult participants and parents of child participants signed informed consent forms. All child participants signed assent forms. Participants were compensated with a small thank you gift choosing between a t-shirt or mug after completing their first session, and between a t-shirt, mug or $15 after the second session. All procedures were approved by the local university institutional review board.

Procedure

Participants completed two EEG recording sessions scheduled one or two weeks apart with the second session being at the same time and day of the week as the first session. Participants were seated in a comfortable chair at a table in front of a computer screen. Two research assistants placed the EEG cap and sensors on the participant. For each participant, the same EEG cap was used for each session, and measurements were performed to assure proper placement each time. Next a research assistant gave participants a brief training on how to reduce production of artifacts from eye blinks and muscle movements. Then, three minutes of eyes-opened resting EEG were recorded while participants stared at a fixation point on the screen. Participants then performed a total of four EEG paradigms lasting approximately one hour. Only the Go-NoGo paradigm used to elicit the CNV, which was the third paradigm, will be discussed in this study. Participants were given a short break of 2 to 4 minutes between each paradigm. Following EEG tasks, participants were given a ten-minute

break before completing approximately one hour of paper and pencil behavioral assessments examining cognitive and executive functioning. Short breaks of 2 to 4 minutes were taken between each assessment. These assessments will not be discussed further in this study. In total, each session lasted approximately two hours.

Visual Go-NoGo Paradigm

The design of the Go-NoGo paradigm used for this study was fairly simple so that participants of all ages could easily understand the instructions and perform the task correctly. All stimuli were presented on a computer screen positioned directly in front of the participant with an approximate 85cm viewing distance. Each trial began with the presentation of a conditional stimulus with participants seeing either a red circle or a green circle for a duration of 250ms. Then, 1750ms after the circle disappeared, the imperative stimulus, an image of a white car, appeared in the center of the screen for a duration of 250ms. Both the red and green circle were 7.5cm in diameter resulting in an approximate visual angle of 5.1°. The car was 5cm tall and 15cm long, thus the stimuli subtended 3.4° vertically and 10.0° horizontally. Participants were instructed to press a button on a response pad placed in front of them with their right index finger as quickly as possible after the car appeared only if a trial began with an image of a green circle, (i.e., a Go trial) and to refrain from pressing the button when the car appeared for trials that began with an image of a red circle (i.e., a NoGo trial). A random, variable inter-trial interval of 2000 to 7000ms was used to avoid eliciting a CNV response between trials.

Participants completed approximately 10 practice trials with feedback that applauded correct trials and offered instruction to correct behavior on incorrect trials. Participants were also shown their reaction times for each correct Go trial and encouraged to try and beat their best time. After the participant was comfortable with the practice and appeared to understand the instructions, the test phase of the paradigm was presented which lasted approximately nine minutes. There were a total of 40 Go trials and 40 NoGo trials presented in the same pseudorandom order during each of the two recording sessions. Performance feedback was not given during the test phase nor at the end of the test phase.

Electrophysiological Recording

EEG recordings were obtained using the BioSemi ActiveTwo system with an Active Two Lycra head cap (BioSemi, Inc., Amsterdam, The Netherlands). Active EEG was recorded from 32 Ag-AgCl sintered electrodes based on the American Electroencephalographic Society nomenclature guidelines (1994) with a Common Mode Sense (CMS) active electrode and a Driven Right Leg (DRL) passive electrode as reference and ground respectively (http://www.biosemi.com/faq/cms&drl.htm). An additional pin-type Ag/AgCl electrode was placed at FCz for a total of 33 scalp sites. Electrodes were placed on the left and right earlobes for off-line referencing, each of the two electrodes had a dedicated preamplifier. Two bipolar electrooculograms (EOG) were used to account for vertical and horizontal eye movements. The vertical EOG was derived from electrodes placed on the supra- and infraorbital regions of the left eye. The horizontal EOG was derived from two electrodes placed on the left and right outer canthi. Data were sampled at a rate of 1024Hz. Electrode offsets were maintained at ±20mV throughout each session.

Electrophysiological Data Reduction

Using BrainVision Analyzer 2.0 software (www.brainproducts.com), data from the continuous EEG recording were re-referenced to the averaged voltage of the two earlobe electrodes, filtered with a .03 to 30Hz bandpass filter (12dB/octave), and then segmented from 200ms prior to the conditional stimulus onset to 2250ms after the conditional stimulus onset. Baseline correction was performed on each segment using the EEG data from -200 – 0ms relative to the conditional stimulus onset. Segments were then divided into correct Go and NoGo trials. Correct Go trials were any trials in which a green circle appeared before the car, and the button was pressed after the car appeared. Correct NoGo trials were any trials in which a red circle appeared before the car, and there was no button press. Trials in which a button was pressed before the onset of the imperative stimulus (i.e., the car) or in which a response was made during a NoGo trial were considered incorrect trials and were not included in the averaging process. A regression procedure used to remove eye blinks was applied to retained segments (Segalowitz, 1996). Following the regression procedure, segments were baseline corrected again using the -200 to 0ms window and then underwent an artifact rejection procedure to remove segments with voltages exceeding ±100μV.

Averaged ERPs for Go and NoGo segments retained after data reduction were calculated for each participant. The data were processed using a Matlab routine which allows for automatic scoring and visual inspection of ERP components, and, when necessary, allows for manual marking of components. All ERP component measurements were carried out by one trained research assistant. Baseline-to-peak amplitudes and latencies for the N1, P2, N2, and P3, and averaged amplitudes in the time windows of the three CNV components were measured using the Matlab routine. All components were measured at site Cz. The N1 was defined as the most negative amplitude in the 70-170ms window, P2 as the most positive peak in the 150-300ms window for both age groups. However, given that N2 and P3 components tended to occur later in children than in adults, separate windows were used for each group. For adults, N2 was defined as the most negative amplitude in the 170-300ms window, and the P3 as the most positive peak in the 250-500ms window. For children, the N2 window was 250-400ms and the P3 was 400-650ms. Peak-to-peak amplitudes were calculated by subtracting the obtained baseline-to-peak component amplitude from the component just prior (e.g., the peak-to-peak P3 amplitude was calculated as P3 amplitude - N2 amplitude). Averaged amplitudes were determined for the 600-800ms window for the O-wave, the 1800-2000ms window for the E-wave, and the 1000-2000ms window for the total CNV. These windows were selected based on prior research (e.g., Kropp et al., 2000). Visual inspection of the grand averages showed that children had a large P300 overlapping the time window of the O-wave, as has been noted in previous literature (Jonkman et al., 2003; Jonkman, 2006), thus the O-wave was not analyzed for children.

Data Analysis

Behavioral data—Behavioral performance on the Go-NoGo task was evaluated based on two dependent measures, number of trials with correct button-press responses, and reaction times of button presses. To determine whether children and adults had a significantly different number of trials with correct responses in the Go-NoGo task and, therefore, leading to a different number of segments included in the averaging process, a 2 (Group: Adults,

Children) × 2 (Session: First, Second) × 2 (Condition: Go versus NoGo) ANOVA (1 between and 2 within factor) was performed. All alpha levels for significance were set to .05. Post-hoc t-tests employing Tukey's studentized range were used to further investigate significant main effects for all ANOVA designs (Kirk, 2012).

Reaction times for correct Go trials were calculated as the time in milliseconds from onset of the imperative stimulus to the button press. A 2 (Group) x 2 (Session) ANOVA was performed to determine whether children and adults had similar reaction times across sessions. Reliability of reaction times was assessed using Pearson correlations as well as two forms of ICC: ICC (3,1) consistency, and ICC (3,1) absolute agreement. ICC analyses are commonly used to assess reliability due to their ability to account for both inter- and intra-individual variability in measurements (e.g., Thesen & Murphy, 2002).

CNV component validity—ANOVAs were performed to determine whether the average amplitude in the time windows of the O-wave, E-wave, and total CNV were significantly different between Go and NoGo trials. Specifically, a 2 (Session) × 2 (Condition) ANOVA was performed with O-wave measurements in adults, and two separate 2 (Session) × 2 (Condition) × 2 (Group) ANOVAs were performed with E-wave and total CNV measurements. Findings from the E-wave and total CNV 2×2×2 ANOVAs were also used to determine whether children had significantly less negative amplitudes compared to adults. A Bonferroni correction was applied resulting in a test-wise error rate of $\alpha = .017$ for the interpretation of ANOVA findings. In some cases (noted in the Results section), the assumption of equality of covariances was violated. However, ANOVA designs are robust to this violation (Cohen, 2008). A second analysis, a Pearson correlation, was used to determine whether CNV amplitudes increased with age within children.

CNV component reliability—Test-retest reliability across the two sessions for the CNV components was assessed using Pearson correlations, ICC (3,1) consistency, and ICC (3,1) absolute agreement. Analyses were only performed for Go trial amplitudes. Reliability of the O-wave was only assessed for adults. In order to address whether the reliability of the CNV and its components differed between children and adults, a series of Fisher's r to z transformations were performed (Cohen, 2008). Specifically, z scores were calculated by subtracting children's z scores from adults' z scores (i.e., $z_{final} = [z_{adult} - z_{children}]/\sigma$).

Conditional stimulus-evoked ERP component reliability—Test-retest reliability across the two sessions for the ERP components evoked by the conditional stimulus (N1, P2, N2, and P3) was assessed using Pearson correlations, ICC (3,1) consistency, and ICC (3,1) absolute agreement. Analyses were performed for only Go trial measurements of baseline-to-peak amplitudes, peak-to-peak amplitudes, and peak latencies.

Results

Behavioral Data

After artifact removal procedures seven children were excluded from the study because they had fewer than 12 correctly performed Go segments retained for the averaging process (< 30% of all Go trials) for either session one or session two. Thus, the final sample included in

analyses consisted of 32 adults and 51 children. In order to examine whether the remaining children had a significantly different task accuracy compared to adults, a 2 (Group) × 2 (Session) × 2 (Condition) ANOVA was performed. Box's test of equality of covariance matrices was violated. On average, children appeared to have similar performance (Go trials: $M = 33.88$, $SD = 4.38$; NoGo trials: $M = 39.01$, $SD = 1.53$) compared to adults (Go trials: $M = 33.40$, $SD = 4.81$; NoGo trials: $M = 39.69$, $SD = .61$). There were no significant differences in the number of correctly-performed trials between groups, regardless of the condition (Go versus NoGo) or the session, $F(1, 81) = .05$, $p = .83$, $\eta^2_p = .001$, indicating similar task accuracy for children and adults.

An analysis was performed in order to determine whether children and adults significantly differed in reaction times to correctly-performed Go trials across sessions. The means and standard deviations of reactions times for Go trials are reported in Table 3 Results of a 2 (Group) × 2 (Session) repeated measures ANOVA showed that adults significantly differed from children in reaction times, $F(1, 81) = 29.86$, $p < .001$, $\eta^2_p = .27$. Equality of covariances was violated for this test. Post hoc analysis showed that adults had significantly faster reaction times when compared to children, $t(82) = -7.94$, $p < .01$. However, there were no significant differences in reaction times across sessions for either group.

Reliability of reaction times was assessed using three different reliability measures: Pearson correlations, ICC (3,1) consistency, and ICC (3,1) absolute agreement. The coefficients of reactions times from session 1 to session 2 ranged from .70 to 74 for children and .64 to .61 for adults groups (see Table 3), suggesting relatively stable performance across sessions for both groups.

CNV Component Validity

The averaged amplitudes in the time windows of the O-wave (600-800ms), the E-wave (1800-2000ms), and the total CNV (1000-2000ms) were examined via ANOVA designs in order to 1) establish the face and construct validity of the CNV components (i.e., Go amplitudes should be more negative than NoGo amplitudes), and 2) to test the hypothesis that adults would have more negative amplitudes than children. A 2 (Condition) × 2 (Session) repeated measures ANOVA was used to examine the O-wave in adults, and two separate 2 (Group) × 2 (Condition) × 2 (Session) ANOVAs were performed to examine the E-wave and the total CNV. The means and standard deviations for the averaged amplitudes of the CNV components are reported in Table 1. ERP tracings can be seen in Figures 1 and 2.

The difference in the average amplitudes of the O-wave for the Go and NoGo trials in adults was not statistically significant given the corrected α level, $F(1, 31) = 4.31$, $p = .046$, $\eta^2_p = .12$. However, the differences between the Go and NoGo amplitudes were significant for both the E-wave, $F(1, 81) = 77.66$, $p < .0005$, $\eta^2_p = .49$, and the total CNV, $F(1, 81) = 56.36$, $p < .0005$, $\eta^2_p = .41$. Post hoc tests indicated that for both components, amplitudes on Go trials were significantly larger (i.e., more negative) than amplitudes on NoGo trials across children and adults (E-wave: $t(82) = -9.05$, $p < .01$; total CNV: $t(82) = -7.12$, $p < .01$).

With respect to developmental trends, the data supported the hypothesis that children would have significantly smaller CNV measures than adults. Adults, in general, had greater negative E-wave amplitudes compared to children, $F(1, 81) = 6.78$, $p = .01$, $\eta^2_p = .08$, though this effect varied as a function of Trial Condition, as indicated by the significant Group × Condition interaction, $F(1, 81) = 7.76$, $p = .007$, $\eta^2_p = .09$. Post hoc tests showed that the difference between Go and NoGo trial amplitudes was greater for adults, $t(31) = -9.83$, $p < .01$, than it was for children, $t(50) = -6.45$, $p < .01$. With respect to the total CNV, the assumption of equality of covariances was violated. As was found for the E-wave, adults had significantly more negative total CNV amplitudes compared to children, $F(1, 81) = 6.23$, $p = .016$, $\eta^2_p = .07$. However, the Group × Session interaction was not statistically significant for the CNV component.

Further exploration of developmental trends supported the hypothesis that among children, CNV component amplitudes would become larger as a function of increasing age. Children's age was negatively correlated with the E-wave (session 1: $r(51) = -.28$, $p = .05$; session 2: $r(51) = -.24$, $p = .10$), and with the total CNV (session 1: $r(51) = -.31$, $p = .03$; session 2: $r(51) = -.20$, $p = .15$). Although not all were statistically significant, the correlations were in the expected direction indicating that older children had averaged amplitude values that were more negative than younger children in both sessions.

CNV Component Reliability

Temporal reliability of the CNV components was assessed using three reliability measures: Pearson correlations, ICC (3,1) consistency, and ICC (3,1) absolute agreement. In general, the analyses revealed low-to-moderate reliability indices for the CNV components (see Table 2). Adults showed the highest reliability coefficients for the O-wave and the E-wave, and the total CNV was the least reliable component. Only the O-wave reliability measurements for adults reached statistical significance, $r = .58$, $p = .001$. For children, the E-wave was the most reliable component followed by the total CNV. Unlike adults, children's reliability statistics were significant for the E-wave and total CNV, $r_{\text{E-Wave}} = .50$, $p < .0005$; $r_{\text{CNV}} = .34$, $p = .016$.

To determine whether the reliability coefficients of the E-wave and total CNV significantly differed between children and adults, two Fisher's r to Z transformations were performed. Though the children had higher reliability coefficients than adults, none of the Z scores calculated from the Pearson correlations were statistically significant, indicating that there were no significant differences in reliability between children and adults; E-wave: $Z = -1.52$, $p = .06$, total CNV: $Z = -1.28$, $p = .10$.

Reliability of ERP Components Elicited by the Conditional Stimulus

Temporal reliability was established for the N1, P2, N2, and P3 following the presentation of the conditional stimulus using measures of both baseline-to-peak and peak-to-peak amplitudes as well as peak latencies. Means and standard deviations for all amplitude and latency measures are reported in Table 3. For adults, in general, baseline-to-peak amplitude measurements showed better reliability than peak-to-peak amplitude measurements (see Table 3). Temporal reliability of adults' ERP amplitudes widely varied with the N1 peak-to-

peak amplitude showing the poorest stability, $r(32) = .28$, $p > .05$, and the P3 baseline-to-peak amplitude showing the strongest reliability, $r(32) = .76$, $p < .0005$. Findings regarding component latencies were also variable among adults with the P3 latency showing the weakest reliability, $r(32) = .30$, $p > .05$, and the N2 latency showing the strongest reliability, $r(32) = .64$, $p < .0005$.

Among children, temporal reliability of the earlier components seemed to be generally lower than observed in adults (see Table 3). Unlike adults, for all components except for the N1, peak-to-peak amplitude measures seemed to exhibit greater stability than baseline-to-peak amplitude measures. Considering amplitude stability among children, the peak-to-peak N2 amplitude showed the strongest reliability, $r(51) = .59$, $p < .0005$, whereas the N1 peak-to-peak amplitude showed the poorest reliability, $r(51) = .21$, $p > .05$. Measures of component latency reliability were consistently low among children with Pearson correlations ranging from .21 to .31.

Discussion

Behavioral Data

Behavioral data analyses indicated that both children and adults tended to perform the Go-NoGo task similarly during both sessions. With respect to the number of correctly-performed trials, ANOVA results indicated no significant differences in task accuracy between children and adults across the two sessions. Additionally, although children's reaction times were significantly longer and more variable than adults, both children's and adults' reaction times were reliable across the two sessions. These data suggest that for both age groups, regardless of any differences in measures of brain processing over time, task performance was consistent over the one-to-two week period, which is expected.

Reliability of ERP Components Elicited by the Conditional Stimulus

Results for the adults' ERP components elicited by the conditional stimulus indicated moderate-to-high reliability across all measures, thus supporting the original hypothesis. The findings agreed with prior studies showing moderate-to-high reliability for the N1 and P2 amplitudes and latencies in auditory and olfactory paradigms in adult samples (Rentzsch et al., 2008; Thesen & Murphy, 2002). However, among children, reliability indices tended to be lower than in adults and ranged from low-to-moderate reliability for the N1 and P2 amplitude and latency measures. To the best of our knowledge, this is the first study to date to report reliability indices for the N1 and P2 components in a sample of children. The closest comparison in the literature is Segalowitz and Barnes' (1993) investigation of the two-year test-retest reliability of N1 and P2 components in a sample of 15-year-olds during an auditory oddball task. The researchers showed low reliability for amplitudes, and moderate reliability for latencies. However, data collected from children in the present study showed moderate reliability for amplitudes, and low reliability for latencies (see Table 3).

One must proceed with caution when comparing the results of the present study to previous literature on the reliability of the N1 and P2 components. Prior literature has, to the best of our knowledge, exclusively examined the reliability of components evoked to sensory

modalities other than visual. In addition, no known studies to date have reported reliability of the ERP components to a conditional stimulus for children. Characteristics of ERP components including amplitude and latency are known to vary as a function of both sensory modality (Kappenman & Luck, 2012) and maturation (Segalowitz & Davies, 2004). Thus, prior studies on the test-retest reliability of the N1, P2, N2, and P3 components may not appropriately compare to the present study.

Overall, the results indicated that adults, in general, showed fairly consistent earlier stimulus processing across sessions. In contrast, children tended to vary in the way that they processed the conditional stimulus information during the task across the two sessions. Previous research suggests that neural processing measured via EEG can change as a result of practice, at least in adult samples. For example, Romero and colleagues (2008) showed that after multiple practice sessions of a complex math task, adults exhibited smaller P300 amplitudes as a result of the paradigm becoming less cognitively taxing and more automatic. Additionally, participants showed differential oscillatory power in the theta and beta bands between practice and test conditions. Other researchers have noted shifts in alpha band power as a result of decreased cortical activation after practice (e.g., Smith, McEvoy, & Gevins, 1999). To the best of our knowledge, no studies to date have examined shifts in cognitive strategies in children indicated by EEG measures over time. Shifts in strategy may be indicated differently in children compared to adults due to differences in neural maturation.

The differences in ERP reliability between the groups in the present study provide evidence that children may have altered their neural processing strategies to a greater extent than adults did over the one-to-two week period despite reliable task performance measures for both groups. In order to further explore this phenomenon, future research should inspect ERP component measures as well as oscillatory power measures across multiple practice sessions in both children and adults.

CNV Component Development

E-wave and total CNV amplitudes were more negative for Go trials than for NoGo trials for both children and adults, thus verifying that the CNV components were present during Go trials, but not in NoGo trials. Though the analysis of the O-wave for adults revealed a p-value of .046, it did not show a significant difference in amplitudes between Go and NoGo trials given the adjusted alpha level. Future investigations should further explore the validity of the visually-evoked O-wave measurement.

The data supported the hypothesis that children would exhibit significantly smaller E-wave and total CNV amplitudes when compared to adults. These data agree with previous developmental literature examining the CNV (Hämmerer et al., 2010; Jonkman, 2006; Jonkman et al., 2003; Segalowitz & Davies, 2004). Additionally, analyses examining developmental trends within children supported the hypothesis that age would be negatively correlated with amplitude, though the correlations were weak. Prior investigations have shown strong relationships between age and CNV component amplitude in children and adolescents (Bender et al., 2002). For instance, Bender and colleagues (2002) examined developmental trends in the auditory-evoked CNV components in a sample of 76 typically-

developing children and adolescents age 6-to-18 years. Regression analyses indicated strong, statistically significant negative relationships between age and the CNV component amplitudes, thus indicating that amplitudes become more negative through the course of development (Bender et al., 2002). It is possible that the developmental trends were weaker in the present investigation due to the more limited age range of children examined (7-to-13 years). Greater variability in age may be required to clearly illustrate the developmental change in the CNV components.

CNV Component Reliability

We hypothesized that for adults, the O-wave would be the most reliable component followed by the E-wave and finally the total CNV. The data supported our hypotheses, though reliability indices were unexpectedly low, especially for adults. Among adults, the O-wave was moderately reliable, and was the only significantly-reliable component. The poor reliability indices for the adult data were surprising given the previous literature on CNV reliability. For example, Kropp et al. (2000) measured the auditory-evoked O-wave, E-wave, and total CNV in 27 adult participants over two recording sessions set ten days apart. Pearson correlations revealed reliability ranging between .63 and .86 across the three components (Kropp et al., 2000). Likewise, previous studies that only measured the total CNV in adults found reliability indices between .60 and .80 (Griesel & Bartel, 1975; Roth et al., 1975). The discrepancy between the current findings and previous literature may be due to differences in the sensory modality of the paradigm presented to the participants. The effects of sensory modality on test-retest reliability warrant further investigation in future studies.

For children, because the O-wave could not be measured, the E-wave was found to be the most reliable component followed by the total CNV. For children, the components reached low-to-moderate reliability, and indices for both the E-wave and total CNV were statistically significant. In contrast to the hypothesis that adults would exhibit better stability in CNV components than children, these data indicated that children seemed to have better reliability indices than adults. However, further inspection via Fisher's r to Z transformations showed no statistically significant differences between children and adults' test-retest reliability. The findings indicate that there is significant variability in cognitive processing across sessions for both children and adults leading to poor reliability in the CNV components. The findings are interesting given that behavioral performance is consistent across sessions in both groups; the neural mechanism underlying stimulus processing and task-attention vary, though the outcome performance is stable over time. These data suggest possible changes in task strategy between sessions in both children and adults. Future investigations should further refine the study methodology in order to more clearly delineate task-strategy differences across sessions.

Limitations

The current investigation employed an entirely visual Go-NoGo paradigm in contrast to prior CNV reliability investigations, making it difficult to directly compare the current findings to prior studies. It is unclear whether these differences in reliability indices are the result of the task modality. In order to better determine the effects of sensory modality on

ERP component reliability, studies should be designed to directly compare visually-evoked ERPs to those evoked by other sensory modalities. Additionally, the current study only investigated 7-to-13 year-old children. It is possible that reliability may vary in children and adolescents of different ages, thus the current findings may not be generalizable to children outside of the 7-to-13 year range. Finally, the task instructions in the present study required participants to respond as quickly as possible during Go trials of the Go-NoGo task. Recent research has shown that emphasizing speed versus accuracy during a Go-NoGo task can lead to differential ERP measures in adult samples (Aasen & Brunner, 2016). To further understand whether the task instructions in the present study may have differentially affected children and adults, future investigations should directly assess the impact of emphasizing speed versus accuracy on children's EEG data during a Go-NoGo task.

Conclusions

Although behavioral performance remained stable across a one-to-two week period, the findings suggest differential neural processing, reflected in CNV components, over a several week timeframe for both children and adults. Among adults, the O-wave was moderately reliable, but neither the E-wave nor the total CNV was reliable. Children tended to have better reliability for the E-wave and total CNV when compared to adults, though there were no statistically significant differences between groups. The poor stability results of the CNV components were unexpected, especially when compared to the stability of the earlier ERP components (i.e., N2, P2 and P3) within the same task, which were at least moderately reliable.

Acknowledgments

References

Aasen IE, Brunner JF. 2016; Modulation of ERP components by task instructions in a cued go/no-go task. Psychophysiology. 53:171–185. DOI: 10.1111/psyp.12563 [PubMed: 26488615]

Bareš M, Rektor I, Kanovsky P, Streitová H. 2000; Cortical and subcortical distribution of cognitive operations: A contingent negative variation depth electrode study. Homeostasis. 40:91–93.

Bareš M, Rektor I, Ka ovský P, Streitová H. 2003; Cortical and subcortical distribution of middle and long latency auditory and visual evoked potentials in a cognitive (CNV) paradigm. Clinical Neurophysiology. 114:2447–2460. DOI: 10.1016/s1388-2457(03)00250-5 [PubMed: 14652105]

Basile LFH, Ballester G, de Castro CC, Gattaz WF. 2002; Multifocal slow potential generation revealed by high-resolution EEG and current density reconstruction. International Journal of Psychophysiology. 45:227–240. DOI: 10.1016/s0167-8760(02)00014-4 [PubMed: 12208529]

Bender S, Resch F, Weisbrod M, Oelkers-Ax R. 2004; Specific task anticipation versus unspecific orienting reaction during early contingent negative variation. Clinical Neurophysiology. 115:1836–1845. DOI: 10.1016/j.clinph.2004.03.023 [PubMed: 15261862]

Cohen, BH. Explaining Psychological Statistics. 3rd. Hoboken, NJ: John Wiley & Sons, Inc; 2008.

Falkenstein M, Hoormann J, Hohnsbein J, Kleinsorge T. 2003; Short-term mobilization of processing resources is revealed in the event-related potential. Psychophysiology. 40:914–923. DOI: 10.1111/1469-8986.00109 [PubMed: 14986844]

Filipovi S, Jahanshahi M, Rothwell J. 2001; Uncoupling of contingent negative variation and alpha band event-related desynchronization in a Go/No-Go task. Clinical Neurophysiology. 112:1307–1315. DOI: 10.1016/S1388-2457(01)00558-2 [PubMed: 11516743]

Gavin, WJ, Davies, PL. Developmental Psychophysiology: Theory, Systems, and Methods. Cambridge University Press; New York, NY: 2008. Obtaining reliable psychophysiological data with child participants; 424–447.

Giard MH, Perrin F, Pernier J, Bouchet P. 1990; Brain generators implicated in the processing of auditory stimulus deviance: A topographic event-related potential study. Psychophysiology. 27:627–640. DOI: 10.1111/j.1469-8986.1990.tb03184.x [PubMed: 2100348]

Griesel RD, Bartel PR. 1975; An assessment of the reliability of measures of the slow cerebral electrical potentials related to conditional expectancy (CNV). Psychologia Africana. 16:1–6.

Hämmerer D, Li SC, Müuuml;ller V, Lindenberger U. 2010; An electrophysiological study of response conflict processing across the lifespan: Assessing the roles of conflict monitoring, cue utilization, response anticipation, and response suppression. Neuropsychologia. 48:3305–3316. DOI: 10.1016/j.neuropsychologia.2010.07.014 [PubMed: 20638396]

Jonkman LM. 2006; The development of preparation, conflict monitoring and inhibition from early childhood to young adulthood; A Go/Nogo ERP study. Brain Research. 1097:181–193. DOI: 10.1016/j.brainres.2006.04.064 [PubMed: 16729977]

Jonkman LM, Lansbergen M, Stauder JEA. 2003; Developmental differences in behavioral and event-related brain responses associated with response preparation and inhibition in a go/nogo task. Psychophysiology. 40:752–761. DOI: 10.1111/1469-8986.00075 [PubMed: 14696728]

Junghöfer M, Elbert T, Tucker DM, Rockstroh B. 2000; Statistical control of artifacts in dense array EEG/MEG studies. Psychophysiology. 37:523–532. [PubMed: 10934911]

Kappenman, ES, Luck, SJ. The Oxford handbook of event-related potential components. New York: Oxford University Press; 2012.

Kirk, RE. Experimental Design: Procedures for the Behavioral Sciences. SAGE Publications; 2012.

Knott VJ, Lapierre YD, De Lugt D, Griffiths L, Bakish D, Browne M, Horn E. 1991; Preparatory brain potentials in major depressive disorder. Progress in Neuro-Psychopharmacology & Biological Psychiatry. 15:257–262. DOI: 10.1016/0278-5846(91)90089-J [PubMed: 1871326]

Kropp P, Kiewitt A, Göbel H, Vetter P, Gerber WD. 2000; Reliability and stability of contingent negative variation. Applied Psychophysiology and Biofeedback. 25:33–41. DOI: 10.1023/a:1009533405695 [PubMed: 10832508]

Rohrbaugh JW, Newlin DB, Varner JL, Ellingson RJ. 1984; Bilateral distribution of the O wave. Annals of the New York Academy of Sciences. 425:267–270. DOI: 10.1111/j.1749-6632.1984.tb23545.x [PubMed: 6588843]

Romero SG, McFarland DJ, Faust R, Farrell L, Cacace AT. 2008; Electrophysiological markers of skill-related neuroplasticity. Biological Psychology. 78:221–230. DOI: 10.1016/j.biopsycho.2008.03.014 [PubMed: 18455861]

Roth W, Kopell B, Tinklenberg J, Huntsberger G, Kraemer H. 1975; Reliability of the contingent negative variation and the auditory evoked potential. Electroencephalography and Clinical Neurophysiology. 38:45–50. [PubMed: 45903]

Segalowitz SJ, Barnes KL. 1993; The reliability of ERP components in the auditory oddball paradigm. Psychophysiology. 30:451–459. DOI: 10.1111/j.1469-8986.1993.tb02068.x [PubMed: 8416071]

Segalowitz SJ, Davies PL. 2004; Charting the maturation of the frontal lobe: An electrophysiological strategy. Brain and Cognition. 55:116–133. DOI: 10.1016/s0278-2626(03)00283-5 [PubMed: 15134847]

Smith ME, McEvoy LK, Gevins A. 1999; Neurophysiological indices of strategy development and skill acquisition. Cognitive Brain Research. 7:389–404. DOI: 10.1016/S0926-6410(98)00043-3 [PubMed: 9838200]

Smith JL, Johnstone SJ, Barry RJ. 2007; Response priming in the Go/NoGo task: The N2 reflects neither inhibition nor conflict. Clinical Neurophysiology. 118:343–355. DOI: 10.1016/j.clinph.2006.09.027 [PubMed: 17140848]

Thesen T, Murphy C. 2002; Reliability analysis of event-related brain potentials to olfactory stimuli. Psychophysiology. 39:733–738. DOI: 10.1111/1469-8986.3960733 [PubMed: 12462501]

Vuillier L, Whitebread D, Szucs D. 2015; ERP evidence of cognitive strategy change in motivational conditions with varying level of difficulty. Neuropsychologia. 70:126–133. DOI: 10.1016/j.neuropsychologia.2015.02.025 [PubMed: 25708173]

Walter WG, Winter AL, Cooper R, McCallum WC, Aldridge VJ. 1964; Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. Nature. 203:380–384. DOI: 10.1038/203380a0 [PubMed: 14197376]

Zimmer H, Demmel R. 2000; Habituation and laterality of orientating processes as reflected by slow negative waves. Biological Psychology. 53:161–176. DOI: 10.1016/S0301-0511(00)00048-X [PubMed: 10967231]
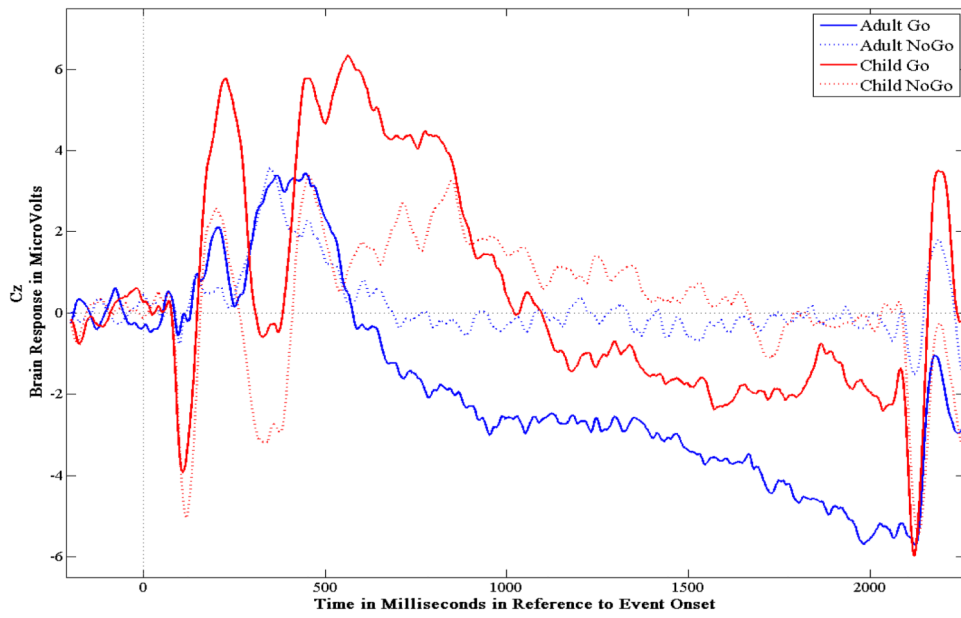
Figure 1.
The grand averaged ERP of children and adults' Go and NoGo trials for session one measured at site Cz.
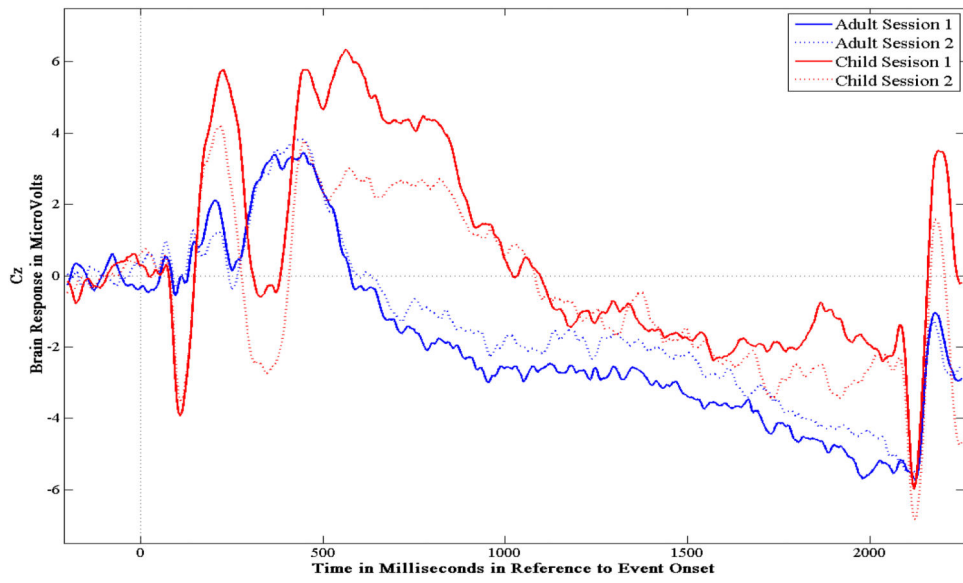
Figure 2.
The grand averaged ERPs for children and adult's Go trials for sessions one and two measured at site Cz.

Table 1

The Means and Standard Deviations of the Averaged Amplitudes (in µV) in the Time Windows of the O-wave, E-wave, and Total CNV for Go and NoGo Trials in Session One and Session Two

		Go Trials		NoGo Trials	
		Session 1	Session 2	Session 1	Session 2
		M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Children	600-800ms (O-Wave)	—	—	—	—
	1800-2000ms (E-Wave)	-1.47 (4.20)	-2.91 (4.23)	0.24 (3.89)	-.20 (2.65)
	1000-2000ms (total CNV)	-1.35 (4.03)	-1.74 (3.85)	.74 (3.80)	.39 (3.09)
Adults	600-800ms (O-Wave)	-1.21 (3.06)	-.66 (3.14)	.11 (1.79)	-.12 (2.33)
	1800-2000ms (E-Wave)	-4.94 (3.04)	-4.24 (3.42)	-.12 (2.31)	-.55 (2.37)
	1000-2000ms (total CNV)	-3.53 (2.56)	-2.74 (2.83)	-.18 (1.95)	-.43 (2.35)

Note: The O-wave was not measured for children due to the overlapping P300 component in the 600-800ms time window.

Table 2

Test-Retest Reliability Indices: Pearson and Intraclass Correlation Coefficients for the CNV Component Amplitudes (μV) in Adults and Children from Session One to Session Two

		r	ICC (3,1) Consistency	ICC (3,1) Absolute Agreement
Children	O-wave	—	—	—
	E-wave	.50 ***	.50 ***	.48 ***
	Total CNV	.34 *	.33 **	.33 **
Adults	O-wave	.58 **	.58 ***	.58 ***
	E-wave	.19	.19	.19
	Total CNV	.05	.05	.05

*
p < .05

**
p < .01

***
p < .001

Note: Reliability for the O-wave was not measured for children due to the overlapping P300 component in the 600-800ms time window.

Table 3

Means, Standard Deviations, and Reliability Indices of Reaction Times (ms) and ERP Component (N1, P2, N2, and P3) Amplitudes (μV) and Latencies (ms) for Go Trials in Session 1 and Session 2

	Session 1	Session 2	Reliability		
	M (SD)	M (SD)	r	ICC (3,1) Consistency	ICC (3,1) Absolute Agreement
Children					
Reaction Time	301.87 (99.71)	312.32 (132.13)	.79***	.79***	.74***
N1 Baseline-to-Peak Amplitude	-5.50 (4.26)	-5.45 (3.97)	.51***	.51***	.51***
N1 Peak-to-Peak Amplitude	-7.68 (4.04)	-7.81 (3.47)	.24	.24*	.24*
N1 Latency	116.56 (20.19)	117.70 (22.30)	.21	.21	.21
P2 Baseline-to-Peak Amplitude	8.67 (3.74)	7.61 (5.22)	.41**	.39**	.38**
P2 Peak-to-Peak Amplitude	14.09 (4.78)	13.06 (5.31)	.54***	.53***	.53***
P2 Latency	222.94 (34.82)	211.47 (35.68)	.31*	.31*	.29*
N2 Baseline-to-Peak Amplitude	-4.18 (6.03)	-9.17 (6.24)	.53***	.53***	.52***
N2 Peak-to-Peak Amplitude	-12.85 (5.47)	-13.44 (5.48)	.59***	.59***	.59***
N2 Latency	336.89 (48.44)	330.00 (44.74)	.22	.21	.22
P3 Baseline-to-Peak Amplitude	10.90 (5.82)	8.17 (4.79)	.49***	.48***	.43***
P3 Peak-to-Peak Amplitude	15.09 (6.00)	14.00 (5.48)	.52***	.52***	.52***
P3 Latency	526.27 (87.98)	519.91 (94.14)	.29*	.29*	.29*
Adults					
Reaction Time	209.11 (43.83)	191.55 (45.27)	.64***	.61***	.62***
N1 Baseline-to-Peak Amplitude	-1.85 (1.82)	-2.06 (2.41)	.43 ***	.41**	.42**
N1 Peak-to-Peak Amplitude	-3.48 (1.75)	-4.41 (2.54)	.28	.26	.25
N1 Latency	115.69 (27.60)	111.45 (31.55)	.52**	.51**	.52**
P2 Baseline-to-Peak Amplitude	4.17 (3.57)	3.62 (3.67)	.74***	.74***	.74***
P2 Peak-to-Peak Amplitude	6.03 (3.49)	5.68 (3.41)	.71***	.71***	.71***
P2 Latency	183.65 (32.60)	176.45 (40.89)	.61***	.59***	.59***
N2 Baseline-to-Peak Amplitude	-2.02 (3.62)	-4.28 (3.22)	.67***	.65***	.65***
N2 Peak-to-Peak Amplitude	-6.19 (3.34)	-6.03 (2.31)	.51**	.47**	.48**
N2 Latency	255.71 (39.58)	242.31 (40.49)	.64***	.64***	.62***
P3 Baseline-to-Peak Amplitude	6.48 (4.35)	6.96 (3.94)	.76***	.75***	.75***

	Session 1	Session 2		Reliability	
	M (SD)	*M (SD)*	*r*	ICC (3,1) Consistency	ICC (3,1) Absolute Agreement
P3 Peak-to-Peak Amplitude	8.50 (3.33)	9.33 (2.97)	.57**	.57***	.56***
P3 Latency	392.21 (63.37)	426.44 (73.85)	.30	.30*	.28*