Clinical Neurophysiology 116 (2005) 1665–1680
www.elsevier.com/locate/clinph

Assessment of working memory abilities using an event-related brain potential (ERP)-compatible digit span backward task☆

Celeste D. Lefebvre[a,*], Yannick Marchand[a,b], Gail A. Eskes[a,c], John F. Connolly[a,c,d,e,**]

[a]Department of Psychology, Dalhousie University, Life Sciences Centre, Halifax NS, Canada B3H 4J1
[b]Institute for Biodiagnostics (Atlantic), Neuroimaging Research Laboratory, National Research Council Canada, Halifax NS, Canada B3H 4J1
[c]Departments of Psychiatry and Psychology, Capital Health Queen Elizabeth II Health Sciences Centre, Halifax NS, Canada B3H 4J1
[d]Department of Pediatrics, IzaakWalton Killam (IWK) Hospital, Halifax NS, Canada B3H 4J1
[e]Divison of Neurology, Department of Medicine, Dalhousie University, Halifax NS, Canada B3H 4J1

Accepted 23 March 2005

Abstract

Objective: This study investigated the effectiveness of an ERP-compatible Digit Span Backward (ERP-DB) task to determine working memory abilities in healthy participants.

Methods: Participants were administered both the standard digit span backward and ERP-DB tasks. The ERP-DB task was divided into two sections, consisting of 2, 4, 6 and 8 (Group 1) and 3, 5, and 7 (Group 2) set sizes. A set of digits was aurally presented, followed by a second set that either corresponded to the reverse order of the first set (correct condition) or had one digit in the sequence replaced by an incorrect digit (incorrect condition).

Results: Two posterior positive components were found to distinguish the two conditions; an earlier positive component (P200/P300) was elicited in the correct condition, whereas a comparatively robust and prolonged positive slow wave (PSW) was elicited in the incorrect condition. Furthermore, the PSW and the difference in PSW amplitude between incorrect and correct conditions (dPSW) dissipated as working memory load increased and were related to working memory capacity.

Conclusions: The PSW, dPSW and P200/P300 components were found to be associated with working memory abilities and may have the potential to act as neurophysiological markers for the assessment of working memory capacity.

Significance: This research lends support for the utility of the ERP-DB task as a means of assessing working memory abilities, which may have implications for testing patients with expressive communication impairments.

Keywords: Digit span; Neuropsychology; Standardized neuropsychological tests; Clinical utility; Working memory capacity; Wechsler Adult Intelligence Scale-Third Edition (WAIS-III).

1. Introduction

After receiving a brain injury, an accurate and valid assessment of a patient's level of cognitive functioning is essential in order to formulate treatment and rehabilitation strategies (Lezak, 1995; Sohlberg and Mateer, 2001). However, most standard neuropsychological tests of cognitive functioning require verbal or behavioral responses from the patient. Unfortunately, following neurological trauma, patients frequently have speech and/or motor disabilities (Morse and Montgomery, 1992; Pedersen et al., 1995; Wade et al., 1986) and thus the ability to

☆ This study was conducted at the Cognitive/Clinical Research Unit, Life Sciences Centre, Department of Psychology, Dalhousie University, Halifax, Nova Scotia. Preliminary data from this study was presented at the 2002 International Organization of Psychophysiology 11th World Congress of Psychophysiology, Montreal, Canada.

* Corresponding author. Tel.: +1 902 494 2961; fax: +1 902 494 6585.
** Corresponding author. Address: Department of Psychology, Dalhousie University, Life Sciences Centre, Halifax NS, Canada B3H 4J1. Tel.: +1 902 494 2531; fax: +1 902 494 6585.
E-mail addresses: clefebvr@dal.ca (C.D. Lefebvre), john.connolly@dal.ca (J.F. Connolly).

1388-2457/$30.00 © 2005 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.clinph.2005.03.015

assess cognitive functioning with standard neuropsychological tests is difficult (or impossible) in patients with communication and/or behavioral impairments.

To bypass some of these assessment challenges, researchers advocated decades ago for computer-automated testing procedures to overcome the need for verbal responses (Miller, 1968; Thompson and Wilson, 1982). However, computerized paradigms of this type have generally required fine motor control (i.e. typing skill) and have had limited clinical applicability to date. Another, more recent approach to overcoming communication barriers is the use of a brain–computer interface (see Kübler et al., 2001 for review), where patients learn to emit specific electrophysiological responses that subsequently drive a communication device (e.g. a spelling device; Birbaumer et al., 1999). This method is still in its rudimentary stages, demands numerous training sessions and imposes a high cognitive load. Therefore, before implementing such a program, it would be sensible to demonstrate that the patient is capable of such a high level of cognitive functioning. Overall, there continues to be a necessity for a method for assessment of cognitive functions independent of verbal and behavioral responses (Connolly et al., 2000).

Event-related brain potentials (ERP) have the potential to significantly contribute to clinical neuropsychology by providing a neurophysiological index of patients' on-line cognitive functioning (Connolly and D'Arcy, 2000; Reinvang, 1999). Also, cognitive ERP paradigms have been modified or developed for application to neurotrauma populations (e.g. Allen et al., 1992; Ellwanger et al., 1996; Kotchoubey et al., 2001; Lang and Kotchoubey, 2002). One way to extend the clinical utility of cognitive ERP paradigms is to adapt standardized neuropsychological tests (see Connolly and D'Arcy, 2000; Connolly et al., 2000 for reviews). This allows one to (1) target the same cognitive processes assessed by the standard test, thereby reducing the inferences necessary to interpret the results; (2) have access to a large normative database for comparison purposes and; (3) assess various patients with the same test materials regardless of their communication impairments.

A research program initiated by Connolly and colleagues in the late 1990's has focused on adapting several standard neuropsychological tests for computer presentation and simultaneous ERP recordings in order to assess language functioning. These tasks include the Peabody Picture Vocabulary Test – Revised (PPVT; Dunn and Dunn, 1981), the Vocabulary and Similarities subtests of the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1991), the Wechsler Adult Intelligence Scale—Revised as a Neuropsychological Instrument (WAIS-R-NI; Kaplan et al., 1991), the Token Test (Boller and Vignolo, 1966; De Renzi and Vignolo, 1964), and the Psycholinguistic Assessments of Language Processing in Aphasia Test (PALPA; Kay et al., 1992).

In all cases, the ERP-adapted tests were designed so that they did not require verbal or behavioral responses (although button-press responses were employed in some studies in order to ensure that performance on the adapted tests was comparable to the original versions). This was achieved by aurally or visually presenting either correct or incorrect answers to the test items. Results demonstrated that different ERP patterns were elicited by correct and incorrect answers when the questions were within participants' ability range but not when the task demands exceeded their capabilities (as determined by their performance on the traditionally administered tests). In addition, performance on the ERP-adapted tests correlated strongly with the standard versions. Irrespective of the type of test, these patterns of results were found in healthy adults (Connolly et al., 1995, 1999a; D'Arcy and Connolly, 1999; D'Arcy et al., 2000), children (Byrne et al., 1995a, 1999) and stroke patients (D'Arcy et al., 2003) as well as in case studies of communication-impaired patients with profound dyslexia, traumatic brain injury (TBI) (Connolly et al., 1999b, 2000) and cerebral palsy (Byrne et al., 1995b). Moreover, using a newly developed statistical method, stroke patients' performance on the standard PPVT test was found to have a high correlation ($r=0.86$) with their ERP patterns (Marchand et al., 2002).

The next step in this research program was to develop a battery of ERP-adapted neuropsychological tests to assess other aspects of cognitive functioning beyond language abilities. In the same manner that a battery of standardized neuropsychological tests is used to highlight a patient's pattern of strengths and weaknesses to guide the development of an individualized rehabilitation program, it would be ideal to have a battery of ERP-adapted neuropsychological tests that are independent of language abilities for the same purpose.

Currently, there are a handful of ERP-adapted standardized tests that have been developed to assess other aspects of cognitive functioning. These include the adaptation of the Wisconsin Card Sorting Task (WCST), the Delayed Recall section of Verbal Paired Associates subtest from the Wechsler Memory Scale-Revised (WMS-R) and the Continuous Visual Memory Test, to assess, respectively, executive functioning (Barceló et al., 1997), recognition memory (Holamon et al., 1995) and figural memory (Retzlaff and Morris, 1996).

Working memory is another key aspect of cognitive functioning, which involves the temporary storage and effortful manipulation of information (Baddeley and Logie, 1999). Working memory is a core cognitive ability shown to be essential to and a strong predictor of learning, and intellectual and fluid reasoning abilities (de Jong and Das-Smaal, 1995; Fry and Hale, 2000; Kyllonen, 1987; Kyllonen and Christal, 1987, 1990; Sternberg, 1980; Woltz, 1988). Because working memory is vital for learning, a necessity for everyday functioning, and deficits are very prominent (Morse and Montgomery, 1992) and disruptive to patients following brain injury (Schwartz et al., 2003), an accurate assessment is critically important for developing individualized rehabilitation and treatment programs (Gioia and Isquith, 2004). In addition, the development of a method to evaluate working memory independent of communication abilities would be of benefit not only to non-communicative patients but also to patients with slow or delayed expressive language abilities because standardized neuropsychological tests of working memory require immediate verbal responses. Lastly, the use of an ERP-adapted measure may reveal differences in neurophysiological processing that may be complementary to, or of benefit over and above the use of standardized tests even in communicative patients.

Numerous studies have investigated the neurophysiological and theoretical aspects of working memory processing using ERP. An early description of the functional significance of the P300 proposed that it was associated with the updating of working memory. Subsequent work supported this proposal by demonstrating a correlation between P300 amplitude and subsequent recall or recognition of items when rote memory strategies are utilized (Fabiani et al., 1985, 1986; Johnson and Donchin, 1985). Further studies have found an increase of P300 amplitude and latency (and/or subsequent posterior positive slow waves (PSW) within 300–1000 ms post-stimulus) as memory load increased (García-Larrea and Cézanne-Bert, 1998; Kusak et al., 2000; Nittono et al., 1999); which has been interpreted as a reflection of the additional processing related to the number of items to be retrieved and manipulated. P300 amplitude was also found to correlate with working memory abilities as assessed by standardized neuropsychology tests (Howard and Polich, 1985; Nittono et al., 1999; Polich et al., 1983) and to increase as task demands increased as long as performance remained high (Johnson and Donchin, 1985; Nittono et al., 1999).

Additional links between the P300 (and/or aspects of the PSW) and various stages of working memory processing have been investigated using an ERP version of the 'Sternberg task' (Sternberg, 1966). For this task, participants are presented with a set of digits/letters to memorize, then following a brief delay, a 'probe' digit is presented and participants indicate by button press whether it was part of the set. There is a sizeable literature demonstrating that a large, sustained, and parietally distributed positivity is elicited by the probe with a linear increase in latency as memory set size increases (e.g. Blumhardt, 1996; Pelosi et al., 1992, 1995, 1998; Starr and Barrett, 1987; Verleger, 1997). Moreover, a P300 has been found to be evoked during the 'study phase' or encoding phase to digits subsequently retrieved, suggesting that elicitation of a P300 during encoding can be used as a predictor of successful retrieval (Chao and Knight, 1996; Golob and Starr, 2004; Kotchoubey et al., 1996; Patterson et al., 1991).

Past ERP studies have been helpful in elucidating the neurophysiological and theoretical aspects of working memory processing and in establishing links between specific ERP components and working memory functions. However, this research does not offer the clinical assessment-related benefits associated with the use of stimuli from standardized neuropsychological tests. Therefore, the purpose of the present study was to investigate the effectiveness and potential clinical utility of an ERP-compatible version of a subsection of the Digit Span task from the Wechsler Adult Intelligent Scale—Third Edition (WAIS-III; Wechsler, 1997a) in order to determine working memory capabilities in healthy participants. The use of healthy participants is an essential step in determining the validity and reliability of this technique before it can be applied to a patient population. The standard Digit Span task involves the recall of a series of digits, either in the order presented (Digit Span Forward [DF]) or reverse order (Digit Span Backward [DB]), and has been shown to have high construct validity and reliability (Wechsler, 1997b).

Similar to the ERP-adapted neuropsychological paradigms by Connolly and colleagues, the major goal of the ERP adapted Digit Span task was to target the same cognitive processes involved in successful completion of the standard task without requiring verbal responses from participants. To accomplish this, the 'recall' segment of the test was presented aurally rather than requiring participants to respond verbally as is done in the traditional version of the tests. For this study, only the DB task was selected for ERP adaptation (ERP-DB task) since it demands more effort from working memory resources than the DF task (Gardner, 1981; Mishra et al., 1985). In addition, due to the nature of the adaptation, the DB task was selected instead of the DF task in order to avoid possible facilitation of working memory performance by the use of a simple auditory pattern-matching strategy from hearing the digits replayed back in the exact order. Successful completion of a DB trial is believed to involve cognitive manipulation (Sattler, 1992; Weschler, 1997b), therefore replaying the digits in the reverse order is not expected to lead to the use of simple strategies that may inflate working memory performance.

For the ERP-DB task digits are played back either in the exact reverse order (correct condition) or with one error (incorrect condition). This is methodologically important because we predict that successful performance will be reflected by the elicitation of different ERP patterns across conditions within participants' ability range; thereby providing a neurophysiological marker capable of assessing a patient's ability level on the task. This prediction is based on the results of the ERP-adapted standardized tests by Connolly and colleagues (discussed above). In further support of our prediction, the ERP-DB task can be viewed as a sequence learning task and previous ERP sequence learning studies have demonstrated the elicitation of a robust parietal positive component to a violation of the expected pattern when the sequence is explicitly known by participants (Lang and Kotchoubey, 2002; Polich, 1985; Schlaghecken et al., 2000; Squires et al., 1976). Specifically, we hypothesized that: (1) healthy subjects' performance on the ERP-DB task will be comparable to the standard test; (2) different ERP patterns will be elicited in correct and incorrect conditions; and, (3) ERP patterns will change as a function of working memory load and will be related to working memory capacity.

2. Methods

2.1. Participants

Twenty university students (11 females) with a mean age of 22.2 years (s.d. = 2.6) were recruited from a departmental subject pool and participated in the study for course credit or $7.00/h. Participants were fluent English speakers with normal hearing and normal or corrected-to-normal vision and no history of neurological or psychiatric conditions. The participants were randomly divided into two groups of 10 (Group 1 and Group 2) and assigned to complete one of the two sections of the ERP-DB task (see below). Informed consent was obtained at the beginning of the experiment and participants were debriefed at the end of the session. The study was approved by the relevant institutional review board.

2.2. Standard digit span backward (DB) procedure

The standard DB test was administered twice to each participant, once before (pre-test) and once after (post-test) the completion of the ERP-DB task. This was done in an effort to account for possible practice or fatigue effects. To prevent the use of the same stimuli for both of the pre- and post-standard DB tests, the stimuli from the DF task was used for one of the DB tests. The standard DB task was administered as specified by the WAIS-III manual (Wechsler, 1997a). Briefly, the experimenter recited a set of digits (at the rate of one digit per second), which the participant repeated in the reverse order. The first set of digits consists of two digits. The set size increased by one digit every two trials. The test stopped when the subject had two consecutive errors at any given set size or when two successful trials at set size 8 had been reached. Success at each set size was determined by successful completion of 1 out of the 2 trials administered. However, because two standard tests (pre and post) were given in this experiment and there was no significant difference between them (refer to the Standard DB task subsection of the Results), completing a set size was defined by successful performance on 2 out of the 4 trials. The maximal set size achieved for each participant will be referred to hereafter as their ‘success level’.

2.3. ERP-DB task procedure

During the ERP-DB test, participants sat in a comfortable chair and the stimuli were presented aurally through earphones. The stimuli were recorded with the NeuroScan™ Incorporated stimulus software package at 90 dB SPL within a fixed duration of 1000 ms and digitally sampled at 20,000 Hz. The numerical stimuli consisted of the digits ‘1’ through ‘9’. Each trial consisted of a study phase and a test phase. The word ‘start’ indicated the beginning of each trial. For the study phase, a set of digits (ranging from 2 to 8 digits in length) was presented at the rate of 1 digit per second. Participants were instructed to listen carefully and to keep track of the order of the digits. The end of the study phase and beginning of the test phase was indicated by a tone that occurred 1 s after the last digit in the study phase. For the test phase, a second set of digits was presented that either corresponded exactly to the reverse order of the digits in the study phase (‘correct condition’) or in which one digit was replaced by an incorrect digit (‘incorrect condition’).

Several types of errors are common during administration of the standard DB task in both healthy and patient populations, including: use of an incorrect digit, reversal of two digits, giving up, repetition of a digit already in the sequence, recall of digits out of order (beyond a reversal error) and leaving out or adding an extra digit. The pattern of errors for participants' standard DB task was examined. It was found that an incorrect digit was the most frequent type of error (accounting for 21% of errors), followed by reversal errors (20%) and giving up (20%). It was speculated that each type of error might lead to different ERP patterns and therefore it was determined that only one type of error should be utilized in an effort to minimize the length of the paradigm. For the current paradigm, the use of an incorrect digit was selected as the error because it is a common mistake (Frankel and Tymckuk, 1974; Warschaudky et al., 1996) and afforded several benefits from an ERP paradigm perspective. For instance, the use of an incorrect digit (1) allowed for the same number of digits in both the study and test phases, (2) was the only type of error that allowed for only one mistaken digit in a sequence and (3) was believed by the authors to be the type of error that would elicit the largest amplitude ERP component. The last point is based on the assumption that an incorrect digit can be interpreted as a ‘complete’ or ‘pure’ violation, in contrast to, for example, a reversal error, that could be regarded as a ‘partial’ error. Numerous studies have demonstrated degraded ERP responses to partial errors versus complete errors (e.g. Coles et al., 2001; Connolly et al., 1999a; Kutas and Hillyard, 1984).

Participants were told to indicate if the trial was correct or incorrect by pressing one of two buttons at the end of the test phase (i.e. participants were told to withhold their button press until the end of the trial irrespective of when the incorrect digit occurred). There was a 2 s response window before the beginning of the next trial. Guessing was discouraged by instructing participants to withhold a response if they could not determine whether the test phase was correct or not. This was done to minimize and/or avoid correct button presses by chance. As accuracy rather than speed of button press responses was emphasized and guessing was discouraged, reaction time was not a relevant measure of performance and was not analyzed in this experiment.

Successful completion of each set size on the ERP-DB task was determined by a 65% behavioral success rate. The success cutoff was set at 65% in an attempt to account for possible inflated scores in the longer set sizes if the incorrect digit occurred early in the sequence.[1] Thus, ERP-DB task success levels were calculated as the maximal set size achieved using a button press success cutoff criterion of 65%.

The ERP-DB task was divided into two sections (10 different subjects participated in each section, forming two independent groups). Section 1 consisted of set sizes that were 2, 4, 6 or 8 digits in length (Group 1) and took 52 min to complete. Section 2 consisted of set sizes that were 3, 5 or 7 digits in length (Group 2) and took 40 min. Three breaks were given and evenly spread out in each of the two sections. Because the ultimate goal of this research is to devise a task suitable for patient populations, the division of the task was deemed necessary to minimize testing time and decrease fatigue. Although having the same subjects complete both sections of the task would have had certain advantages, the use of two independent groups allows for replication and increased generalization of the data within the same study.

For both sections of the ERP-DB task, the presentation of the stimuli mimicked the standard test such that two trials were given at each set size (starting with two digits in Section 1, and three digits in Section 2), and increased incrementally (up to a maximum of eight digits in Section 1, and a maximum of seven digits in Section 2). This pattern was then repeated 28 times for each of the sections for a total of 224 trials in Section 1 and 168 trials in Section 2. Due to the division of the ERP-DB task, the set sizes in the ERP-DB task increased by increments of two digits compared to the standard test that uses increments of one digit. In line with the standard format, stimuli in the study phase were presented pseudorandomly with the restriction that a digit could not be repeated in any one sequence. For the test phase, the digit sets were presented in the exact reverse order of the study phase 50% of the time (correct condition) or with one of the digits replaced with an incorrect digit (incorrect condition) for the remaining 50% of the time. The incorrect digit was randomly selected from the remaining digits not already in the digit set. The incorrect digit had an equal probability of replacing a digit anywhere in the sequence and occurring at anytime throughout the testing session.

[1] The 65% cutoff was calculated based on the statistical estimation, taking into consideration participants' error rates, of the expected inflation of accuracy due to an early onset of the incorrect digit in an incorrect trial. For example: for an 8 set size sequence, it was predicted that subjects would obtain inflated accuracy if the incorrect digit occurred within the first 3 digits of the reverse sequence, since, in general, participants were able to hold up to 3 digits in working memory without difficulty, as reflected by participants' > 90% accuracy at set sizes 2 and 3.

2.4. Electrophysiological recording procedures

EEG was recorded using Ag/AgCl electrodes, from three midline sites (Fz, Cz, and Pz) referred to linked earlobes (in accord with the 10–20 electrode system, Jasper, 1958). The electrooculogram (EOG) was recorded from electrodes placed above and below the left eye (vertical EOG) and from electrodes lateral to each eye (horizontal EOG). A ground electrode was placed on the right forearm. Electrode impedances were kept below 7 kΩ. The EEG was recorded with a bandpass of 0.01–100 Hz and digitally sampled at 500 Hz for 1000 ms (including a 100 ms prestimulus baseline) from digit onset and filtered off-line with a bandpass of 0.1–20 Hz.

2.5. Statistical procedures

2.5.1. Standard vs. ERP-DB task performance

For comparative analyses of participant performance on the standard and ERP-DB tasks, only set sizes presented in both versions were compared for Groups 1 and 2. A 'percent match' was calculated by comparing the success levels on both the standard and ERP-DB task. No correlations comparing success levels were performed due to the restricted range of the set sizes (four sets for Groups 1 and 3 for Group 2). In order to assess possible practice or fatigue effects, participants' pre- and post-success levels on the standard task were compared. For the ERP-DB task, performance on the first and second halves of the task was compared separately for Groups 1 and 2 using paired *t* tests. In addition, the potential impact of the location of the incorrect digit embedded within an incorrect trial on button press performance was assessed by a t-test for set size 2 and a series of one-way repeated measures ANOVAs (one for each additional set size), followed by LSD post-hoc tests with a Bonferroni correction when appropriate.

2.5.2. Grand average waveforms and ANOVA analyses

EEG trials with EOG voltages greater than ± 75 μV were discarded from the analyses. Following EOG artifact rejection, a mean of 83% of the data (with a range of 70–99%) was retained for the analyses in the correct condition for Group 1 and 80% (with a range of 59–97%) for Group 2. For the incorrect condition 82% of the data (range 58–98%) was retained for Group 1 and 82% (range 64–99%) for Group 2. Grand average waveforms were created for the correct and incorrect conditions at each set size (2, 4, 6, and 8 for Group 1, and 3, 5, and 7 for Group 2) at each of the three electrode sites (Fz, Cz, and Pz). All the digits from the test phase of correct trials were averaged together,[2] whereas only the incorrect digits from incorrect trials were averaged together to create the grand average waveforms for each condition. The main objective of the analyses was to determine if different ERP components were elicited in the correct and incorrect conditions and if the ERP patterns related to working memory load and capacity. Upon visual inspection of the grand averages, two major centro-parietal positive deflections emerged and were selected for analyses, (1) a positive slow wave (PSW) elicited in the incorrect condition with an average peak latency of 599 ms post-stimulus onset and, (2) an earlier positive component (P200/P300) elicited in the correct condition with an average peak latency of 296 ms post-stimulus onset.

Separate analyses for the PSW and P200/P300 were carried out for Groups 1 and Group 2. The PSW was analyzed by a three-way repeated measures analysis of variance (RM ANOVA) with CONDITION (correct and incorrect, C/I), SET SIZE (2, 4, 6, and 8 for Group 1 or 3, 5, and 7 for Group 2) and SITE (Fz, Cz and Pz) as factors.[3] The PSW was defined as the most positive peak occurring between 450 and 750 ms. The P200/P300 was analyzed by a four-way RM ANOVA, with the same factors and levels as the PSW analyses with the addition of an INTERVAL factor (200–250, 250–300, 300–350, and 350–400 ms). For this analysis, the most positive peak amplitude for each 50 ms latency interval was selected. A positive deflection within 200–250 ms was interpreted as a P200, whereas a positive deflection within 250–400 ms was interpreted as a P300. These ANOVA analyses were subjected to a Greenhouse-Geisser conservative degrees of freedom correction (Greenhouse and Geisser, 1959). Relevant significant main effects and interaction effects were further analyzed by Tukey Honestly Significant Difference (HSD) post-hoc comparisons with a probability level set at $P<0.05$ for statistical significance. Only main and interaction effects considered germane to the experimental hypotheses were investigated (i.e. significant effects collapsed across condition were not explored).

[2] Significantly more stimuli were averaged in the correct compared to the incorrect condition; a common issue in many paradigms such as the oddball design. It is noted that this may lead to certain errors of measurement due to S/N ratio differences or even a 'washing out' of some smaller effects in the correct condition. To address this concern, a pseudorandom subsection of correct digits (that were directly selected and matched for the location and N of the corresponding incorrect digits) were averaged and compared to the correct condition grand average to ensure that there were no significant or visually apparent pattern differences ($P<0.05$). No differences were observed and thus all data from the correct condition were included for the analyses.

[3] The PSW was also analyzed with a four-Way RM ANOVA with an additional INTERVAL factor (450–550, 550–650 and 650–750 ms intervals). No significant main or relevant INTERVAL interaction effects were found using peak or mean interval scoring. As it contributed nothing to the understanding of these data the INTERVAL variable was discarded from all statistical analyses. The major findings relating to the PSW were significant across the two forms of analyses.

3. Results

3.1. Behavioral performance

3.1.1. Standard DB task

Table 1 depicts success levels for each participant on the standard DB test, as well as the mean and standard deviation for Groups 1 and 2, respectively. For both Groups 1 and 2, participants' average success level was between set size 5 and 6 (Group 1: mean = 5.6; s.d. = 0.84 and Group 2: mean = 5.8; s.d. = 1.03). No practice or fatigue effects were found between participants' pre- and post-standard DB performance for Groups 1 and 2 ($P>0.05$).

3.1.2. ERP-DB task

Table 1 depicts the percentage of correct button press responses for each of the set sizes for Groups 1 and 2, respectively. For both Groups 1 and 2, the percentage of correct button press responses decreased as set size increased. The mean success level achieved was 6.0 for both Groups 1 and 2. No practice or fatigue effects were found between participants' performance during the first and second half of the ERP-DB task for either Group 1 or 2 ($P>0.05$).

The one-way RM ANOVAs to assess the impact of the location of the incorrect digit on button press performance indicated that regardless of the location of the incorrect digit within the sequence, there were no significant differences in percent correct button presses for set sizes 2 ($P>0.05$) and 4 ($F(3)=0.353$, $P>0.05$) for Group 1 and at set sizes 3 ($F(2)=1.0$, $P>0.05$) and 5 ($F(4)=2.32$, $P>0.05$) for Group 2. For Group 1, at set size 6 ($F(5)=3.067$, $P<0.05$) the one-way RM ANOVA was significant, however, the LSD post-hoc tests with the Bonferroni correction indicated that none of the comparisons were significantly different from each other, although there was a trend that approached statistical significance for increased performance when the incorrect digit was in the first location compared to the other locations. At set size 8 ($F(7)=3.266$), performance was significantly increased when the incorrect digit occurred in the first location compared to the fourth position. For Group 2, participants' button press performance was significantly greater at set size 7 ($F(6)=8.37$, $P>0.05$) when the incorrect correct digit occurred in the first location compared to the last 4 locations.

3.1.3. Standard vs. ERP-DB task

A percentage was computed for the number of participants that obtained the same success level on both the standard and ERP-DB tasks. The success level matched in 7 out of 10 participants for Group 1 (70%, Table 1) and for 9 out of 10 participants in Group 2 (90%, Table 1).

3.2. ERP analyses

Figs. 1 and 2 depict the grand average waveforms for the correct and incorrect conditions for each electrode site at each set size for Groups 1 and 2, respectively.

Table 1
Percentage of correct button press responses for each set size on the ERP-DB task as well as the success levels achieved on both the ERP-DB and standard tasks for each participant in A. Group 1 and B. Group 2

	% Correct button press				Success level	
	Set sizes				ERP-DB task	Standard task
	2	4	6	8		
A. Group 1						
S01	98.2	82.1	69.6	48.2	6	6
S02	96.4	69.6	67.9	37.5	6	6
S03	100.0	92.9	71.4	57.1	6	4
S04	98.2	91.1	69.6	51.8	6	6
S05	100.0	80.4	69.6	58.9	6	4
S06	100.0	100.0	85.7	66.1	8	8
S07	94.6	92.9	96.4	60.7	6	6
S08	100.0	87.5	58.9	42.9	4	6
S09	91.1	85.7	80.4	60.7	6	6
S10	100	96.4	75.0	41.1	6	6
Mean	97.9	87.9	74.5	52.5	6.0	5.6
SD	3.0	8.9	10.6	9.7	0.9	0.8
Percent match					70%	

	% Correct button press			Success level	
	Set sizes			ERP-DB task	Standard task
	3	5	7		
B. Group 2					
S11	91.1	73.2	53.6	5	5
S12	96.4	85.7	71.4	7	7
S13	96.4	98.2	69.6	7	5
S14	96.4	94.6	67.9	7	7
S15	96.4	85.7	60.7	5	5
S16	98.2	91.1	73.2	7	7
S17	96.4	94.6	51.8	5	5
S18	92.9	80.4	53.6	5	5
S19	98.2	94.6	67.9	7	7
S20	92.9	85.7	55.4	5	5
Mean	95.5	88.4	62.5	6	5.8
SD	2.4	7.7	8.4	1.1	1.0
Percent match				90%	

The percent match between the standard and ERP-DB task success levels are also presented.

3.3. PSW

3.3.1. Group 1

The three-way RM ANOVA (refer to Table 2A) indicated a significant main effect of C/I, reflecting a significant increase in amplitude in the incorrect compared to the correct condition. The interaction effects for Group 1 are best interpreted within the significant C/I × SITE × SET SIZE interaction. Therefore, post-hoc tests were computed only on this interaction effect. In terms of topography, the results clearly indicate the presence of a parietally distributed PSW in the incorrect condition (Pz > Cz and Fz for all set sizes and Cz > Fz for set sizes 2, 4 and 8, $P < 0.05$). Comparing across conditions, PSW amplitude was significantly greater in the incorrect condition compared to the correct condition at all three electrode sites for set sizes 2, 4 and 6. The same pattern was evident at set size 8, but was only significant at Cz and Pz. In the correct condition, there were no significant amplitude differences across set sizes or electrode sites.

In addition, PSW amplitude in the incorrect condition declined as set size increased. Thus, at Pz and Cz sites, PSW amplitude in the incorrect condition was significantly greater at set sizes 2 and 4 compared to set sizes 6 and 8 as well as from set sizes 4 and 6 compared to set size 8. At the Fz site, the PSW amplitude in the incorrect condition at set sizes 2, 4, and 6 was significantly larger than at set size 8.

Fig. 3 plots both the PSW amplitude for the incorrect condition and the difference in PSW amplitude between incorrect and correct conditions (hereafter referred to as 'the difference PSW or dPSW') for each set size at each of the three electrode sites. From Fig. 3, the significant decreases in PSW and the dPSW amplitudes as set size increases are apparent; the only exception being for set size 2, which has lower PSW and dPSW amplitudes than set size 4. Furthermore, the decrease in PSW and dPSW amplitude

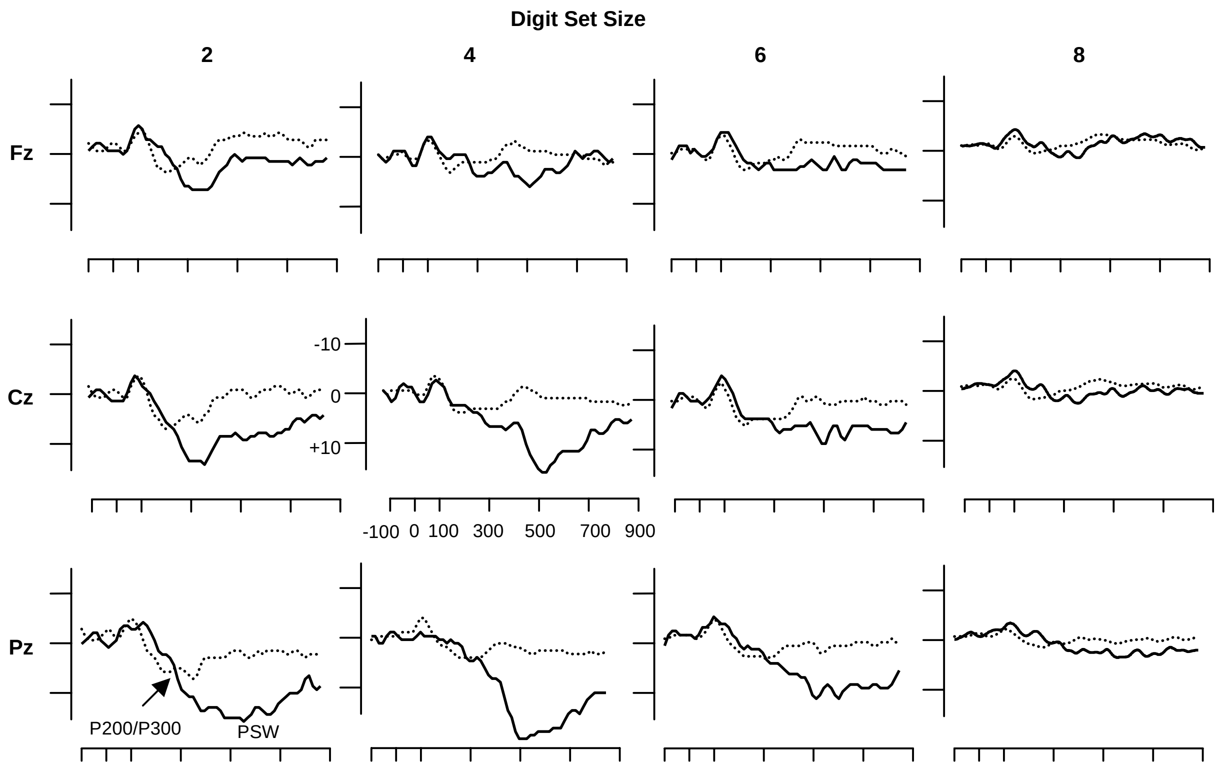

Fig. 1. Grand average waveforms from the ERP-DB task for Group 1 depicting the response to correct and incorrect conditions at each set size (2, 4, 6, and 8) for each of the electrode placement sites (Fz, Cz, and Pz). The *X*-axis represents time (ms) and the *Y*-axis represents amplitude (μV).

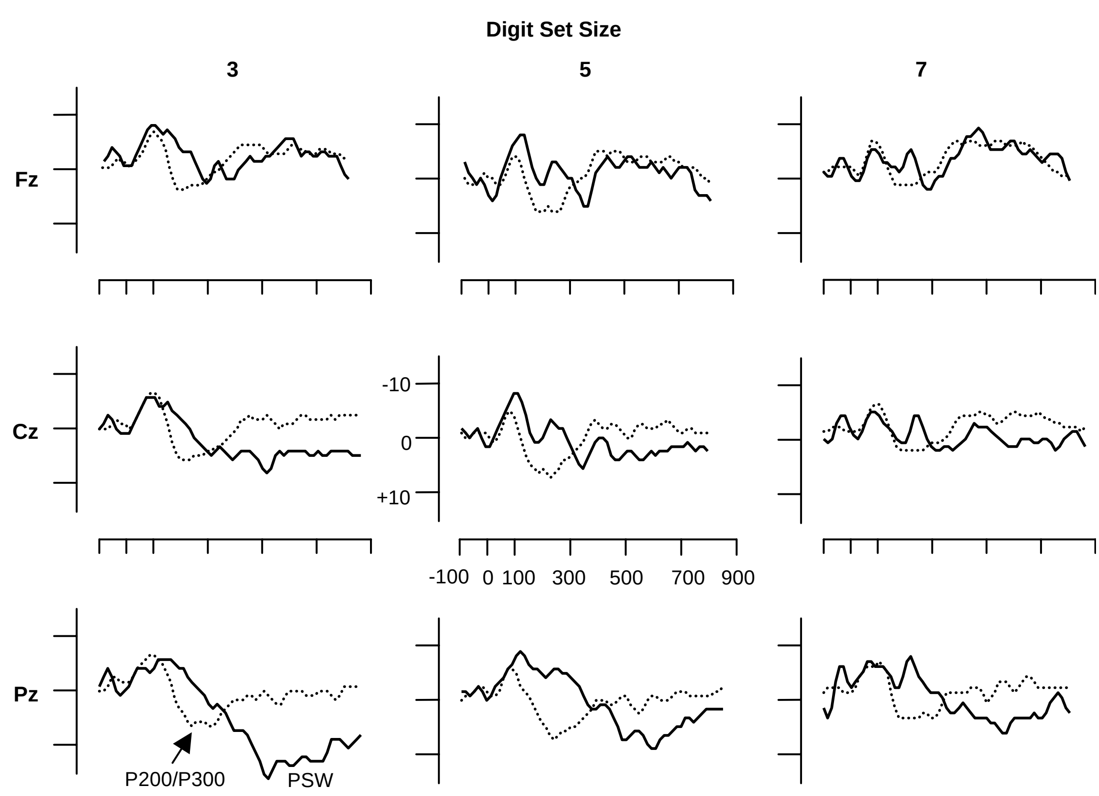

Fig. 2. Grand average waveforms for the ERP-DB task for Group 2 depicting the response to correct and incorrect conditions at each set size (3, 5, and 7) for each of the electrode sites (Fz, Cz, and Pz). The *X*-axis represents time (ms) and the *Y*-axis represents amplitude (μV).

Table 2A
RM ANOVA for the PSW with CONDITION (C/I), SITE (S) and SET SIZE (SS) as factors for Groups 1 and 2

Effects	Group 1		Group 2	
	df	*F*	df	*F*
C/I	1, 9	38.20***	1, 9	55.46***
SITE	2, 18	87.43***	2, 18	21.78***
SET SIZE	3, 27	4.02*	2, 18	2.05
C/I×S	2, 18	20.93**	2, 18	12.75**
C/I×SS	3, 27	3.19	2, 18	0.33
S×SS	6, 54	5.04*	4, 36	1.83
C/I×S×SS	6, 54	3.35*	4, 36	1.99

as set size increases was most prominent at Pz and became smaller in a posterior to anterior direction (i.e. Pz > Cz > Fz; Fig. 3). To further explore the relation between PSW amplitude and working memory capacity, the change in amplitude (i.e. slope) between successive set sizes was compared to participants' maximum success level on the standard DB task. The steepest decline in PSW amplitude was found to occur in 70% of participants when they reached their maximum performance level. The dPSW measure is important because it can provide a neurophysiological index to determine if participants are able to distinguish correct from incorrect conditions.

3.3.2. Group 2

The three-Way RM ANOVA (refer to Table 2A) indicated a significant main effect of C/I, where the PSW amplitude in the incorrect condition was significantly larger than in the correct condition. There was also a significant interaction effect for C/I×SITE, indicating a significant parietally focused positivity in the incorrect compared to the correct condition. Specifically, post-hoc tests revealed a significant increase in positive amplitude in the incorrect compared to the correct condition at all three electrode sites. In addition, the PSW amplitude in the incorrect condition was significantly greater at Pz compared to Cz and Fz.

For Group 2, no SET SIZE interaction effects were significant, indicating that the PSW amplitude did not change significantly with working memory load. However, a trend of decreased PSW and dPSW amplitude as set size increases is apparent (Fig. 3). In addition, this trend is more apparent at posterior than anterior sites (Pz > Cz > Fz). As done with Group 1, to further explore the relation between PSW amplitude and working memory capacity, the slope between successive levels was compared to participants' success level on the standard DB task. The steepest decline in PSW amplitude occurred in 70% of participants when they reached their maximum success level on the standardized task.

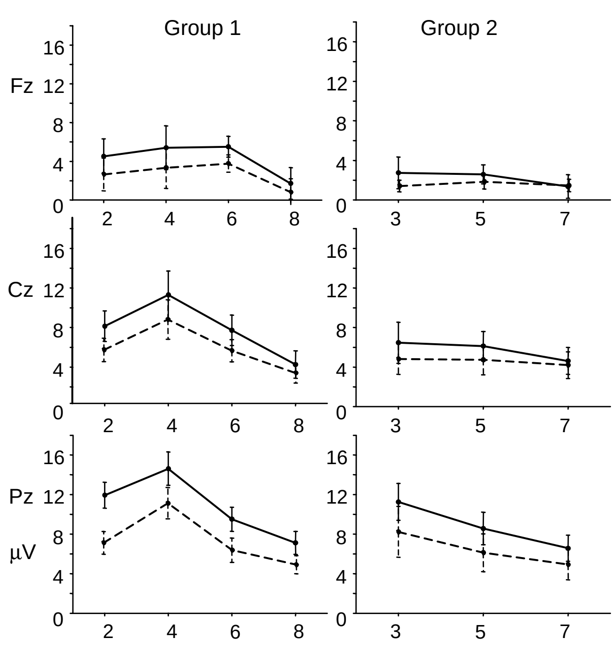

Fig. 3. The PSW peak amplitude for the incorrect condition and the difference in PSW amplitude between incorrect and correct conditions (dPSW) peak amplitudes at each of the electrode sites for each set size. The *X*-axis represents the set size and the *Y*-axis represents amplitude (μV).

3.4. P200/P300

3.4.1. Group 1

The four-Way RM ANOVA (refer to Table 2B) indicated a significant main effect of C/I. The P300 amplitude in the incorrect condition was significantly larger than in the correct condition. Although the analysis revealed both a significant two-way C/I×INTERVAL and three-way C/I×SITE×INTERVAL interactions, the effects are best interpreted within the three-way interaction. In terms of topography, there is evidence of a broad centrally located

Table 2B
RM ANOVA for the P300 with CONDITION (C/I), SITE (S), SET SIZE (SS) and INTERVAL (I) as factors for Groups 1 and 2

Effects	Group 1		Group 2	
	df	*F*	df	*F*
C/I	1, 9	6.50*	1, 9	1.77
SITE	2, 18	7.51*	2, 18	2.31
SET SIZE	2, 18	4.88*	2, 18	1.86
INTERVAL	4, 36	1.86	3, 27	0.31
C/I×S	2, 18	2.44	2, 18	10.02**
C/I×SS	2, 18	0.19	2, 18	0.41
C/I×I	4, 36	29.21***	3, 27	14.36***
S×SS	4, 36	2.02	4, 36	1.89
S×I	8, 72	8.79***	6, 54	4.58*
SS×I	8, 72	1.48	6, 54	0.52
C/I×S×SS	4, 36	0.64	4, 36	0.34
C/I×S×I	8, 72	6.34**	6, 54	2.25
C/I×SS×I	8, 72	0.88	6, 54	1.26
S×SS×I	16, 144	1.06	12, 108	0.82
C/I×S×SS×I	16, 144	0.88	12, 108	1.14

(* $P<0.05$, ** $P<0.01$ and *** $P<0.001$ after Greenhouse-Geisser correction).

P200 component within the 200–250 ms interval as well as evidence of a parietally focused P300 component within the 250–400 ms intervals for the correct condition. Specifically, for the 200–250 ms interval, there were no significance differences in amplitude across the three electrode sites, however, mean amplitude was greatest at Cz compared to Fz and Pz. Within the 250–400 ms intervals, amplitude was largest at Pz and decreased anteriorly across the scalp (reaching significance for Pz vs. Fz at all three time intervals and Pz vs. Cz for the 300–400 ms intervals; mean amplitude at Cz was also significantly larger than Fz for all three time intervals). For Fz the amplitude in the correct condition was greatest at 200–250 ms and decreased significantly across the latency intervals (250–400 ms). A similar pattern was seen at Cz, where the amplitude was largest at 200–300 ms, but significantly decreased from 300 to 400 ms. In contrast, at Pz amplitude was maintained across all the four latency intervals.

When comparing across conditions, the amplitude of the P200 (200–250 ms interval) is significantly larger for the correct than incorrect condition at Pz. However, within the 250–400 ms interval there is evidence of positive-going responses for both correct and incorrect conditions but amplitudes were found to be significantly larger for the incorrect than correct condition at 250–400 ms for Fz and Cz and from 300 to 400 ms for Pz. For the incorrect condition, the amplitude significantly increased across all the four latency intervals at all sites (except for the 300–350 ms interval compared to the 350–400 ms interval at Fz and Cz). Although a positivity was found in both conditions, the positivity in the correct condition had a peak latency at approximately 300 ms and was interpreted as a P300, whereas the positivity elicited in the incorrect condition was prolonged and peaked only within the 450–750 ms latency range and was therefore interpreted as the onset of the PSW.

3.4.2. Group 2

The 4-Way RM ANOVA (refer to Table 2B) indicated a significant interaction effect for C/I×SITE and C/I×INTERVAL. In terms of topography, post-hoc tests from both interactions revealed evidence of a parietally focused positive component across the 200–400 ms interval within the correct condition. Specifically, post-hoc tests from the C/I×SITE interaction indicated that the amplitude in the correct condition was significantly greater at Pz compared to Cz and Fz, while post-hoc tests from the C/I×INTERVAL interaction revealed that the positive amplitude in the correct condition was larger at 200–300 ms compared to 350–400 ms. For the incorrect condition, there was evidence of a negative shift within the 200–300 ms (N2) time intervals and a positivity within the 300–400 ms intervals. This was evidenced by post-hoc tests from the C/I×INTERVAL interaction demonstrating a significant increase in positive amplitude at 300–400 ms compared to 200–300 ms. Comparing across conditions, post-hoc tests from the C/I×SITE interaction indicated a significant increase in positive amplitude in the correct compared to the incorrect condition at Pz, while the post-hoc tests from the C/I×INTERVAL interaction indicated that, at 200–300 ms, the positive amplitude in the correct condition was significantly larger than in the incorrect condition.

Taken together, the results from the two significant interaction effects were attributable to the parietal positive component within the 200–400 ms range for the correct condition. This response was interpreted as an amalgamation of a P200/P300 response. However, in the incorrect condition, an N2 was present in the 200–300 ms range followed by a postivity; which is interpreted as the onset of the PSW (that subsequently peaked within the 450–750 ms period).

4. Discussion

The findings from this experiment support our three hypotheses. Firstly, this study provides support that behavioral performance on the ERP-DB task is comparable to performance on the standard version of the test. Secondly, it was demonstrated that, in a normal population, it is possible to distinguish correct and incorrect conditions in the ERP-DB task by a parietally distributed PSW elicited by the incorrect condition and a P200/P300 component elicited by the correct condition. Finally, the PSW and dPSW peak amplitudes change as a function of working memory load (however, this trend was only statistically significant for Group 1) and are related to working memory capacity.

4.1. DB behavioral performance

Participants' maximal success level on both the standard and ERP-DB task was found to match in 70% of Group 1 and 90% of Group 2 participants. Participants scored within the average range for their age group when comparing their standard and ERP-DB task scores to those of the Canadian normative database (ages 20–29) provided in the WAIS-III Canadian Technical manual (Wechsler, 2001). The Canadian average digit span backward level for this age group ranged between 5 and 6 digits (mean = 5.32). In accordance with these norms, participants from our study had an average digit span performance of approximately 5 or 6 digits (mean = 5.5 for Group 1 and 5.8 for Group 2, excluding levels not administered in the ERP-DB task). On the ERP-DB task, using the 65% success criterion, 80% of subjects in Group 1 achieved up to set size 6 (mean = 6.0) and in Group 2, 50% of participants achieved up to set size 5 and 50% up to set size 7 (mean = 6.0).

Although participants' success levels on both the standard and ERP-DB task were well matched and corresponded with the Canadian norm base, it is important to consider the modifications from the standardized procedures that were necessary for the ERP adaptation.

Since the ultimate goal of this research is to assess working memory abilities in non-communicative neurotrauma patients, it was essential to design the ERP paradigm in such a way that it did not require verbal responses from the participant. This was accomplished by playing back the sequence of digits to the participant rather than having them verbally repeat them. This necessary modification may alter the cognitive strategies or processes that are important for completion of the standard DB task. Successful completion of the standard DB task is presumed to involve the ability to (1) understand and encode the digits; (2) retain the information in working memory; (3) manipulate and organize the information in working memory (into the reverse order); and, (4) retrieve and recall the information from working memory.

In the ERP-DB task, this fourth cognitive process is altered so that the participant retrieves and compares the information stored in their working memory with the digits being played back in order to determine the correctness of the trial rather than recalling the information verbally. It is possible that this type of comparison process is less demanding than recall and, therefore, the ERP-DB task may be easier than the standard version. However, there was no evidence of increased behavioral accuracy on the ERP-DB compared to the standard test when the 65% success cut-off criterion was applied. In addition, the and standard ERP-DB versions demonstrated a similar trend in the error rate; participants rarely made incorrect responses on set sizes 2, 3, and 4 in either the standard or ERP-DB task.

As outlined above, the majority of the cognitive processes thought to be involved in the standard test are met by the conditions of the ERP-DB task, with the exception of a comparison/recognition process rather than recall. The most fundamental commonality between the two versions is that both tasks involve working memory processes, which include storage, encoding, manipulation and retrieval of information from working memory. The importance of ensuring that the modifications required for ERP adaptation do not significantly affect performance has been emphasized (Byrne et al., 1995b; Connolly et al., 1995). In this experiment, the comparable behavioral accuracy between the standard and ERP-DB tasks satisfies this requirement. Therefore, we can conclude that the ERP-DB task demonstrates validity when compared to the criterion-standardized measure.

In addition, no practice or fatigue effects were demonstrated, suggesting that the increased number of trials and length of the ERP-DB task did not influence behavioral performance over time. The lack of practice and fatigue effects on the ERP-DB task as well as the lack of pre to post differences on the standard DB task demonstrates that DB test performance within individuals is relatively stable across both extended and multiple testing situations; a finding that confirms the high test–retest stability coefficient (0.86) reported in the WAIS-III and WMS-III Technical Manual (Wechsler, 1997b).

Moreover, the impact of the location of the incorrect digit did not have any significant impact on button press performance on digit set sizes that were within participants' ability range (e.g. set sizes 2, 4, and 6 for Group 1 and set sizes 3 and 5 for Group 2). However, as the set sizes became more difficult and beyond participants ability range (set size 8 for Group 1 and set size 7 for Group 2), there was a significant increase in participants' accuracy when the incorrect digit was the first digit played back. This suggests that participants' used the location of the incorrect digits as a strategy to increase button press performance only at levels that exceeded their ability range. Therefore, there is no evidence of artificially inflated performance as a result of the modifications to the DB task for ERP adaptation at digit set sizes within participants' ability range, and only minimal inflation at the higher set sizes.

4.2. Grand average neural responses

The grand average waveforms from the ERP-DB task demonstrate that two distinguishable posterior positive components are elicited to each of the two conditions for both Groups 1 and 2; an earlier P200/P300 component (peaking between 200 and 400 ms post-digit onset) elicited in the correct condition and a comparatively robust and prolonged PSW (peaking between 450 and 750 ms post digit onset) elicited in the incorrect condition. In addition, the PSW and the dPSW decreased in amplitude as working memory load and task difficulty increased. For Group 1, the PSW and the dPSW peak amplitudes were greatest at the lower set sizes (set size 2 and 4), but significantly reduced as working memory load increased (set size 6 and 8). Specifically, for Group 1, both the PSW and dPSW were present at set size 2, greatest at set size 4, reduced at set size 6 and further attenuated at set size 8. This pattern was strongest parietally and decreased steadily across the scalp in an anterior direction. The only exception to the reduction in PSW amplitude as set size increased was that PSW amplitudes were larger at set size 4 compared with set size 2. This effect could be explained by task complexity and neural resource demands, such that successful completion of trials containing 2 digits may have been easier and required less neural resources than trials at set size 4. Whereas, subsequent decreases in PSW amplitudes at set sizes higher than 4 are interpreted as being due to task demands, which challenged or exceeded participants' working memory abilities.

Similar to Group 1, Group 2 showed a decrease in the PSW and the dPSW peak amplitudes in the incorrect condition as set size increased, although these effects were not statistically significant. Also, like Group 1, the PSW in Group 2 showed a parietal distribution with decreasing amplitudes at anterior sites. The comparability of the results for both Groups 1 and 2 reflect the reliability of the ERP-DB task for two independent groups within the same study. The lack of significant PSW change across set size in Group 2 may be due to the restricted range of the set sizes in Group 2 compared to Group 1. That is, the maximal level of difficulty in Group 2 (set size 7) did not challenge participants' working memory capacity to the same degree as for Group 1 (set size 8). The button press performance scores support this interpretation since 50% of the participants successfully completed set size 7 in Group 2, yet in Group 1, only 10% of subjects successfully completed set size 8.

To further explore the link between PSW amplitude and working memory capacity, it was found that the steepest decline in PSW amplitude occurred in 70% of participants, for both Groups 1 and 2, when they reached their maximum success level on the standardized task. This result supports the idea that the slope of PSW amplitude decline may be used as a neurophysiological marker associated with working memory capacity.

Interestingly, for Group 1, both the PSW and the dPSW amplitudes were significantly attenuated at levels of difficulty that challenged (e.g. set size 6) or surpassed participants' working memory capacity (e.g. set size 8) compared to the less challenging levels (e.g. set sizes 2 and 4). Studies adapting stimuli from standardized tests of vocabulary abilities (PPVT-R and Vocabulary section of the WAIS-R-NI) also found that reliable ERP differentiation between correct and incorrect conditions were attenuated or reversed when the difficulty level of the vocabulary words exceeded the participants' abilities (Byrne et al., 1995b; Connolly et al., 1995, 1999a). This pattern of differentiation between tasks that are 'within' and tasks that are 'beyond' a participant's ability level is important because such ERP patterns may serve as a marker for successful or unsuccessful performance, respectively. Importantly, observation of such patterns in non-communicative patients would have the potential to identify the performance limits of patients who are otherwise impossible to assess.

Although there was a significant decrease in the PSW from all set sizes compared to set size 8 (the level that exceeded 90% of participants working memory capacity), there was still a significant dPSW peak at set size 8 at Pz and Cz sites. Group 1 participants were still able to achieve an average of 52.5% correct button press responses at set size 8; a performance level that likely contributed to the elicitation of an PSW even at this set size. In addition, although it is not possible to investigate the impact of the incorrect digit location on the PSW due to the limited number of ERP trials for each location in the higher set sizes, the inflated button press accuracy when the incorrect digit was the first digit to be replayed at set size 8 likely contributed to the elicitation of an PSW at this set size. The potential inflation in PSW amplitude at higher set sizes could be avoided by excluding trials where the incorrect digit is in the first location. However, in this study, exclusion would have resulted in too few accepted, artifact-free trials to achieve an adequate signal-to-noise ratio.

In accordance with past studies, a parietal positive wave was the major component influenced by working memory load (García-Larrea and Cézanne-Bert, 1998; Nittono et al., 1999; Ruchkin et al., 1990). In studies utilizing a modified Sternberg task, a major parietal positive wave was also the major component influenced by working memory load and elicited to probe digits, irrespective of whether the probe was part of the memory set or not (e.g. Ford et al., 1979; Grippo et al., 1996; Pratt et al., 1989). However, in our study, there are two parietal positive components within the 200–800 ms post-stimulus latency range (a P200/P300 elicited by correct digits, and a PSW elicited by incorrect digits).

The results of this study could be interpreted within an oddball framework, such that for each trial, the incorrect digits may be interpreted as deviant stimuli within an array of standard (correct) stimuli. From this perspective, the PSW to incorrect digits could be interpreted as a classic P300 response to deviants; while the P200/P300 elicited by correct digits could be similar to a P250 component, described by García-Larrea et al. (1992), associated with processing non-relevant stimuli during an odd-ball task and believed to be related to indexing aspects of stimulus classification. However, the P250 response is widely distributed across the scalp with maximal amplitude at the vertex and frontal regions. In contrast, in our paradigm, the positive component evoked by correct digits (within 250–400 ms) shares a similar parietal maximum to a classic P300 response. In addition, successful completion of the task demands processing all the stimuli, therefore correct digits are meaningful task-relevant stimuli. Thus, it is believed that the processing of correct digits involves both stimulus identification and a working memory component; processes that have been associated with a classic P300 response (Kok, 1997).

Similar to our results, in a two-choice reaction time working memory task, Nittono et al. (1999) also found evidence for two parietal-maximal positive components (each referred to as a P300 response); an earlier component elicited to the task-relevant frequent stimuli and a prolonged and increased amplitude component to the rare stimuli. ERP sequencing studies have also demonstrated that P300 (and/or PSW) components are elicited to both expected (correct) sequence items and violations (incorrect) of task-relevant sequence items (Baldwin and Kutas, 1997; Eimer et al., 1996; Lang and Kotchoubey, 2002; Polich, 1985; Schlaghecken et al., 2000). Generally, the P300 to violations was prolonged and of increased amplitude compared to correct items, although in some cases, the P300 was similar in amplitude and latency in response to both correct and incorrect stimuli (Baldwin and Kutas, 1997; Eimer et al., 1996; Lang and Kotchoubey, 2002).

An additional difficulty of interpreting the ERP-DB task within an odd-ball paradigm perspective is that probability effects alone cannot account for the PSW results in this study. For the ERP-DB task, the probability of an incorrect digit changes with each digit set size. Although it is well known that a classic odd-ball P300 is directly affected by stimulus probability (Johnson, 1986), such a direct relationship between PSW amplitude and probability was not observed in this study. Instead, changes in PSW amplitude are better accounted for by working memory performance levels. Even though the likelihood of an incorrect digit became rarer at higher set sizes, amplitude decreased as participants reached the limitations of their working memory capacities. The notion of decreased information transmission, including equivocation or decreased attention allocation (another factor documented to affect P300 amplitude) may be applicable (Johnson, 1986). Thus, amplitude attenuation at the higher digit set sizes to incorrect digits may be associated with participants' inability to adequately absorb/maintain all the information provided by each stimulus. Alternatively, participants may have maximized their attentional resources and were unable to allocate additional processing to the new incoming stimuli.

Another stimulus-related feature widely demonstrated to affect P300 amplitude is the unexpectedness of an event, regardless of probability effects (Johnson, 1986). The overlapping onset, similar topography and prolonged nature of the PSW in the incorrect condition (relative to the P300 in the correct condition) could be interpreted as a secondary or additive P300 response, involving an initial positivity relating to stimulus identification and working memory, followed by secondary processing related to a violation of expectation. The notion of multiple P300s or overlapping/prolonged positive components during memory tasks has been discussed by various authors (Fabiani et al., 1986; García-Larrea and Cézanne-Bert, 1998; Johnson and Donchin, 1985; Rugg, 1995).

The interpretation that the PSW elicited to incorrect digits is associated with secondary processing related to a violation of expectancy is in line with ERP sequencing studies, which have demonstrated a larger and prolonged posterior positive wave to the incorrect ending of a consecutive sequence of digits/letters, even though the endings had an equal probability of being correct or incorrect (Lang and Kotchoubey, 2002; Polich, 1985). We interpret the PSW decreases at higher set sizes as due to limitations of working memory capacity, where participants were unable to develop an expectation of the upcoming digits, which in turn, resulted in PSW attenuation. This interpretation is supported by findings from a sequence learning task where participants were presented with a recurring, 16-letter sequence, in which an occasional deviant letter replaced one of the letters (Schlaghecken et al., 2000). Deviant stimuli elicited N2b (230–330 ms) and P3b (400–600 ms) components, which were enhanced in explicitly learned chunks of the sequence compared to unlearned sections. Schlaghecken et al. (2000) concluded that the N2b/P3b components reflect a violation of explicit knowledge of learned chunks, and hence can provide a measure of explicit knowledge. Therefore, the P3b elicited by deviant stimuli in explicitly learned sequence chunks can be seen as analogous to the elicitation of the PSW in our study to incorrect digits within participants' span. Correspondingly, the reduction in PSW amplitude in the present study as set size increased beyond participants' ability can be interpreted as a reflection of participants' inability to explicitly learn the set. In a visual-spatial sequencing task, Baldwin and Kutas (1997) also found that a N2/P3 complex reflected both the confirmation and violation of sequential expectancy; a small positivity between 250 and 350 ms corresponded to an 'expected target effect' and a robust and delayed P300 to ungrammatical targets in unexpected locations. This effect was also amplified when participants had explicit awareness of the sequence structure. With these facts in mind, it is possible that the PSW observed in our study is quite similar, to the P3b observed in these other studies.

Moreover, previous studies have demonstrated links between P300 amplitude and explicit memory performance, causing amplitude to increase along with task demands as long as performance remains high, but if performance declines then the amplitude is attenuated (Johnson and Donchin, 1985; Nittono et al., 1999; Ruchkin et al., 1980; Smith and Guster, 1993). Nittono et al. (1999) further investigated the link between P300 and performance and found participants with high reading span scores demonstrated increased P300 amplitude on a working memory task, compared to participants with low reading span scores. They concluded that the high group was able to allocate more attentional and processing resources compared to the low group.

It could be argued that the PSW elicited by incorrect digits might be a reflection of trial completion rather than a component elicited by a correctly identified incorrect digit. That is, the PSW may reflect a participant's perception that a trial has ended by virtue of the occurrence of an incorrect item rather than a response to the incorrect item per se. To assess this interpretation, the last digits of all correct trials were averaged together. The resulting waveform did not exhibit an PSW and did not significantly differ from the grand average waveforms for all the correct digits. Thus, it is unlikely that the PSW reflects the perception of a trial as ended.

Due to the use of an incorrect digit as the error for incorrect trials, it may have been possible for participants' to wait for an out-of-set digit; a strategy that would not be dependant on mentally reversing the original sequence. However, this strategy would still have required holding and re-ordering the digits in working memory. In other words, the use of this strategy would still involve the key components of working memory (e.g. encoding, storage and manipulation of information). After completion of the task, participants were asked about strategies and no participant acknowledged the use of this strategy. The most common strategies reported included rote memorization, rehearsal, visualization, or chunking of the sequence; all of which are strategies commonly employed during the standard DB administration (Sattler, 1992; Wechsler, 1997a).

4.3. General discussion

Overall, this study demonstrates that the combining of ERP and neuropsychological methods has the potential to provide a neurophysiological index of working memory abilities. Nonetheless, there are a few issues that warrant consideration. Firstly, there are concerns associated with attributing a lack of ERP differentiation between correct and incorrect conditions as a reflection of working memory disabilities and not other cognitive difficulties. This issue becomes even more relevant when testing non-communicative patients with motor difficulties since button press data cannot be collected. This limitation also applies to other research that has adapted standard neuropsychological tests and to neuropsychological testing in general, since many standard neuropsychological tests often rely on the use of several cognitive processes and therefore a series of tests are needed to tease these abilities apart.

Secondly, working memory tasks demand attention and concentration and, if a patient is not motivated, ERP neural patterns may be attenuated (Bennington and Polich, 1999; Polich and Kok, 1995). The ERP-DB task used in this study involves additional demand on attention and concentration over the standard task due to the length of the testing session (52 min for Group 1 and 40 min for Group 2). Therefore, fatigue and/or waning of attention and concentration may attenuate ERP differentiation. In the present study fatigue and practice effects were not demonstrated and there is no reason to believe that the differentiation demonstrated was dampened by these effects. However, decreased concentration and fatigue effects may be more of a concern in clinical populations with neurological conditions (deGroot et al., 2003).

Despite the obvious benefits of an accurate working memory assessment tool that is both independent of language abilities and provides valuable neuropsychological information, it is clear that use of such a measure would not be the first line of approach to assess of patients with presumed severe cognitive disabilities (e.g. patients in a persistent vegetative state (PVS) or 'locked in' syndrome). The ERP-DB task is seen as having the most relevance for patients with demonstrated language comprehension abilities who suffer from expressive (verbal and behavioral) communication impairments that prevent them for being accurately assessed using traditional measures. This task may be useful in PVS or 'locked-in' patients, for example, but only after successful recordings have been made using simpler ERP paradigms (e.g. a tone-based odd-ball or MMN-type task, followed by basic ERP language comprehension tasks) that demonstrate the existence of requisite sensory, attentional, and language abilities.

Moreover, the task may need further modification to increase its clinical utility for assessing non-communicative patients. For example, the length and concentration demands of the ERP-DB task could be modified on a patient-by-patient basis (e.g. by presenting only the lower set sizes) which, if successful, could be followed by inclusion of the higher set sizes at a later date.

Recently, innovative research aimed at developing brain-computer interfaces (BCI) to allow patients with 'locked-in' syndrome or amyotrophic lateral sclerosis (ALS) to communication has shown promise (e.g. Birbaumer et al., 1999, 2000; Kübler et al., 2001 for review). Currently, the use of this technique involves several months of training and substantial cognitive demands on the patients. The use of initial assessment measures that tap into complex cognitive functioning, such as the ERP-DB task or other elements of our research program, may be of great benefit in determining which patients may be most appropriate for this type of intensive investigation.

Overall, this research supports the clinical utility of the ERP-DB task as a means of assessing working memory abilities. Nevertheless, before this test could be applied clinically, future research is required to replicate these results and further investigate the strength and reliability of the relationship between the PSW, dPSW, and P200/P300 components and working memory capacity. For example, administration of all the set sizes ranging from 2 to 8 digits to participants (rather than the division of the task into the two sections) may overcome some of the restricted range difficulties encountered in the present study and produce a pattern that strengthens the association between the PSW or dPSW and working memory capacity. Lastly, the current sample was quite homogenous in terms of age, education and DB performance. The ERP-DB task should be assessed in other groups characterized by greater heterogeneity in performance to enable the results to be generalized to a wider range of populations.

Acknowledgements

This work was supported by grants from the Scottish Rite Charitable Foundation of Canada, the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council of Canada (NSERC), and Fonds de la Recherche en Santé du Québec – Fonds pour la formation de Chercheurs et l'Aide à la Recherche (FRSQ – FCAR). We would like to thank Paula Miles for administrative assistance and The Psychological Corporation for permission to computer-adapt the Digit Span Backward subtest from the Wechsler Adult Intelligence Scale: Third Edition.

References

Allen JJ, Iacono WG, Danielson KD. The identification of concealed memories using the event-related potential and implicit behavioral measures: a methodology for prediction in the face of individual differences. Psychophysiology 1992;29(5):504–22.

Baddeley AD, Logie RH. Working memory: the multiple component model. In: Miyake A, Shal P, editors. Models of working memory: mechanisms of active maintenance and executive control. New York: Cambridge; 1999.

Baldwin KB, Kutas M. An ERP analysis of implicit structured sequence learning. Psychophysiology 1997;34(1):74–86.

Barceló F, Sanz M, Molina V, Rubia FJ. The Wisconsin card sorting test and the assessment of frontal function: a validation study with event-related potentials. Neuropsychologia 1997;35(4):399–408.

Bennington JY, Polich J. Comparison of P300 from passive and active tasks for auditory and visual stimuli. Int J Psychophysiol 1999;34(2):171–7.

Birbaumer N, Ghanayim N, Hinterberger T, Iversen I, Kotchoubey B, Kübler A, Perelmouter J, Taub E, Flor H. A spelling device for the paralysed. Nature 1999;398(6725):297–8.

Birbaumer N, Kübler A, Ghanayim N, Hinterberger T, Perelmouter J, Kaiser J, Iversen I, Kotchoubey B, Neumann N, Flor H. The thought translation device (TTD) for completely paralyzed patients. IEEE Trans Rehabil Eng 2000;8(2):190–3.

Blumhardt LD. The clinical utility of event-related potentials to a short-term memory task. Electroenceph Clin Neurophysiol Suppl 1996;46: 73–84.

Boller F, Vignolo LA. Latent sensory aphasia in hemisphere-damaged patients: an experimental study with the token test. Brain 1966;89(4): 815–30.

Byrne JM, Dywan CA, Connolly JF. Assessment of children's receptive vocabulary using event-related brain potentials: development of a clinically valid test. Child Neuropsychol 1995a;1(3):211–23.

Byrne JM, Dywan CA, Connolly JF. An innovative method to assess the receptive vocabulary of children with cerebral palsy using event-related brain potentials. J Clin Exp Neuropsychol 1995b;17(1):9–19.

Byrne JM, Connolly JF, MacLean SE, Dooley JM, Gordon KE, Beattie TL. Brain activity and language assessment using event-related potentials: development of a clinical protocol. Dev Med Child Neurol 1999;41: 740–7.

Chao LL, Knight RT. Prefrontal and posterior cortical activation during auditory working memory. Brain Res Cogn Brain Res 1996;4(1): 27–37.

Coles MG, Scheffers MK, Holroyd CB. Why is there an ERN/Ne on correct trials? Response representations, stimulus-related components, and the theory of error-processing Biol Psychol 2001;56(3):173–89.

Connolly JF, D'Arcy RCN. Innovations in neuropsychological assessment using event-related brain potentials. Int J Psychophysiol 2000;37(1): 31–47.

Connolly JF, Byrne JM, Dywan CA. Assessing adult receptive vocabulary with event-related potentials: an investigation of cross-modal and cross-form priming. J Clin Exp Neuropsychol 1995;17(4):548–65.

Connolly JF, Major A, Allen S, D'Arcy RCN. Performance on WISC-III and WAIS-R NI vocabulary subtests assessed with event-related brain potentials: an innovative method of assessment. J Clin Exp Neuropsychol 1999a;21(4):444–64.

Connolly JF, Mate-Kole CC, Joyce BM. Global aphasia: an innovative assessment approach. Arch Phys Med Rehabil 1999b;80(10):1309–15.

Connolly JF, D'Arcy RCN, Newman RL, Kemps R. The application of cognitive event-related brain potentials (ERPs) in language-impaired individuals: review and case studies. Int J Psychophysiol 2000;38(1): 55–70.

D'Arcy RCN, Connolly JF. An event-related brain potential study of receptive speech comprehension using a modified token test. Neuropsychologia 1999;37(13):1477–89.

D'Arcy RCN, Connolly JF, Eskes GA. Evaluation of reading comprehension with neuropsychological and event-related brain potential (ERP) methods. J Int Neuropsychol Soc 2000;6(5):556–67.

D'Arcy RCN, Marchand Y, Eskes GA, Harrison ER, Phillips SJ, Major A, Connolly JF. Electrophysiological assessment of language function following stroke. Clin Neurophysiol 2003;114(4):662–72.

deGroot MH, Phillips SJ, Eskes GA. Fatigue associated with stroke and other neurological conditions: implications for stroke rehabilitation. Arch Phys Med Rehabil 2003;84:1714–20.

de Jong PF, Das-Smaal EA. Attention and intelligence: the validity of the star counting test. J Educ Psychol 1995;87(1):80–92.

De Renzi A, Vignolo LA. Token test: a sensitive test to detect receptive disturbances in aphasics. Brain 1964;85:665–78.

Dunn LM, Dunn LM. Peabody picture vocabulary test-revised. Minnesota: American Guidance; 1981.

Eimer M, Goschke T, Schlaghecken F, Stürmer B. Explicit and implicit learning of event sequences: evidence from event-related brain potentials. J Exp Psychol Learn Mem Cogn 1996;22(4):970–87.

Ellwanger J, Rosenfeld JP, Sweet JJ, Bhatt M. Detecting simulated amnesia for autobiographical and recently learned information using the P300 event-related potential. Int J Psychophysiol 1996;23(1–2):9–23.

Fabiani M, Karis D, Donchin E. Effects of strategy manipulation on P300 amplitude in a von Restoff paradigm. Psychophysiology 1985;22: 588–9.

Fabiani M, Karis D, Donchin E. P300 and recall in an incidental memory paradigm. Psychophysiology 1986;23(3):298–308.

Ford JM, Roth WT, Mohs RC, Hopkins III WF, Kopell BS. Event-related potentials recorded from young and old adults during a memory retrieval task. Electroencephalogr Clin Neurophysiol 1979;47(4): 450–9.

Frankel F, Tymckuk AJ. Digit recall of mentally retarded and nonretarded children under three presentation rates. Am J of Ment Deficiency 1974; 79:311–9.

Fry AF, Hale S. Relationships among processing speed, working memory, and fluid intelligence in children. Biol Psychol 2000;54(1–3):1–34.

García-Larrea L, Cézanne-Bert G. P3, positive slow wave and working memory load: a study on the functional correlates of slow wave activity. Electroenceph Clin Neurophysiol 1998;108(3):260–73.

García-Larrea L, Lukaszewicz AC, Mauguière F. Revisiting the oddball paradigm. Non-target vs neutral stimuli and the evaluation of ERP attentional effects. Neuropsychologia 1992;30(8):723–41.

Gardner RA. Digits forward and digits backward as two separate tests: normative data on 1567 school children. J Clin Child Psychol 1981; 10(2):131–5.

Gioia GA, Isquith PK. Ecological assessment of executive function in traumatic brain injury. Dev Neuropsychol 2004;25(1–2):135–58.

Golob EJ, Starr A. Serial position effects in auditory event-related potentials during working memory retrieval. J Cogn Neurosci 2004; 16(1):40–52.

Greenhouse SW, Geisser S. On methods in the analysis of profile data. Psychometrika 1959;24:95–112.

Grippo A, Pelosi L, Mehta V, Blumhardt LD. Working memory in temporal lobe epilepsy: an event-related potential study. Electroencephalogr Clin Neurophysiol 1996;99(3):200–13.

Holamon B, Morris G, Retzlaff P. Event-related potentials during delayed recognition of Wechsler memory scale-R paired associate learning. J Clin Psychol 1995;51(3):391–5.

Howard L, Polich J. P300 latency and memory span development. Dev Psychol 1985;21(2):283–9.

Jasper HH. The ten-twenty electrode system of the international federation. Electroencephalogr Clin Neurophysiol 1958;10:371–5.

Johnson RJ. A triarchic model of P300 amplitude. Psychophysiology 1986; 23(4):367–84.

Johnson RJ, Donchin E. Second thoughts: multiple P300s elicited by a single stimulus. Psychophysiology 1985;22(2):182–94.

Kaplan E, Fein D, Morris R, Delis DC. The WAIS-R as a neuropsychological instrument. New York: The Psychological Corporation; 1991.

Kay J, Lesser R, Coltheart M. Psycholinguistic assessments of language processing in aphasia (PALPA) test. Hove, UK: Lawerance Erlbaum Associates; 1992.

Kok A. Event-related-potential (ERP) reflections of mental resources: a review and synthesis. Biol Psychol 1997;45(1–3):19–56.

Kotchoubey BI, Jordan JS, Grözinger B, Westphal KP, Kornhuber HH. Event-related brain potentials in a varied-set memory search task: a reconsideration. Psychophysiology 1996;33(5):530–40.

Kotchoubey B, Lang S, Baales R, Herb E, Maurer P, Mezger G, Schmalohr D, Bostanov V, Birbaumer N. Brain potentials in human patients with extremely severe diffuse brain damage. Neurosci Lett 2001;301(1):37–40.

Kübler A, Kotchoubey B, Kaiser J, Wolpaw JR, Birbaumer N. Brain-computer communication: unlocking the locked in. Psychol Bull 2001; 127(3):358–75.

Kusak G, Grune K, Hagendorf H, Metz AM. Updating of working memory in a running memory task: an event-related potential study. Int J Psychophysiol 2000;39(1):51–65.

Kutas M, Hillyard SA. Brain potentials during reading reflect word expectancy and semantic association. Nature 1984;307(5947):161–3.

Kyllonen PC. Theory based cognitive assessment. In: Zeidner J, editor. Human productivity enhancement: organizations, personnel and decision making, vol. 2. New York: Praeger; 1987. p. 338–81.

Kyllonen PC, Christal RE. Cognitive modeling of learning abilities: a status report of LAMP. In: Dillon R, Pelligrino JW, editors. Testing: theoretical and applied issues. New York: Freeman; 1987.

Kyllonen PC, Christal RE. Reasoning ability is (little more than) working memory capacity? Intelligence 1990;14:389–433.

Lang S, Kotchoubey B. Brain responses to number sequences with and without active task requirement. Clin Neurophysiol 2002;113(11): 1734–41.

Lezak MD. Neuropsychological assessment. 3rd ed. New York: Oxford University Press; 1995.

Marchand Y, D'Arcy RCN, Connolly JF. Linking neurophysiological and neuropsychological measures for aphasia assessment. Clin Neurophysiol 2002;113(11):1715–22.

Miller E. A case for automated clinical testing. Bull Br Psychol Soc 1968; 21:75–8.

Mishra SP, Ferguson BA, King PV. Research with the Wechsler digit span subtest: implications for assessment. Sch Psychol Rev 1985;14(1): 37–47.

Morse PA, Montgomery CE. Clinical syndromes in adult neuropsychology: the practitioner's handbook. Amsterdam: Elsevier; 1992.

Nittono H, Nageishi Y, Nakajima Y, Ullsperger P. Event-related potential correlates of individual differences in working memory capacity. Psychophysiology 1999;36(6):745–54.

Patterson JV, Pratt H, Starr A. Event-related potential correlates of the serial position effect in short-term memory. Electroenceph Clin Neurophysiol 1991;78:424–37.

Pedersen PM, Jørgensen HS, Nakayama G, Raaschou HO, Olsen TS. Aphasia in acute stroke: incidence, determinants and recovery. Ann Neurol 1995;38:659–66.

Pelosi L, Holly M, Slade T, Hayward M, Barrett G, Blumhardt LD. Wave form variations in auditory event-related potentials evoked by a memory-scanning task and their relationship with tests of intellectual function. Electroenceph Clin Neurophysiol 1992;84(4):344–52.

Pelosi L, Hayward M, Blumhardt LD. Is 'memory-scanning' time in the Sternberg paradigm reflected in the latency of event-related potentials? Electroencephalogr Clin Neurophysiol 1995;96(1):44–55.

Pelosi L, Hayward M, Blumhardt LD. Which event-related potentials reflect memory processing in a digit-probe identification task? Brain Res Cogn Brain Res 1998;6(3):205–18.

Polich J. N400s from sentences, semantic categories, number and letter strings? Bull Psychon Soc 1985;23(4):361–4.

Polich J, Kok A. Cognitive and biological determinants of P300: an integrative review. Biol Psychol 1995;41(2):103–46.

Polich J, Howard L, Starr A. P300 latency correlates with digit span. Psychophysiology 1983;20(6):665–9.

Pratt H, Michalewski HJ, Barrett G, Starr A. Brain potentials in a memory-scanning task. I. Modality and task effects on potentials to the probes. Electroencephalogr Clin Neurophysiol 1989;72(5):407–21.

Reinvang I. Cognitive event-related potentials in neuropsychological assessment. Neuropsychol Rev 1999;9(4):231–48.

Retzlaff PD, Morris GL. Event-related potentials during the continuous visual memory test. J Clin Psychol 1996;52(1):43–7.

Ruchkin DS, Sutton S, Kietzman ML, Silver K. Slow wave and P300 in signal detection. Electroencephalogr Clin Neurophysiol 1980;50(1–2): 35–47.

Ruchkin DS, Johnson Jr R, Canoune H, Ritter W. Short-term memory storage and retention: an event-related brain potential study. Electroenceph Clin Neurophysiol 1990;76(5):419–39.

Rugg MD. ERP studies of memory. Electrophysiology of mind: event-related brain potentials and cognition. In: Rugg MD, Coles MGH, editors.. London: Oxford University Press; 1995. p. 133–70.

Sattler JM. Assessment of children, revised and updated third edition. San Diego: Jermone M. Sattler, Publisher, Inc; 1992 p. 154.

Schlaghecken F, Stürmer B, Eimer M. Chunking processes in the learning of event sequences: electrophysiological indicators. Mem Cognit 2000; 28(5):821–31.

Schwartz L, Taylor HG, Drotar D, Yeates KO, Wade SL, Stancin T. Long-term behavior problems following pediatric traumatic brain injury: prevalence, predictors, and correlates. J Pediatr Psychol 2003;28(4): 251–63.

Smith ME, Guster K. Decomposition of recognition memory event-related potentials yields target, repetition, and retrieval effects. Electroencephalogr Clin Neurophysiol 1993;86(5):335–43.

Sohlberg MM, Mateer CA. Cognitive rehabilitation: an integrative neuropsychological approach. New York, NY: Guilford Press; 2001.

Squires KC, Wickens C, Squires NK, Donchin E. The effect of stimulus sequence on the waveform of the cortical event-related potential. Science 1976;193(4258):1142–6.

Starr A, Barrett G. Disordered auditory short-term memory in man and event-related potentials. Brain 1987;110(Pt 4):935–59.

Sternberg S. High-speed scanning in human memory. Science 1966;153: 652–4.

Sternberg S. Factor theories of intelligence are all right almost. Educ Res 1980;9:6–18.

Thompson JA, Wilson SL. Automated psychological testing. Int J Man Machine Studies 1982;17:279–89.

Verleger R. On the utility of P3 latency as an index of mental chronometry. Psychophysiology 1997;34(2):131–56.

Wade DT, Hewer RL, David RM, Enderby PM. Aphasia after stroke: natural history and associated deficits. J Neurol Neurosurg Psychiatry 1986;49(1):11–16.

Warschaudky SD, Kewman DG, Selim A. Attentional performance of children with traumatic brain injury: a quantitative and qualitative analysis of digit span. Arch of Clin Neuropsych 1996;11(2):147–53.

Wechsler D. Wechsler intelligence scale for children-III. San Antonio: The Psychological Corporation. Harcourt Brace and Company; 1991.

Wechsler D. WAIS-III administration and scoring manual. San Antonio: The Psychological Corporation. Harcourt Brace and Company; 1997.

Wechsler D. The WAIS-III and WMS-III technical manual. San Antonio: The Psychological Corporation. Harcourt, Brace and Company; 1997.

Wechsler D. WAIS-III Canadian Technical Manual. Toronto: The Psychological Corporation. Harcourt Brace and Company; 2001.

Woltz DJ. An investigation of the role of working memory in procedural skill acquisition. J Exp Psychol Gen 1988;117(3):319–31.