Intern. J. Neuroscience, 115:1605–1630, 2005
Copyright © 2005 Taylor & Francis Inc.
ISSN: 0020-7454 / 1543-5245 online
DOI: 10.1080/00207450590958475

THE TEST-RETEST RELIABILITY OF A STANDARDIZED NEUROCOGNITIVE AND NEUROPHYSIOLOGICAL TEST BATTERY: "NEUROMARKER"

L. M. WILLIAMS

The Brain Dynamics Centre
Westmead Hospital
NSW, Australia

E. SIMMS

The Brain Dynamics Centre
Westmead Hospital
NSW, Australia

and

School of Psychology
University of Sydney
NSW, Australia

C. R. CLARK

Flinders University
Adelaide, South Australia, Australia

R. H. PAUL

Brown Medical School
Department of Psychiatry
Centers for Behavioral and Preventive Medicine
Providence, Rhode Island, USA

Received 1 December 2004.

Address correspondence to L. M. Williams, The Brain Dynamics Centre, Westmead Hospital, NSW 2145, Australia. E-mail: lea@psych.usyd.edu.au

D. ROWE
E. GORDON

The Brain Dynamics Centre
Westmead Hospital
NSW, Australia

and

Brain Resource International Database
Brain Resource Company
Ultimo, NSW, Australia.

NeuroMarker combines EEG and ERP measures with neurocognitive tests in a fully computerized and standardized testing system. It is designed for use across the lifespan and has a large normative database of over 1,000 subjects. This study was a preliminary evaluation of "NeuroMarker" in subjects spanning four decades. Twenty-one healthy subjects (12–57 years) were tested at baseline and four weeks later. From the "Neuromarker" battery, the authors analyzed EEG data (eyes open and closed) and ERPs elicited during auditory oddball (N100, P200, N200, P300) and working memory (P150, P300) tasks. Concomitant neuropsychological data, acquired using a touch-screen system, comprised measures of sensori-motor, attention, verbal, executive, and memory function. Test-retest data were examined using analyses of variance and correlational procedures (corrected for multiple comparisons), with parallel analyses of age. EEG data did not differ across sessions, and showed high test-retest reliability (.71–.95), particularly for theta and delta (>.85). ERP components also showed sound reliability, particularly for sites where components are maximal: fronto-central N100 (.76–.77), centro-parietal P300 (.78–.81) to oddball targets, N100 and P200 (.74–.86) to oddball non-targets, and P150 amplitude and latency (.84–.93) to working memory stimuli. Neuropsychological tests showed a similarly sound level of consistency (on average, .70), with the most consistent tests tapping simple motor function, estimated intelligence, switching of attention (Part 2), verbal interference response time and memory intrusions (.71–.89). Age and sex did not have a differential impact on reliability for EEG, ERP, or neuropsychology measures. These findings provide preliminary evidence that the "NeuroMarker" battery is reliable for test-retest assessments. The results suggest that the standardized approach has utility for providing sensitive clinical and treatment evaluations across age groups.

Keywords EEG, ERP, reliability, neuropsychology, oddball test-retest

INTRODUCTION

Neurophysiological and neuropsychological measures provide complementary information for screening and clinical assessment, monitoring of health and disease, and treatment evaluation (Dikmen et al., 1999; Kinoshita et al., 1996). Neuropsychological assessment is recommended for evaluating cognitive impairment (Neurology Practice Parameters; Petersen et al., 2001), and is a valuable tool in the development and evaluation of new medications. Among neurophysiological measures, the electroencephalograph (EEG) is recommended for routine assessment of neurological conditions such as epilepsy, dementia, and cerebrovascular disease, and for critical care monitoring and presurgical analysis (Nuwer, 1997). Evoked potentials have typically been used by clinicians to evaluate sensory processes (Chiappa, 1997). A large pool of evidence has demonstrated the sensitivity of event-related potentials (ERPs) to cognitive processes, with the P300 component showing particular sensitivity to aspects of attention and memory (Polich & Kok, 1995). Although the application of EEG and ERPs in neuropsychiatric evaluation is largely investigatory (Nuwer, 1997), ERPs show promise as a tool for cognitive assessment (Connolly & D'Arcy, 2000).

To date, limitations to the utility of EEG and ERPs in clinical contexts include the lack of standardized testing protocols, large normative databases for comparison (Thatcher, 1998) and the evaluation of psychometric properties such as test-retest reliability. The development of neuropsychological test batteries has addressed these features in more detail, facilitating their use in clinical evaluations, yet these batteries are still constrained by the lack of large standardized databases that cover the range of core cognitive domains and testing formats appropriate for use across the lifespan (for review see Lezak, 1995, Spreen & Strauss, 1988).

This study used a new testing system, "NeuroMarker," which integrates a comprehensive and fully automatic neuropsychological test battery with EEG measures and tasks known to elicit robust cognitive ERPs (Gordon, 2003a,b). NeuroMarker uses fully standardized testing protocols for both hardware and software, and includes a large normative database of over 1,000 subjects spanning nine decades and screened rigorously for mental and physical health confounds. Touch-screen hardware, with instructions presented both visually and via audio files ensures independence from computer and keyboard skills, making the testing system appropriate for use in young and elderly groups, and in the cognitively impaired. This study undertook a preliminary evaluation of test-retest reliability for NeuroMarker, as part of a series of studies evaluating its psychometric properties, and establishing age and sex norms for the associated database (Paul et al., 2005; Clark et al., 2005).

Neuropsychological tests have been shown to have good to high test-retest reliability (.70–.90). One exception has been the memory domain, where lower reliability coefficients have been observed, even over short intervals of less than 12 months (Dikmen et al., 1999). These findings have typically been interpreted as an inconsistent practice effect, reflecting variation in initial scores and demographic factors. Subjects who initially score poorly tend to show a much smaller cognitive improvement than those who initially performed well (Dikmen et al., 1999). Age, sex, and years of education each impact on verbal memory and learning (Bolla-Wilson & Bleecker, 1986), and younger and better-educated subjects, with strong test performance, may benefit more from practice (Ushiyama et al., 1995). These practice effects are usually minimized with longer testing intervals, although they have been observed at intervals of 12 months (Ushiyama et al., 1995). Executive functions, reflected in "strategic learning," and variations in test item novelty may also contribute to inconsistency in memory (Dikmen et al., 1999). By contrast, measures of sensori-motor function, attention, and verbal and performance ability tend to have consistent practice effects and thus, a minimal effect on reliability over time (Ushiyama et al., 1995; Watson et al., 1994).

Studies to date suggest that the EEG is stable over periods of one day to 18 weeks, particularly for the theta, alpha, and beta bands. Similarly robust reproducibility has been observed across these bands for resting eyes closed and eyes open conditions at 17–18 weeks retest in adult samples (Pollock et al., 1991). The highest reliability has been observed for theta (0.80) and alpha (0.84) (Gasser et al., 1985; Pollock et al., 1991). Indeed, alpha has been reported as the most stable EEG feature at both 5-min and 12–16-week intervals (Salinsky et al., 1991), and across recording sites (Burgess & Gruzelier, 1997).

A number of studies have reported changes in EEG activity with changes in cognitive performance or state. Theta activity has been found to increase with better performance on memory tasks tapping encoding, working memory, and retrieval (Klimesch, 1997; 2001; 2004). Increases in frontal midline theta, with decreases in alpha, have also been observed with increasing task difficulty (e.g., Gevins, 1979; 1997), whereas increases in diffuse theta, with decreases in fast posterior alpha have been observed with fatigue (Gevins et al., 1977; Makeig & Jung, 1995). Notably, however, task-related EEG has shown good test-retest reliability, suggesting it is appropriate for evaluating clinical changes in cognitive status. For instance, in a study of 20 young adults (18–29 years), McEvoy et al. (2000) demonstrated that EEG alpha and theta activity was highly reliable both within conditions (1 h apart) and between sessions (1 week apart) for tasks tapping vigilance (correlations >.8) and working memory at two difficulty levels (correlations >.9).

Although there have been relatively few studies of ERP reliability, the existing reports suggest that ERPs are generally less reproducible than the EEG. Most reliability studies have used the auditory oddball task, and the importance of using a standard task for comparative purposes has been highlighted, given that ERPs vary with task conditions (Segalowitz & Barnes, 1993). Studies of young adults (18–31 years) have reported stable ERP amplitudes within session (Polich, 1986) and over relatively short periods of one to two weeks (Fabiani et al., 1987; Maeda et al., 1995). Similarly sound reliability has been observed in adolescents (14–16 years) over longer periods of one to two years (Segalowitz & Barnes, 1993). In older adults (60–86 years), the N1, P2, and P3 have shown high reliabilities (averaging around .75) over these longer periods (Sandman & Patterson, 2000). ERP latency has generally demonstrated lower reliability than amplitude, although P300 latency has shown reproducibility across one to four week intervals, even with changes in arousal level (Fabiani et al., 1987; Polich, 1986; Segalowitz & Barnes, 1993; Waldhovd & Fjell, 2002).

Although these studies collectively encompass a wide range of ages, to the authors' knowledge only one study to date has examined reliability in a sample covering the adult life span (21–92 years) (Waldhovd & Fjell, 2002). In this study, reliabilities for N1, P2, and P3 amplitude were significant over 2 test-retest intervals, ranging from .43 to .89 at 12 months and .58 to .67 at 24 months (Waldhovd & Fjell, 2002). Analysis of ERP data across consecutive 50 ms intervals showed that reliabilities peaked around the windows of these conventional components, providing a validation of the N1, P2, and P3. Reliability also tends to follow the topographical distribution of these components, and is greatest where they are maximal (Waldhovd & Fjell, 2002).

The consistency of auditory oddball ERPs across a number of repeat testing sessions has been demonstrated in a sample of 10 adults (29–57 years) (Kinoshita et al., 1996). Over eight test-retest sessions (7–10 days apart), the reliability for P300 amplitude ranged from .54 to .68 and for P300 latency, from .12 to .46. N100 and N200 were also found to be stable across sessions (Kinoshita et al., 1996).

To date, test-retest reliability studies have focused on a particular measure (such as EEG or ERP or neuropsychological tests), rather than the concurrent reproducibility of these "brain-behaviour" measures in the same subjects.

These studies also typically include samples with a restricted range of ages. The aim of this study was to determine the test-retest reliability of the "NeuroMarker" battery of integrated EEG/ERP and neuropsychological measures in a sample encompassing adolescence to later adulthood.

METHODS

Subjects

The present study makes use of data acquired as part of the Brain Resource International Database (BRID; http://www.brainresource.com). Twenty-one healthy individuals undertook recordings at both test and retest four weeks later. These subjects were aged 12 to 57 years (mean = 27.76 years), with years of education 7 to 15 years (mean = 12.14), and comprised 10 females matched to the 11 males. A complementary study had demonstrated that years of education has little impact on the "NeuroMarker" measures (Clark et al., 2005). Exclusion criteria included a history of mental illness, substance abuse, neurological disorder and brain injury, other serious medical condition, and left-handedness. The SPHERE questionnaire (Hickie et al., 1998) was used to screen out individuals with a potential Axis I disorder (APA, 1994), and subjects were also excluded if they reported a first-degree family member with an Axis I diagnosis.

Procedure

Subjects gave written informed consent, in accordance with medical research council guidelines.

They were asked to refrain from smoking and caffeine for 2 h prior to testing, and to refrain from alcohol for 12 h prior to testing. EEG/ERP and cognitive testing was undertaken within one testing session. The re-test session followed the initial testing session after a tightly controlled period of four weeks for each subject.

EEG Testing Protocol

All participants were seated in a comfortable chair in a dimly lit room. EEG data were recorded from 26 scalp electrode sites according to the NuAmps International 10–20 electrode system using a Quikcap with sintered Ag/AgCl electrodes, and a continuous acquisition system, with a sample rate of 500 Hz.

The study focused on midline site (Fz, Cz, Pz) data. These data were recorded relative to the virtual ground, but referenced offline to linked mastoids. Horizontal eye movements were recorded with electrodes placed 1.5 cm lateral to the outer canthus of each eye. Vertical eye movements were recorded with electrodes placed 3 mm above the middle of the left eyebrow and 1.5 cm below the middle of the left bottom eye-lid. Skin resistance was <5 kOhms. Data were EOG corrected offline using the Gratton et al. (1983) procedure.

EEG data were acquired under two conditions: resting quietly with their eyes closed and then with their eyes open (2 min each).

ERP Testing

ERP data were acquired using the NuAmps EEG system. ERPs were elicited in response to an auditory oddball task, providing a standard comparison to previous studies, and a continuous performance (working memory) task, to assess reliability across tasks.

In the auditory oddball task, subjects were presented with a series of high (1000 Hz) "task relevant" target tones and lower (500 Hz) "task irrelevant" tones of 50 ms duration (rise and fall time, 5 ms), at 75 dB, with an ISI of 1 s. Target probability was relatively low (12%), consistent with reports that reliability is maximized at lower target probabilities (Waldhovd & Fjell, 2002). The two tones were presented in a quasi-random order, with the only constraint being that two task-relevant tones cannot appear consecutively. Instructions were presented in a standardized format using audio files. Subjects were instructed to button press with the index finger of each hand (to counterbalance for possible motor effects) in response to target and to ignore the lower tones, and the speed and accuracy of response were equally stressed.

The working memory task consisted of a series of 125 letters (B, C, D, or G) presented for 200 ms on a computer screen, with an ISI of 2.5 s. Of the 125 letter stimuli, 20 were target letters (i.e., repetitions of the previous letter) and 85 were non-target letters (i.e., non-repeated letters). Via standardized audio file instructions, subjects were instructed to respond if the same letter appeared twice in a row (target letters), via button press with the index finger of each hand. Speed and accuracy of response were again equally stressed.

Neuropsychological Testing

The cognitive tests were part of a fully computerized and standardized battery (Gordon, 2003a,b); "IntegNeuro." This battery is presented on a touch-screen computer and does not rely on keyboard or computer skills, thereby making it readily applicable to subjects across a wide range of ages. The IntegNeuro test battery was administered in a sound-attenuated testing room, with participants seated in front of the touch-screen computer (NEC MultiSync LCD 1530V). Standardized task instructions were presented visually on the screen and using concurrent audio files (via headphones). Nonverbal responses (reaction time and accuracy) were recorded via the touch-screen computer and verbal responses were recorded via a microphone and recording system attached to the headphones. The test battery has well established convergent validity relative to conventional paper-and-pencil tests tapping equivalent domains of cognitive function, as well as divergent validity relative to tests tapping distinct domains of function (Paul et al., 2005). Correlations between the IntegNeuro tests and equivalent measures (such as the WAIS-III digit span tasks, California Verbal Learning Test, Weschler Memory Scale III, Trail Making Tests, Rey Complex Figure test, Verbal Fluency, and Finger tapping) have been found to be significant for each test across younger and older subjects, with correlations ranging in magnitude from .53 to .77 (Paul et al., 2005).

The cognitive tests, indexing sensori-motor, attention, executive function and estimated intelligence, language, and memory domains of function, were as follows:

Sensori-Motor

Simple Motor Tapping Task. Participants were required to tap a circle on the touch-screen with their index finger, as fast as possible for 60 s. The dependent variable was total number of taps with the dominant hand.

Choice Reaction Time Task. Participants attended to the computer screen as one of four target circles was illuminated in a pseudorandom sequence. For each trial, the subject used their hand of preference (right hand in each case) to touch the illuminated circle as quickly as possible following presentation. Twenty trials were administered with a random delay between trials of 2–4 s. The dependent variable was the mean reaction time across trials.

Attention

Span of Visual Memory. This provides a computerized adaptation of the Corsi Blocks task (Milner, 1970) and the Dot Location task variant (Roth & Crosson, 1985). It represents a visual analogue of a digit span test. Participants were presented with squares arranged in a random pattern on the computer screen.

The squares were highlighted in a sequential order on each trial. Participants were required to repeat the order in which the squares were highlighted. Both forward and reverse trials were conducted. The total correct was the dependent variable.

Digit Span. Participants were presented with a series of digits on the touchscreen (e.g., 4, 2, 7, etc., 500 ms presentation), with a one second interstimulus interval. Subjects were asked to enter the digits immediately on a numeric keypad on the touch-screen. In Part 1, subjects were required to recall the digits in forward order (Digit Span Forwards); in Part 2, they were required to recall them in reverse order (Digit Span Reversed). In each part, the number of digits in each sequence was gradually increased from 3 to 9, with 2 sequences at each level. Performance was indexed by the maximum number of digits recalled without error.

Switching of Attention Task; Part 1. This test provides a computerized adaptation equivalent to the Trail Making Test (Reitan, 1958). In Part 1, subjects are presented with a pattern of 25 numbers in circles and asked to touch them in ascending numerical sequence (i.e., 1 2 3 . . .). As each number is touched in correct order, a line is drawn automatically to connect it to the preceding number, allowing subject to visualize the path. Performance in terms of psychomotor speed and the basic ability to hold attention on a simple task is indexed by time to completion and number of errors.

Executive Function

Switching of Attention Task; Part 2. In Part 2, subjects are presented with a pattern of 13 numbers (1–13) and 12 letters (A–L) and required to touch numbers and letters alternatively in ascending sequence (i.e., 1 A 2 B 3 C . . .). The additional demands of Part 2 reflect the requirement to switch attention between the respective tasks and mental sets induced, indexed by time to completion.

Verbal Interference. This task taps the ability to inhibit automatic and irrelevant responses and is similar to the Stroop task (Golden, 1978). Subjects were presented with colored words with incongruent color-word combinations. The words were presented serially, one at a time, drawn from the following set of lower case words: red, yellow, green, and blue. The color of each word is drawn from the following set of colors: red, yellow, green, and blue. In Part 1, the subject was required to identify the name of each word as quickly as possible after it was presented on the screen (using a response pad with the four possible words displayed in black and in fixed format). In Part

2, the subject was required to name the color of each word as quickly as possible (using the response pad). Performance was indexed in terms of reaction time and number of errors.

Language/Verbal Function

Letter Fluency. Participants were required to generate by speech, words that began with the letters F, A, and S, allowing 60 s for each letter and excluding proper nouns. Fluency was indexed by the total number of correct words generated across the three trials.

Animal Fluency. Participants were required to name animals as quickly as possible for 60 s, indexed similarly by total number correct.

Memory

Verbal List-Learning. Participants were required to read a list of 12 English words, and asked to memorize and recall them verbally. The first list was presented and recalled on 4 trials to test immediate recall and learning. For trial 5, a new list was presented and recalled only once, to act as a distractor list. In trial 6 the original list was presented once again and recalled. Following a 30-min delay filled with distractor tasks, participants were required to recall as many words as possible from the original list (delayed recall). Finally, participants were required to discriminate the 12 target words from 12 foils on a recognition trial (recognition recall).

Words were closely matched on concreteness, number of letters, and frequency. The subject was asked to recall (verbally) the word lists after immediate and delayed recall time intervals. For each trial, scores represented the number of words correctly recalled within 30 s. Memory performance was indexed by the words correctly recalled across the four learning trials, the immediate recall trial and the delayed recall trial, and the total number of correctly identified word on the recognition trial.

Estimated Intelligence Domain

Spot-the-Real-Word Task. This task is a computerized adaptation of the Spot the Real Word test (Baddeley et al., 1993). On each trial, a valid English word was presented simultaneously with a non-word foil. The subject was required to select (via the touchscreen) that was the valid word. The number of words correctly recognized was the dependent measure.

To estimate intelligence, the total correct score was entered into a regression formula that factored education, and age to render an estimated intelligence quotient (Sullivan et al., 2000). This estimate has been found to be highly correlated ($r = .8$) with the WAIS test of intellectual ability (Paul et al., 2005) and is a good indicator of premorbid IQ.

EEG DATA SCORING AND REDUCTION

Average power spectra were computed for eyes closed and eyes open conditions. For each condition, the two minutes of EEG were divided into adjacent intervals of four seconds. Power spectral analysis was performed on each interval by first applying a Welch window to the data, and then performing a Fast Fourier Transform (FFT) to provide an average power spectrum for each midline electrode site. For each spectrum, the power was calculated for the following four frequency bands: delta (1.5–3.5 Hz) theta (4–7.5 Hz), alpha (8–13 Hz) and beta (14.5–30 Hz).

ERP Data Scoring and Reduction

ERP data were extracted for target stimuli in the auditory oddball task, and non-target stimuli in the oddball and working memory task, using a window from –300 ms to 700 ms (relative to the stimulus onset). Data were filtered with a 25 Hz low-pass filter, and baselined relative to the 300 ms pre-stimulus window. Peak components for the auditory oddball task were determined according to the maximal response within latency windows shown previously to provide the key components for this task (Bahramali et al., 1999; Lim et al., 1999; Williams et al., 2000). The relevant components and windows, elicited by target stimuli in the oddball task, included: N100 (80–140 ms), P200 (140–270 ms), N200 (180–320 ms), P300 (270–550 ms). For non-target stimuli, the N100 and P200 were focused on. A parallel procedure was applied to the working memory paradigm, with a corresponding focus on the P150 (115–190 ms) and P300 (285–600 ms) of background (non-target) stimuli shown previously to provide the key components for the updating of visuoverbal working memory (Clark et al., 1998).

Neuropsychological Data Scoring and Reduction

For each test within each domain, the relevant performance scores were extracted.

Data Analyses

Consistent with previous studies (e.g., McEvoy et al., 2000), test-retest effects were analyzed in two ways: via repeated measures analysis of variance (ANOVA) to examine any differences across sessions and via correlational procedures to examine the degree of shared association. Whereas ANOVA may highlight any group practice effects (regardless of variation across individuals), correlations highlight the stability of individual scores.

In ANOVAs, the within-subjects factor was session (with repeated measures, session 1 and session 2) for EEG/ERP and neuropsychological measures. For EEG/ERP data, an additional within-subjects factor of recording site (midline Fz,Cz,Pz) was included. Age and sex were included as covariates in these analyses, providing a means to examine any differential impact of these factors on the test-retest reliability of dependent measures.

Pearson product moment correlation analyses were then undertaken to examine associations between session 1 and session 2 for each measure. Parallel partial correlation analyses were also used to control for any systematic effects of age and sex.

RESULTS

EEG data

Figure 1A–F depicts the stability of EEG alpha, beta, theta, and delta activity across test and retest sessions. There were no significant effects for session (test-retest) for EEG bands across conditions: eyes closed alpha ($F_{(1,18)} = .05$, $p = .83$), beta ($F_{(1,18)} = .77$, $p = .39$), theta ($F_{(1,18)} = .29$, $p = .60$) and delta ($F_{(1,18)} = .14$, $p = .71$), and eyes open alpha ($F_{(1,18)} = .12$, $p = .74$), beta ($F_{(1,18)} = .09$, $p = .77$), theta ($F_{(1,18)} = .05$, $p = .82$) and delta ($F_{(1,18)} = .01$, $p = .94$). There were also no significant interactions between session and age or sex for EEG bands, or between session and electrode site. Significant effects for site were observed for eyes closed alpha ($F_{(2,17)} = 5.45$, $p = .015$), due to increasing power from frontal to posterior midline sites, and for eyes closed beta ($F_{(2,17)} = 4.13$, $p = .035$), due to relatively greater power at centro-parietal sites (Figure 1A–C).[1]

---

[1] When age and sex were not included as covariates, the following session effects were observed: Theta power: Session effect ($f = 16.62$, $p = .001$); Alpha power: Session by midline interaction ($f = 3.87$, $df = 2.19$, $p = .039$), at the uncorrected level. This pattern of results suggests that any significant effects involving session are due to the contribution of demographic (age, sex) factors rather than to practice effects per se.

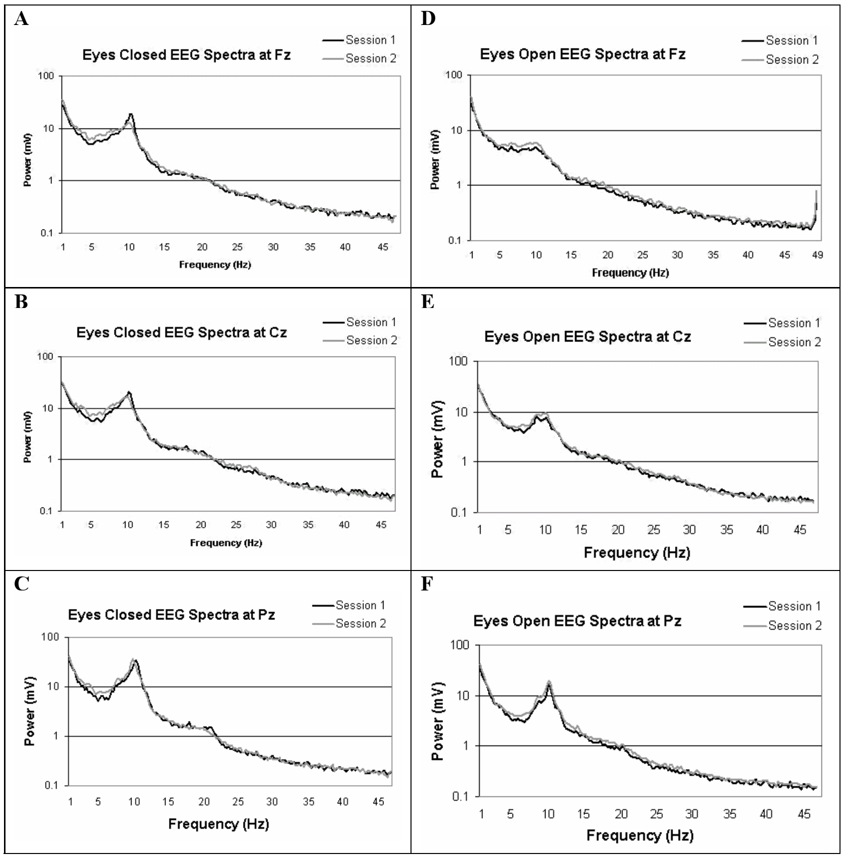

Figure 1. EEG spectra for the eyes closed condition across midline sites (A, Fz, B, Cz and C, Pz) and for the eyes open condition across midline sites (D, Fz, E, Cz, F, Pz).

ERP Data

Figure 2A–D depicts the ERP waveforms for auditory oddball targets and non-targets, and Figure 2E–F depicts ERP waveforms for working memory stimuli, across test and retest sessions.

There were no significant session effects for the amplitude of oddball target ERP components, N100 ($F_{(1,18)} = .004$, $p = .95$), P200 ($F_{(1,18)} = .88$, $p = .36$), N200 ($F_{(1,18)} = .75$, $p = .40$) and P300 ($F_{(1,18)} = .007$, $p = .93$). There were also no significant session effects for non-target ERPs, N100 ($F_{(1,18)} = .039$, $p = .85$) and P200 ($F_{(1,18)} = .039$, $p = .85$). Significant effects for site were observed for P300 amplitude ($F_{(2,17)} = 5.0$, $p = .02$), due its standard

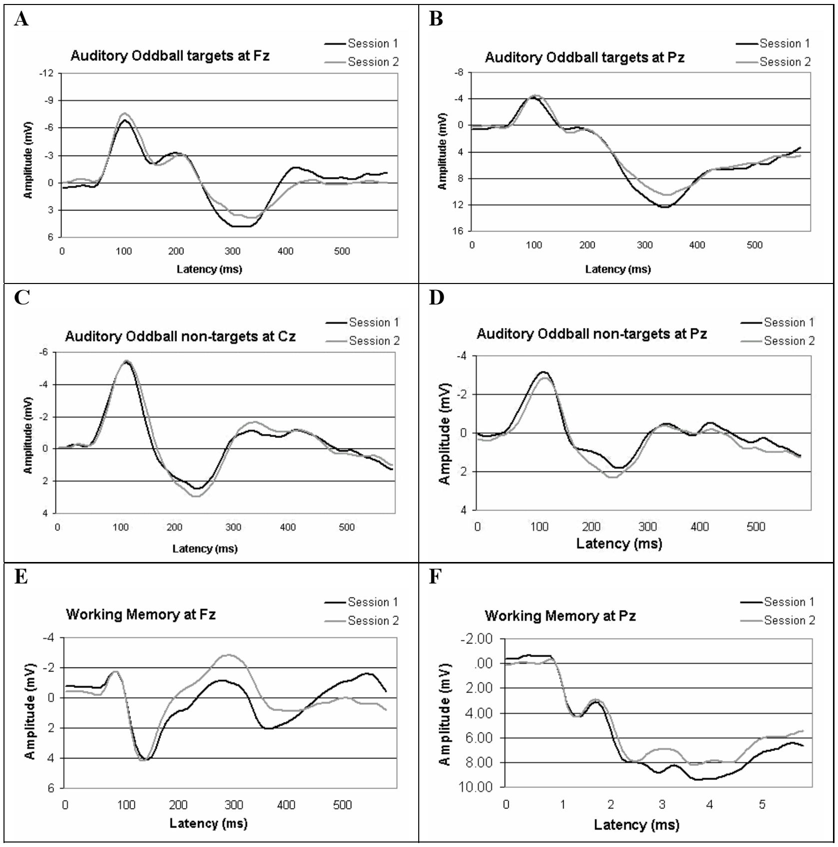

Figure 2. ERP averaged waveforms for oddball target data at Cz (A) and Pz (B), for oddball non-target data at Cz (C) and Pz (D), and working memory stimuli at Fz (E) and Pz (F).

topographical distribution of increasing response from frontal to parietal midline sites (Figure 2A–C).

For ERP latency to oddball targets, there were no significant session effects for N100 ($F_{(1,18)} = .024$, $p = .88$), P200 ($F_{(1,18)} = .85$, $p = .37$), N200 ($F_{(1,18)} = .34$, $p = .57$) and P300 ($F_{(1,18)} = .09$, $p = .77$), as well as no significant site effects.

---

[2] When age and sex were not controlled for, both N200 latency ($f = 4.90$, $df = 1.18$, $p = .042$) and P300 latency ($f = 4.84$, $df = 1.17$, $p = .042$) for targets were slightly longer at re-test. For nontargets, N100 latency was also slightly longer at re-test (by about 5 ms, $f = 4.92$, $df = 1.18$, $p = .04$). Together, these data suggest a slight latency shift of the whole waveform at re-test, a shift that interacts with demographic data.

For ERP amplitude and latency for target and non-target components, neither age nor sex showed a significant interaction with session. Site effects also did not interact with the session factor for the amplitude and latency of oddball ERP components.

For the working memory task, there were no significant effects over test-retest for the amplitude of the frontal P150 ($F_{(1,18)} = 2.98$, $p = .10$) and parietal P300 ($F_{(1,18)} = .69$, $p = .42$) components. There were also no significant session effects for the latency of the P150 ($F_{(1,18)} = .17$, $p = .68$) and P300 ($F_{(1,18)} = 1.13$, $p = .30$) components. Age and sex did not show significant differential effects on session for the amplitude and latency of these working memory components.

Neuropsychological Data

For the 10 neuropsychological tests, mean performance scores for key dependent measures across test and retest are presented in Figure 3A–C.

There were no significant session effects for any of the neuropsychological test scores across the five cognitive domains: sensori-motor [motor tapping ($F_{(1,18)} = .01$, $p = .92$), choice reaction time ($F_{(1,18)} = .06$, $p = .82$), choice reaction time ($F_{(1,18)} = 2.98$, $p = .10$)]; attention [span of visual memory ($F_{(1,18)} = .11$, $p = .74$), digit span *F*orwards ($F_{(1,18)} = .001$, $p = .98$), digit span reversed ($F_{(1,18)} = 2.51$, $p = .13$), switching of attention, Part 1 completion time ($F_{(1,18)} = 1.15$, $p = .30$), switching of attention, Part 1 errors ($F_{(1,18)} = .03$, $p = .87$)]; executive function [switching of attention, part 2 completion time ($F_{(1,18)} = .13$, $p = .72$), switching of attention, Part 2 errors ($F_{(1,18)} = .98$, $p = .34$), verbal interference, Part 1 reaction time ($F_{(1,18)} = .81$, $p = .38$), verbal interference, Part 1 errors ($F_{(1,18)} = .32$, $p = .58$), verbal interference, part 2 reaction time ($F_{(1,18)} = .61$, $p = .45$), verbal interference, Part 2 errors ($F_{(1,18)} = .05$, $p = .83$)]; language/verbal function [letter fluency ($F_{(1,18)} = .12$, $p = .73$), animal fluency ($F_{(1,18)} = .62$, $p = .44$)]; memory [list learning, trials 1–4 ($F_{(1,18)} = .87$, $p = .36$), list learning, total recall ($F_{(1,18)} = 2.4$, $p = .14$), list learning intrusions ($F_{(1,18)} = .79$, $p = .39$), list learning repeats ($F_{(1,18)} = .51$, $p = .48$)], and spot the real word reaction time ($F_{(1,18)} = 1.28$, $p = .26$) and errors ($F_{(1,18)} = .11$, $p = .75$).

---

[3] When age and sex were not controlled, only one test-retest main effect was observed at the corrected alpha level: Switching of Attention, Part 2 ($f = 9.47$, $df = 20$, $p = .006$). At the uncorrected alpha level, the following test-retest effects were observed: Spot the Real Word accuracy ($f = 7.18$, $p = .014$), but ot the IQ estimate; Word generation score (FAS) ($f = 5.69$, $p = .027$); Total accuracy for Memory Recall ($f = 10.27$, $p = .004$).

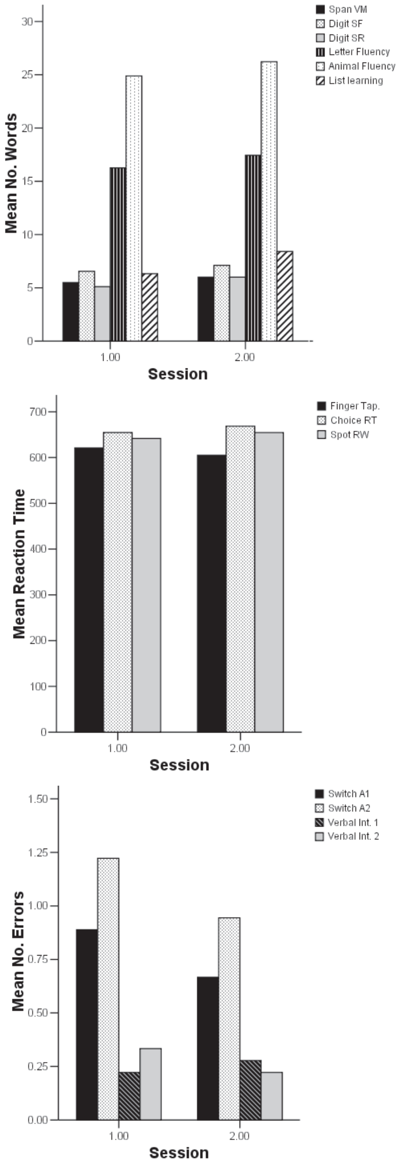

Figure 3. Neuropsychological test performance, in terms of accuracy, reaction time (RT), and number of words generated. Abbreviations: Switch A (Switching of attention, Parts 1 and 2), Verbal Int. (Verbal Interference test, Parts 1 and 2), Tap (Motor tapping RT), Choice RT (Choice Reaction Time test), Spot RW (Spot the Real Word test), Span VM (Span of Visual Memory test), Digit SF (Digit Span Forwards), Digit SR (Digit Span Reversed), Letter Fluency (Letters F,A,S), Animal Fluency (Animal Categories), and List Learning (Recall, Lists 1 to 4).

Neither age nor sex showed any significant interactions with test scores across each of these domains.

EEG Test-Retest Correlations

Results for the correlations between test and retest EEG measures (depicting the partial and zero-order correlation coefficients with shared variance due to r^2) are shown in Table 1.

Table 1. Test-retest reliability coefficients (zero-order correlations, with shared variance) and partial correlations (controlling for age) for each EEG band (eyes closed and eyes open conditions).

EEG power	Site	Shared variance (R^2)	Zero-order correlations	Partial correlations	*p* value
Eyes closed					
Delta	Fz	84	.89	.91	<.0001
	Cz	73	.81	.85	
	Pz	84	.92	.90	
Theta	Fz	81	.90	.89	<.0001
	Cz	76	.96	.96	
	Pz	87	.93	.92	
Alpha	Fz	61	.78	.78	<.0001
	Cz	75	.85	.87	
	Pz	74	.86	.85	
Beta	Fz	49	.70	.71	<.0001
	Cz	61	.73	.78	
	Pz	70	.81	.83	
Eyes open					
Delta	Fz	79	.87	.81	<.0001
	Cz	89	.94	.92	
	Pz	92	.96	.94	
Theta	Fz	86	.92	.89	<.0001
	Cz	89	.94	.93	
	Pz	92	.96	.95	
Alpha	Fz	73	.86	.85	<.0001
	Cz	77	.88	.87	
	Pz	84	.91	.90	
Beta	Fz	68	.82	.81	<.0001
	Cz	67	.81	.82	
	Pz	77	.88	.88	

Individual scores were highly stable over test-retest sessions for all four EEG bands, and for both eyes open and eyes closed conditions (Table 1). However, reliability coefficients were generally even larger for the eyes open (>0.8) compared to the eyes closed condition. Across conditions, the test-retest stability of individual subjects also tended to be relative greater for theta and delta power (0.86–0.95 and 0.85–0.94, respectively) than for the alpha and beta bands (.78–.90 and .71–.88, respectively).

ERP Test-Retest Correlations

Correlation coefficients (partial, zero-order, and shared variance) for test and retest ERP data for both auditory oddball and working memory tasks are shown in Tables 2A and B.

For the auditory oddball task, ERP components tended to show greater stability for amplitude than latency, particularly for N100, P200, and P300 (Tables 2A and B). The stability of individual test-retest ERP data tended to follow the preferential topography of each component; for example, relatively increasing stability of the P300 from Fz to Pz. Conversly, the least stable test-retest associations across individuals were for sites where ERP components are least robust, particularly for latency (such as N100 latency at Pz). For oddball targets, the P300 showed comparatively greater reproducibility than other components, whereas for non-targets the P200 was somewhat more consistent than the N100 (Tables 2A and B).

For the working memory task, the P150 was particularly stable, and showed the greatest shared variance across sessions. Consistent with the oddball task, the P300 showed greater reproducibility for amplitude than for latency.

Neuropsychological Test-Retest Correlations

Results for the correlations (partial, zero-order and shared variance) for the association of test-retest neuropsychological data are shown in Table 3.

Neuropsychological tests each showed significant correlations between test and retest, with reliability coefficients ranging from sound to high (.52–.89). Tests showing the greatest consistency for individual scores were from sensori-motor (simple motor tapping, .89), attention (spot the real word response time, .79; estimated intelligence, .80), executive function (switching of attention part 2, .72–.73; verbal interference response time, .71–.79) and memory (list learning intrusions, .77) domains (Table 3).

Table 2A. Test-retest reliability coefficients for (zero-order correlations, with shared variance) and partial correlations (controlling for age) for each ERP component elicited by targets and non-target stimuli in the auditory oddball paradigm

EEG power	Site	Shared variance (R^2)	Zero-order correlations	Partial correlations	*p* value
Oddball targets					
N100 amplitude	Fz	61	.78	.76	<.0001
	Cz	53	.71	.77	.001
	Pz	25	.50	.49*	.036
N100 latency	Fz	52	.72	.72	.001
	Cz	8	.28	.56	.238
	Pz	3	.16	.28	.510
P200 amplitude	Fz	58	.76	.68	<.0001
	Cz	56	.75	.79	<.0001
	Pz	1	.09	.13	.757
P200 latency	Fz	13	.36	.13	.157
	Cz	36	.63	.68	<.001
	Pz	9	.30	.28	.303
N200 amplitude	Fz	46	.68	.47	.003
	Cz	3	.19	.23	.673
	Pz	3	.17	.17	.540
N200 latency	Fz	50	.71	.71	.002
	Cz	18	.43	.43	.075
	Pz	9	.31	.28	.189
P300 amplitude	Fz	19	.44	.45	.082
	Cz	69	.83	.78	<.001
	Pz	69	83	.81	<.0001
P300 latency	Fz	34	.58	.56	.015
	Cz	32	.57	.69	.017
	Pz	42	.65	.66	.003
Oddball non-targets					
N100 amplitude	Fz	58	.76	.74	.001
	Cz	64	.80	.74	<.0001
	Pz	49	.70	.69	.001
N100 latency	Fz	40	.63	.63	.010
	Cz	19	.44	.69	.052
	Pz	8	.28	.33	.238
P200 amplitude	Fz	69	.83	.82	<.0001
	Cz	67	.82	.86	<.0001
	Pz	38	.62	.64	.002
P200 latency	Fz	34	.58	.62	.009
	Cz	17	.41	.49	.060
	Pz	14	.37	.33	.137

*Significance increased to $p < .01$.

Table 2B. Test-retest reliability coefficients for (zero-order correlations, with shared variance) and partial correlations (controlling for age) for each ERP component elicited by the working memory paradigm

EEG power	Site	Shared variance (R^2)	Zero-order correlations	Partial correlations	*p* value
P150 amplitude	Fz	76	.87	.84	<.0001
P150 latency	Fz	86	.93	.93	<.0001
P300 amplitude	Pz	34	.58	.55	.015
P300 latency	Pz	31	.56	.52	.019

DISCUSSION

In this study, the reliability of "NeuroMarker" was established; it is a new battery of neurophysiological and neuropsychological measures, which relies on a standardized methodology implemented in the brain resource international database (BRID). The neuropsychological test in "NeuroMarker" spans five domains of cognitive function (sensori-motor, attention, verbal, executive function/planning, and memory), and the concomitant neurophysiological measures comprise resting EEG (eyes open, eyes closed) and ERPs elicited during auditory oddball and working memory tasks. NeuroMarker tests and measures were found to be stable over a four-week period in healthy subjects. EEG power was highly reliable across all bands (r = .71–.95), particularly for the lower frequency bands, theta and delta (Table 1). Although the reliablity of ERP components was generally sound, it was on average lower than that for EEG data. The most consistent components were the fronto-central N100 and centro-parietal P300 elicited by oddball targets, the fronto-central N100 and P200 elicited to oddball non-targets, and the P150 to working memory stimuli (.76–.93; Table 2A,B). Neuropsychological tests showed moderate to high consistency across domains of function, with particularly stable scores in simple motor tapping, spot the real word and estimated intelligence, switching of attention part 2, verbal interference response time, and list learning intrusions (.71–.89; Table 3).

The high reliability coefficients observed for EEG data over the four-week period are consistent with previous studies (Gasser et al., 1985; McEvoy et al., 2000; Pollock et al., 1991). Notably, it was found that reliability tended to increase with lower frequency bands (Table 1), whereas previous reports have revealed comparatively lower reliability for the delta band (Gasser et al., 1985; Pollock et al., 1991). The difference in delta reliability has been observed at test intervals of over four months (Pollock et al., 1991). The

Table 3. Test-retest reliability coefficients for neuropsychological tests

Test (and dependent measure)	Shared variance (R^2)	Zero-order correlations	Partial correlations	*p* value
Simple Motor tapping (mean RT)	79	.87	.89	<.0001
Choice Reaction Time (mean RT)	46	.71	.68	<.001
Spot the Real Word (total correct)	48	.70	.69	<.001
Spot the Real Word (mean RT)	62	.76	.79	<.0001
Spot the Real Word (estimate of intelligence)	64	.78	.80	<.0001
Span of Visual Memory (total correct)	37	.59	.61	.002
Digit Span Forwards (maximum correct)	27	.50	.52	.009
Digit Span Reversed (maximum correct)	40	.60	.63	.001
Switching of Attention (Part 1—completion time)	28	.59	.53	.008
Switching of Attention (Part 1—errors)	46	.74	.68.	<.001
Switching of Attention (Part 2—completion time)	52	.76	.72	<.0001
Switching of Attention (Part 2—errors)	53	.77	.73	<.0001
Letter Fluency (FAS, total number)	61	.71	.71	<.0001
Animal Fluency (animals, total number)	30	.57	.55	.006
Verbal Interference (Part 1—mean RT)	50	.70	.71	<.0001
Verbal Interference (Part 1—errors)	34	.60	.58	.003
Verbal Interference (Part 2—mean RT)	62	.85	.79	<.0001
Verbal Interference (Part 2—errors)	35	.60	.59	.005
List Learning (Number correct, Trials 1–4)	29	.52	.54	.007
List Learning (total correct, Trials 1–7)	27	.51	.52	.009
List Learning (repeats)	29	.48	.54	.007
List Learning (intrusions)	59	.73	.77	<.0001

present findings suggest that there is minimal difference in the reliability of EEG bands over relatively short periods such as the four-week period employed in this study, consistent with findings comparing intervals of five minutes to several months (Salinksy et al., 2003).

To date, EEG reliability studies have focused on adult samples, particularly in the 18 to 40 years range (Burgess & Gruzelier, 1997; Gasser et al., 1985; McEvoy et al., 2000). In the present study, which spanned a wide range of ages, neither age nor sex were found to impact on EEG stability over four weeks (Table 1). This observation suggests that the EEG is a reliable tool for clinical assessment across the lifespan, although confirmation in a larger sample that includes children is warranted.

In line with previous reports (e.g., Waldhovd & Fjell, 2002), the degree of ERP reliability varied with topography, and was most robust for sites where ERP components were maximal. For instance, P300 amplitude and latency were more reliable at centro-parietal sites, than at the frontal site. Similarly, the N100 was more stable frontally, where it is typically maximal. Topographical distribution largely accounted for the instances in which ERP reliability coefficients did not reach significance.

This study provided a comparison between the reliability of standard oddball task ERPs to those elicited by a working memory, which requires greater vigilance. Although reliability was sound in both tasks, parietal P300 reproducibility was somewhat reduced for working memory, suggesting that attentional demands may increase its variability over time. By contrast, the earlier P150 component showed a high level of consistency, indicating that components tapping sustained vigilance are comparatively stable over repeated testing sessions. ERPs elicited by additional tasks are available in the "NeuroMarker" battery, and provides a means to assess the generality of ERP reliability across a wider range of cognitive load, and in regard to affect-related functions. From the current findings, the generally sound level of ERP consistency for sites of maximal activity suggests that the standardized methodology established in the BRID has utility for application in clinical evaluations. Given that instability in ERPs may suggest the presence of disease processes, and that low consistency may index cognitive decline and dementia in elderly subjects (Fabiani et al., 1987; Sandman & Patterson, 2000), the present methodology provides a sound normative framework to examine these clinical processes and changes.

Consistent with the previous studies, neuropsychological test reliability was generally sound across cognitive domains (Dikmen et al., 1999; Ushiyama et al., 1995; Watson et al., 1994).

It is noted that the high stability revealed for the spot the real word estimate of intelligence supports the view that this measure provides a reliable brief index of intellectual ability (Baddeley et al., 1993). Although significant test-retest correlations were found for all tests, the authors did, however, observe relatively lower reliability for digit span forwards performance and list learning recall, which tap aspects of memory, consistent with previous reports that memory may be more vulnerable to the differential effects of practice across individual subjects (Dikmen et al., 1999). Importantly, however, age and sex did not show a differential impact on test-retest reliability for these tests, indicating that they are applicable for retest assessments across age groups and across males and females. Indeed, our results suggest that the neuropsychological tests in "NeuroMarker" in general do not show a differential impact from age or sex over retest, for relatively short periods such as the four weeks used in this study.

CONCLUSION AND SIGNIFICANCE

Taken together, this study provides preliminary support for the consistency of the "NeuroMarker" battery over time, and the utility of using a standardized approach for neurocognitive assessments of clinical conditions and treatment across a wide range of ages. Future studies are warranted to examine reliability explicitly in clinical groups, the effects of age and sex using between-groups designs in large sample sizes, and reproducibility over longer retest intervals.

REFERENCES

American Psychological Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) (DSIM-IV). APA: Washington, DC.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psycholology*, *32,* 55–65.

Bahramali, H., Gordon, E., Lagopoulos, J., Lim, C. L., Li, W., Leslie, J., & Wright, J. (1999). The effects of age on late components of the ERP and reaction time. *Experimental Aging Research, 25,* 69–80.

Bolla-Wilson, K., & Bleecker, M. L. (1986). Influence of verbal intelligence, sex, age, and education on the Rey Auditory-Verbal Learning Test. *Developmental Neuropsychology*, *2,* 203–211.

Burgess, A. P., & Gruzelier, J. H. (1993). Individual reliability of amplitude distribution in topographical mapping of EEG. *Electroencephalography and Clinical Neurophysiology*, *86,* 219–223.

Burgess, A. P., & Gruzelier, J. H. (1997). How reproducible is the topographical distribution of EEG amplitude? *International Journal of Psychophysiology*, *26,* 113–119.

Chiappa, K. H. (1997). *Evoked potentials in clinical medicine* (3rd ed.). New York: Lippincott-Raven.

Clark, C. R., Orr, R. S., Wright, E. K., & Weber, D. L. (1998). Working memory updating to visual verbal stimuli: A high resolution ERP study. In Y. Koga, K. Nagata, and K. Hirata (Eds.), *Brain topography today* (pp. 173–178). Tokyo: Elsevier Science.

Clark, C. R., Paul, R. H., Williams, L. M., Arns, M., Sumich A., Handmer, C., & Gordon, E. (2005). Standardized assessment of cognitive functioning during development and aging using an automated touch screen battery. *Archives of Clinical Neuropsychology* (under review).

Connolly J. F., & D'Arcy J. C. N. (2000). Innovations in neuropsychological assessment using event-related potentials. *International Journal of Psychophysiology*, *37,* 31–47.

Dikmen, S. S., Heaton, R. K., Grant, I., & Temkin, N. R. (1999). Test-retest reliability and practice effects of Expanded Halstead-Reitan Neuropsychological test battery. *Journal of the International Neuropsychological Society*, *5,* 346–356.

Fabiani, M., Gratton, G., Karis, D., & Donchin, E. (1987). Definition, identification, and reliability of measurement of the P300 component of the event-related brain potential. *Advances in Psychophysiology, 2,* 1–78.

Gasser, T., Bächer, P., & Steinberg, H. (1985). Test-retest reliability of spectral parameters of the EEG. *Electroencephalography and Clinical Neurophysiology, 60,* 312–319.

Gevins, A. S., Zeitlin, G. M., Ancoli, S., & Yeager, C. L. (1977). Computer rejection of EEG artifact. II: Contamination by drowsiness. *Electroencephalography and Clinical Neurophysiology, 42,* 31–42.

Gevins, A. S., Zeitlin, G. M., Doyle, J. C., Yingling, C. D., Schaffer, R. E., Callaway, E., & Yeager, C. L. (1979). EEG correlates of higher cortical functions. *Science, 203,* 665–668.

Gevins, A., Smith, M. E., McEvoy, L., & Yu, D. (1997). High resolution EEG mapping of cortical activity related to working memory. *Cerebral Cortex, 7,* 374–385.

Golden, C. J. (1978). *Stroop Color and Word Test*. Wood Dale, IL: Stoelting Company.

Gordon, E. (2003a). Integrative neuroscience and psychiatry. *Neuropsycopharmacology, 28*(1), S2–S8.

Gordon, E. (2003b). Integrative neuroscience in psychiatry: The role of a standardized database. *Australasian Psychiatry*, *11*(2), 156–163.

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

Hickie, I., Hadzi-Pavlovic, D., Scott, E., on behalf of the SPHERE National Scretariat. (2001). SPHERE: A National Depression Project. *The Medical Journal of Australia, 175,* 52–55.

Kinoshita, S., Inoue, M., Maeda, H., Nakamura, J., & Morita, K. (1996). Long-term patterns of change in ERPs across repeated measurements. *Physiology Behavior, 60,* 1087–1092.

Klimesch, W., Doppelmayr, M., Schimke, H., & Ripper, B. (1997). Theta synchronization and alpha desynchronization in a memory task. *Psychophysiology, 34,* 169–176.

Klimesch, W., Doppelmayr, M., Wimmer, H., Schwaiger, J., Roehm, D., Gruber, W., & Hutzler, F. (2001). Theta band power changes in normal and dyslexic children. *Clinical Neurophysiology, 112,* 1174–1185.

Klimesch, W., Schack, B., Schabus, M., Doppelmayr, M., Gruber, W., & Sauseng, P. (2004). Phase-locked alpha and theta oscillations generate the P1-N1 complex and are related to memory performance. *Cognitive Brain Research, 19,* 302–316.

Lezak, M. D. (1995). *Neuropsychological Assessment* (3rd ed.). Oxford: New York.

Lim, C. L., Gordon, E., Rennie C., Wright, J. J., Bahramali, H., Li, W. M., Clouston, P., & Morris, J. G. (1999). Dynamics of SCR, EEG, and ERP activity in an oddball paradigm with short interstimulus intervals. *Psychophysiology, 36,* 543–551.

Maeda, H., Morita, K., Nakamura J., Masayuki, I., Kinoshita, S., Kodama, E., Maki, S., & Nakazawa, Y. (1995). Reliablity of the task-related component (P3b) of P3 event-related potentials. *Psychiatry and Clinical Neuroscience, 49,* 281–286.

Makeig, S., & Jung, T. P. (1995). Changes in alertness are a principal component of variance in the EEG spectrum. *NeuroReport, 7,* 213–216.

McEvoy, L. K., Smith, M. E., & Gevins, A. (2000). Test-retest reliability of task-related EEG. *Clinical Neurophysiology, 1,* 457–463.

Milner, B. (1970). Memory and the medial temporal regions of the brain. In K. H. Pribram and D. E. Broadbent (Eds.), *Biology of memory* (pp. 29–48). New York: Academic Press.

Nuwer, M. (1997). Assessment of digital EEG, quantitative EEG, and EEG brain mapping. *Report of the American Academy of Neurology and the American Clinical Neurophyiology Society, 49,* 277–292.

Paul, R. H., Lawrence, J., Williams, L. M., Clark, C. R., Cooper, N., & Gordon, E. (2005). A standardized battery for computerized neurocognitive assessment: Methods and construct validity of the "IntegNeuro™." *International Journal of Neuroscience, 115*(11), 1549–1567.

Petersen, R. C., Stevens, J. C., Ganguli, M., Tangalos, E. G., Cummings, J. L., & DeKosky, S. T. (2001). Practice parameter: Early detection of dementia: Mild cognitive impairment (an evidenced-based review). Report of the Quality Standards Subcommittee of the American Academy of Neurology. *Neurology, 8,* 1133–1142.

Polich, J. (1986). Normal variation of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology, 65,* 236–240.

Polich J., & Kok, A. (1995). Cognitive and behavioral determinants of P300: An integrative review. *Biological Psychology, 41,* 103–146.

Pollock, V. E., Schneider, L. S., & Lyness, S. A. (1991). Reliability of topographic quantitative EEG amplitude in healthy late-middle-aged and elderly subjects. *Electroencephalography and Clinical Neurophysiology*, *79,* 20–26.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8,* 271–276.

Roth, D. L., & Crosson, B. (1985). Memory span and long-term memory deficits in brain-injured patients. *Journal of Clinical Psychology, 41,* 521–527.

Salinsky, M. C., Oken, B. S., & Morehead, L. (1991). Test-retest reliability in EEG frequency analysis. *Electroencephalography and Clinical Neurophysiology, 79,* 382–392.

Sandman, C. A., & Patterson, J. V. (2000). The auditory event-related potential is a stable and reliable measure in elderly subjects over a 3 year period. *Clinical Neurophysiology, 111,* 1427–1437.

Segalowitz, S. J., & Barnes, K. L. (1993). The reliability of ERP components in the auditory oddball paradigm. *Psychophysiology, 30,* 451–459.

Spreen, O., & Strauss, E. (1988). *A compendium of neuropsychological tests: Administration, norms, and commentary*. New York: Oxford University Press.

Sullivan, R., Senior, G., & Hennessy, M. (2000). Australian age-education and premorbid cognitive/intellectual estimates for the WAIS-III. *6th Annual Conference of the APS College of Clinical Neuropsychologists*, Hunter Valley, NSW, Australia.

Thatcher, R. W. (1998). Normative EEG databases and EEG biofeedback. *Journal of Neurotherapy*, *2,* 8–39.

Ushiyama, C. L., D'Elia, L. F., Dellinger, A. M., Becker, J. T., Selnes, O. A., Wesch, J. E., Chen, B. B., Satz, P., van Gorp, W., & Miller, E. N. (1995). Alternate forms of the Auditory-Verbal Learning Test: Issues of test comparability, longitudinal reliability, and moderating demographic variables. *Archives of Clinical Neuropsychology, 10,* 133–145.

Walhovd, K. B., & Fjell, A. M. (2002). One-year test-retest reliability of auditory ERPs in young and old adults. *International Journal of Psychophysiology, 46,* 29–40.

Watson, F. L., Pasteur, M. A. L., Healy, D. T., & Hughes, E. A. (1994). Nine parallel versions of four memory tests: An assessment of form equivalence and the effects of practice on performance. *Human Psychopharmacology, 9,* 51–61.

Williams, L. M., Gordon, E., Bahramali, H., Wright, J. J., & Meares, R. A. (2000). Late component ERPs are associated with distinct schizophrenia syndromes. *International Journal of Neuroscience, 105,* 37–52.