

Published in final edited form as:

Clin Neurophysiol. 2011 January ; 122(1): 114–120. doi:10.1016/j.clinph.2010.06.010.

A cognitive and neurophysiological test of change from an individual's baseline

Alan Gevins^{a,*}, Michael E. Smith^{a,b}, Linda K. McEvoy^{a,c}, Aaron B. Ilan^a, Cynthia S. Chan^a, An Jiang^a, Lita Sam-Vargas^a, and Gordon Abraham^a

^aSan Francisco Brain Research Institute & SAM Technology, San Francisco, CA

Abstract

Objective—An automated cognitive neurophysiological test is presented that characterizes how an individual was affected by a drug or treatment. The test calculates sub-scores for working memory task performance, cortical activation, and alertness, and combines the sub-scores into an overall score.

Methods—The test was applied in a double-blind, placebo-controlled study of alcohol, caffeine, diphenhydramine, and sleep deprivation in 16 healthy adults.

Results—The between- and within-day variability of the sub-scores and overall scores for placebo were all near zero, suggesting that the scores are stable. All treatments affected the overall score, while differential effects on sub-scores highlighted the added value of EEG measures.

Conclusions—The test is sensitive to relatively mild alterations in cognitive function. Its automation makes it suitable for use in large-scale clinical trials.

Significance—By combining task performance with EEG brain function measures, the test may prove to have better sensitivity and specificity in detecting changes due to drugs or other treatments than comparable neuropsychological test batteries that do not directly measure brain function signals.

Keywords

brain function; cognition; EEG; working memory; clinical trials

1. Introduction

Cognitive brain function is affected by many diseases, by the intended and unintended effects of treatment medications, and by a variety of stressors such as disturbed sleep. There are many batteries for assessing cognition based on performance on tests of cognitive function and rating scales. However, behavior is the end product of many neural systems, some of which may be recruited or adapted in some way to compensate for deficits. For

© 2010 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

*Corresponding author. Address: Alan Gevins, San Francisco Brain Research Institute & SAM Technology, 536 Waller St., San Francisco, CA 94117, alan@sfbri.org, Tel: (415) 837-1600 ext 1.

^bCurrent affiliation: CorTechs Labs, San Diego, CA

^cCurrent affiliation: Dept. Radiology, University of California, San Diego, CA

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

instance, a motivated impaired person may make a greater effort and not show signs of impairment, and a person who is simply drowsy may have a poor cognitive test score but not have a disease affecting cognitive brain function per se. Tests that do not directly measure brain function signals have difficulty accounting for factors such as motivation and alertness, and therefore have limited sensitivity and specificity.

The lack of a clinical standard for testing cognitive brain function has been cited as a major confounding factor in the discrepancies between the results of different clinical trials (Vermeulen and Aldenkamp, 1995). Standardized neurologically based tests of an individual's cognitive brain function have the potential to make such evaluations more sensitive and efficient, and could be helpful to researchers, clinicians, and patients. Towards this goal, we present an automated cognitive neurophysiological test, the Sustained Working Memory Test (SWMT) that combines cognitive test performance measures with EEG measures. The multivariate analysis combines task response accuracy and speed measures with task-related and resting EEG measures to arrive at an assessment of how an experimental drug or stressor, or a disease and its treatment, has changed an individual's neurocognitive functional status. This paper focuses on the initial version of the SWMT that assesses treatment-related changes relative to an individual's placebo or pre-treatment baseline test.

1.1. Working memory, EEG and the scientific basis of the SWMT

Working memory (WM) is the fundamental cognitive function of controlling attention and actively sustaining its focus on a particular mental representation (Baddeley, 1992; Engle et al., 1999). It is essential for reasoning, planning, learning, and other higher cognitive functions, and is highly correlated with performance on psychometric tests of cognitive ability such as IQ tests (Carpenter et al., 1990; Gevins and Smith, 2000; Kyllonen and Christal, 1990).

Research on the neurophysiological signals of WM often employs "n-back" tasks in which participants respond to simple stimuli presented at different locations on a computer monitor once every few seconds (Gevins and Cutillo, 1993; Gevins et al., 1990). The load imposed on WM is varied across easy and more difficult versions, while perceptual and motor demands are kept constant (Gevins et al., 1979 b; Gevins et al., 1980). A spatial n-back WM task is used in the SWMT to minimize language-dependent cultural bias in the testing. In the easier 1-back task, participants have to decide whether the location of the current stimulus (a dot) is the same as on the previous trial (1-back); in the more difficult 2-back task, the current location of the dot has to be compared with the remembered position of the dot two trials ago in a continuous block of 50 trials. This requires constant updating of the information to be remembered on each trial, as well as focused attention to new stimuli and maintenance of representations of recently presented stimuli. To be successful when WM demands are high, as in the 2-back task, participants typically must make a significant and continuous mental effort. In this regard, the easier version of the task serves as a control condition.

Functional neuroimaging studies reliably demonstrate that n-back WM tasks activate circuitry in the frontal lobes critical to the control of attention and the maintenance of representations in WM (Cohen et al., 1994; Jansma et al., 2000; Jonides et al., 1993; McCarthy et al., 1994), and that the magnitude and extent of this activation is directly related to increasing load in n-back tasks (Braver et al., 1997). The discriminant validity of the n-back task as a measure of concentration is illustrated by the task impairment exhibited by groups with deficits suggesting impaired frontal lobe function, including patients with schizophrenia and children with head injury or ADHD (McCallister et al., 2001; Perlstein et al., 2003; Shallice et al., 2002). Abnormalities in frontal lobe activation during n-back task

performance in such groups have been noted even in cases where performance measures were insensitive (Callicott et al., 2003; McCallister et al., 2001). Such findings provide a strong rationale to use such tasks to gauge executive dysfunction and its neural correlates.

The spectral characteristics of the EEG display regular patterns of difficulty-related modulation during n-back task performance (Gevins et al., 1997). Fig. 1 illustrates regional differences in EEG in response to manipulations of WM load on spectral power (left) and high-resolution topographic maps of spectral peaks (right). At the frontal midline site, power in a 5–7Hz (theta) band is increased during the high load task. This “frontal-midline theta” signal is known to increase in difficult, attention demanding tasks requiring a sustained focus of concentration (Miyata et al., 1990). Topographic analyses and source modeling (Ishii et al., 1999) point to the anterior cingulate cortex as the likely origin of this signal. This region plays an important role in attention control (Posner and Rothbart, 1992), and activation in this region is known to increase with task difficulty (Paus et al., 1998). The attenuation of signals in the 8–13 Hz (alpha) band in the high relative to the low load n-back WM task has been observed in numerous studies (e.g. Gundel and Wilson, 1992), suggesting that the magnitude of alpha activity is inversely proportional to the quantity of cortical neurons recruited into a transient functional network for purposes of task performance (Mulholland, 1995; Pfurtscheller and Klimesch, 1992). Convergent evidence is also provided by observations of a negative correlation between alpha power and regional brain activation as measured with PET (Larson et al., 1998; Sadato et al., 1998) or fMRI (Goldman et al., 2002).

The EEG and, to a somewhat lesser extent, the performance measures during n-back tasks are highly reliable (Salinsky et al., 1991). In one study (McEvoy et al., 2000), average test-retest reliabilities were greater than .9 for EEG spectral features between two n-back task sessions one week apart ($p < .001$), and .86 for response speed ($p < .001$) and .47 for response accuracy ($p < .05$; the relatively low reliability observed for accuracy was due to a ceiling effect). Multivariate combinations of such EEG variables can identify specific cognitive states in individual participants accurately and reliably (Gevins et al., 1979 a; Gevins et al., 1979 c). For instance, multivariate EEG-based functions trained on one set of WM data and then cross-validated on new data correctly identified high vs. low load conditions with over 95% accuracy ($p < .001$, Gevins et al., 1998). Such results illustrate that EEG measures can reliably recognize different levels of task-related attention engagement.

A number of studies have reported how EEG signals during the n-back WM task are affected by fatigue and sleep loss (Smith et al., 2002), by medications that affect cognition and alertness (Gevins et al., 2002; McEvoy et al., 2001), and by recreational drugs including alcohol and marijuana (Ilan and Gevins, 2001; Ilan et al., 2004). Using a variety of analysis methods, detection of the effect of a drug or sleep loss was consistently most accurate when EEG measures were combined with task performance measures. For instance, sensitivity was 96% and specificity 100% in distinguishing the relatively strong neurocognitive effects of a widely prescribed anti-epileptic drug (carbamazepine) from those of a newer drug (levetiracetam) with milder side effects using EEG and task performance measures, but sensitivity and specificity were only 75% and 75%, respectively, using measures from conventional neuropsychological tests and subjective questionnaires (Meador et al., 2007).

The SWMT therefore combines EEG and n-back task performance measures to quantify how a treatment has affected cognitive brain function. The reliability of the SWMT scores is illustrated here by computing between-day and within-day variability in a large sample of healthy adults who performed the test multiple times without an active drug or other treatment. As an example of the application of the SWMT, these no-treatment variability values are then used to assess the significance of the effects of caffeine, alcohol, the

antihistamine diphenhydramine, and sleep deprivation. Based on the well-known effects of these drugs and sleep loss on cognition and brain function, the SWMT scores would be expected to show negative effects of alcohol, diphenhydramine, and sleep deprivation, and positive effects of caffeine.

2. Methods

2.1. EEG recording

EEG signals were recorded during task and resting conditions with a stretchable nylon cap with electrodes over bilateral and midline dorsolateral prefrontal locations (F9, F10, Fp1, Fp2, FpZ, F3, F4, Fz), midline sensorimotor cortex (Cz), lateral superior parietal cortex (P3, P4) and midline parieto-occipital cortex (POz), referenced to digitally linked mastoids. These locations were selected for their sensitivity to variations in working memory load on the basis of cognitive EEG studies with 40 or 100 electrodes (e.g. Gevins et al., 1996; Gevins et al, 1997). Vertical and horizontal eye movements were monitored by the electrodes above and at the outer canthus of each eye. Signals were sampled at 256 Hz and band-pass filtered from 0.1 to 35 Hz.

2.2. Cognitive testing

EEG was recorded during a test battery consisting of easier and more difficult versions of a spatial n-back WM task. In the WM task a dot stimulus was displayed for 200 ms in one of 6 positions on each trial with a mean inter-stimulus interval of 4 sec (range 3500 – 4500 ms). In the easier (low-load) version of the task, participants had to decide whether the spatial location of the dot on each trial matched the location of the dot on the immediately preceding trial. In the more difficult (high-load) version, each dot was compared to the dot that appeared two trials before. Participants responded “match” or “no-match” on each trial with the left and right mouse buttons, respectively. Participants received sufficient practice to stabilize performance prior to the first testing session (150 trials of the easier and 200 trials of the more difficult WM task). For the actual test, approximately 3.5 minute blocks of 50 trials each (25 match and 25 no-match, randomly ordered) were presented for the low-load and for the high-load tasks. Resting EEG was then recorded for 90 sec each in eyes-open and eyes-closed conditions.

2.3 Data analysis

Following each test, data were uploaded over the internet to a central data analysis server. There, automated algorithms removed artifact contaminated data, computed parameters from the task performance data and background EEG spectra, computed multivariate functions combining the parameters, and reported and stored intermediate and final results.

2.3.1. Artifact decontamination—Fourth-generation algorithms detected a variety of different types of artifacts including eye movements and blinks, scalp muscle activity, head and body movements and bad electrode contacts. In a formal evaluation of performance on a database of ~40,000 artifacts, the algorithms detected 98.3% of the artifacts with a false detection rate of 2.9%, whereas the consensus of 3 expert human judges found 96.5% of the artifacts with a 1.7% false detection rate. After artifacts were detected, adaptive filters with and without noise reference signals (Du et al., 1994) were applied to remove the contaminants when possible. All raw and decontaminated data and EEG spectra were visually inspected following the automated decontamination.

2.3.2. Combined EEG and cognitive task performance analysis—For convenience, an overall score that combined task performance and two EEG sub-scores was computed to indicate how the participant’s state changed from baseline, where baseline

consisted of one or more tests administered to the participant prior to treatment or during a placebo condition. Statistical significance was determined by comparing an individual's score to the distribution of scores from a population of participants tested on multiple occasions in the absence of any treatment.

The overall score was computed as the mean of three sub-scores, one based on WM task performance measures (performance sub-score) and the other two on EEG (cortical activation and alertness sub-scores). Like the overall score, each of these sub-scores indicated whether the participant scored lower or higher relative to his or her baseline, using standard deviation units. The performance sub-score quantified the speed and accuracy of the responses made in the WM tasks. The cortical activation sub-score measured the neural workload, and by inference the attentional effort, exerted to produce the observed level of WM task performance. The alertness sub-score provided a neurophysiological measure of how alert a participant was when resting and not challenged with a cognitive task. It was included in the overall score to quantify the influence of drowsiness, a common effect of many medications and disorders. Consideration of the three sub-scores provided insight into these different aspects of cognitive performance and brain function.

Combining the sub-scores into a single score provided a convenient overall assessment of change in a participant's state from baseline. However, the three sub-scores could not merely be averaged together since there is no one-to-one relationship between measures of brain function and task performance and a particular behavior can be produced by the brain in a variety of different ways. For instance, a person may produce relatively poor performance on a particular task because of brain disease, or a low level of effort due to drowsiness, lack of motivation, etc. Therefore, algorithmically combining task performance and brain function measures into an overall score necessitated use of expert knowledge of contextual interpretation of changes in both types of measures taken together. For the SWMT, three simple rules were applied to the signs of the sub-scores so that the overall score more accurately reflected a positive or negative change in the participant's neurocognitive state from baseline. One rule addressed dissociation between performance and alertness sub-scores and two others addressed dissociation between performance and cortical activation sub-scores. For the alertness rule, if WM performance decreased from baseline while alertness increased, the sign of the alertness sub-score was inverted before being averaged into the overall score, the rationale being that increased alertness was not helpful to overall neurocognitive status if performance was worse. For cortical activation, if performance increased from baseline while cortical activation decreased, the sign of the cortical activation sub-score was inverted before being averaged into the overall score, the rationale being that less cortical activation in the presence of better performance reflected less effort required to perform the task better and hence improved overall neurocognitive status. Conversely, if cortical activation increased from baseline while performance decreased, then the sign of the cortical activation sub-score was inverted before being averaged into the overall score. The rationale was that more cortical activation with worse performance reflected a neurocognitive state in which the patient was trying harder but performing worse.

Each sub-score itself consisted of a combination of several individual variables. The difference in each raw variable between the participant's current test and baseline test was calculated. The resulting "change score" was then divided by a normative standard deviation of that change score. These normative standard deviations of change scores were calculated from a large database of participants who had performed the SWMT multiple times without a drug or other intervention, and provide an approximation of how much a particular variable can be expected to vary from baseline to follow-up test in the absence of an active intervention. Dividing a participant's change score by a normative standard deviation puts

all the individual measures on a comparable scale — standard deviation units — before they are combined with other measures into sub-scores and an Overall score. Although the individual raw variables are measured in different units (e.g. percent correct, milliseconds, decibels, etc.) transforming each variable into standard deviation units before combining them makes the resulting sub-scores and overall score scale-independent.

The performance sub-score indicated how well the participant performed the easier and more difficult WM tasks relative to that participant's baseline. It was computed as the difference between the current test and the baseline test in measures of performance accuracy (percent correct) and mean and standard deviation of reaction time (milliseconds). Because the difficult version of the task places higher demands on focused attention and working memory (Gevins et al., 1997), it was given more emphasis in the sub-score. To summarize the performance on the WM tasks succinctly, we compacted these key variables into three measures, one reflecting performance (accuracy and reaction time) in the low load WM task, and two reflecting performance in the high load WM task (one accuracy, the other reaction time). When the mean of these three measures was used as the performance sub-score, the high load WM task thus received twice the weight of the low load WM task. During test development using a variety of data sets, these measures were reliable and sensitive to factors known to affect WM; different variables may prove to be more suitable measures of WM performance in other studies.

The cortical activation sub-score reflected the difference between the current test and the baseline in the divergence (see below) between EEG power spectral variables, in decibels, measured during the easier vs. the more difficult WM task, i.e., the difference in EEG measures in the degree to which large cortical neuronal populations were recruited to perform the more difficult version of the WM task relative to the easier task (Gevins et al., 1997; Gevins and Smith, 2008). A positive cortical activation score indicated a larger neuronal population mediating performance of the more difficult task relative to baseline, whereas a negative cortical activation score indicated a smaller neuronal population recruited for difficult task performance relative to baseline. First, within each 2-sec window over the task performance interval, EEG power was computed across theta, alpha, and beta frequency bands (Gevins et al., 1997) for three frontal and three parietal channels. Then, for the frontal and parietal regions separately, a multivariate divergence analysis selected the subset of four EEG power variables that, in combination, yielded the greatest differentiation of the easy from the more difficult WM task. Multivariate divergence analysis (Smith et al., 2001) is a type of discriminant analysis that performs an exhaustive search over all possible subsets of variables from a set of candidate variables to find the particular subset that maximizes the multivariate distance between two sets of data, in this case EEG power spectral measures from the easy vs. the more difficult WM tasks. The specific subset of EEG variables was chosen in this way for each participant, and the cortical activation sub-score was the mean of the resulting frontal and parietal divergence measures. Only variables known a priori to be most sensitive to task-load modulation and less influenced by drowsiness or drug effects were considered (Fig. 1). For instance, alpha band power was only selected for an individual participant if power was larger in the easier WM task for that individual, and frontal theta measures were only selected if power was larger in the more difficult WM task (Gevins et al., 1997).

The alertness sub-score was a neurophysiological measure that indicated how the participant's alertness during the current test session differed from the baseline test session. It used EEG variables recorded while the participant was resting and not challenged with the WM tasks because such cognitive tasks are activating, and their concurrently recorded EEG measures reflect factors other than alertness. The alertness sub-score was computed as the mean of three well-established neurophysiological markers of alertness (Davis et al., 1937;

Makeig and Jung, 1995; Matousek and Petersen, 1983; Oken and Salinsky, 1992): a ratio of alpha band power between eyes closed and eyes open resting tasks, slow horizontal eye movement activity during the resting task, and resting low frequency EEG power (measured in decibels).

3. Results

3.1. Normal variability of test scores in the absence of a treatment

The variability of the overall score and the 3 sub-scores in the absence of a drug or other experimental treatment was computed for 127 healthy adults (mean age 34 years, range 18–70) who were tested in seven studies with drug and sleep deprivation interventions (Gevins et al., in preparation; Gevins et al., 2002; Ilan et al., 2005; McEvoy et al., 2001; Meador et al., 2007; Smith et al., 2002; Smith et al., 2006). The experiments were conducted according to protocols approved by an NIH-registered Institutional Review Board and were in compliance with the Helsinki Declaration. All studies with drugs were double-blind, randomized, crossover experiments. The number of test days varied from 2 to 6 across studies and the number of times the SWMT was administered on a test day also varied from 2 to 6, providing a representative sample for computing variability of the test in the absence of a treatment.

Table 1a shows the mean, standard deviation and confidence interval of the overall score and the 3 sub-scores across test sessions in which no treatment was administered, relative to each participant's first test. The between-day variability was computed for 279 tests from 127 participants based on the first test session on subsequent test days relative to the first test session on a participant's initial test day. Subsequent test days followed the initial test day by 1 week to 6 months. (Only the initial test session of each day was used because drugs were administered during some subsequent sessions.) The within-day variability was computed for 362 tests from 118 of the 127 healthy adults who participated in the 6 studies in which the test was performed multiple times throughout a test day without a treatment. (There were fewer tests for the Alertness sub-score because one of the studies did not have applicable data.) The within-day variability was based on all follow-up tests relative to the first test on a participant's initial testing day. The total variability includes follow-up sessions with no intervention that occurred on a later day or any time on the same day as the initial session. Between-day, within-day and total variability of the overall score and three sub-scores were all close to zero, with standard deviations all below 1.

Table 1b shows the mean, standard deviation and confidence interval of between-day and within-day variability of the overall score for each of the seven studies comprising the data shown in Table 1a. The variabilities in each of the studies were also close to zero with standard deviations below 1. Similar results were found for the three sub-scores in each of the seven studies.

3.2. Test scores for four treatments that affected attention and alertness

Sixteen healthy adults participated in a randomized, double-blind, placebo-controlled crossover study with multiple drug treatments as well as an overnight sleep deprivation session (Gevins et al., 2002; Smith et al., 2002). Test sessions were separated by at least one week. Each session involved a baseline recording prior to administration of a study drug and two placebo drugs. In different sessions, the active study drug was: 1) alcohol, a 500 cc drink containing 0.88g/kg of 95% ethanol mixed in fruit juice, sufficient to raise blood/breath alcohol concentration (BBAC) to 0.08; 2) 200 mg of caffeine, equivalent to approximately 2 cups of coffee; or 3) 50mg of diphenhydramine, an antihistamine that commonly induces drowsiness. Placebo pills were capsules of powdered sugar; the placebo drink contained 495

cc of fruit juice with 5cc of ethanol floated on top to mimic the taste and smell of the active alcohol treatment. The sleep deprivation treatment consisted of staying up all night performing the SWMT and other tasks at regular intervals. Table 2 presents the overall and sub-scores at the time of peak effect in these four conditions (as well as at equivalent points in the placebo session), which was 1.5 hrs after drug ingestion in the cases of caffeine and alcohol, 2.5 hrs after drug ingestion in the case of diphenhydramine, and at 5 a.m. during the overnight sleep deprivation session (with 11 p.m. for comparison). Post-caffeine, alcohol, diphenhydramine, and placebo scores were computed relative to the pre-drug baseline test performed on the same day. The sleep deprivation session scores were computed relative to an average of four pre-drug afternoon tests that same participant performed on different test days.

The overall score was sensitive to the effects of all three drugs and sleep loss. Caffeine increased the overall score ($p < .01$), whereas diphenhydramine ($p < .001$), sleep loss ($p < .001$), and alcohol ($p < .05$) decreased it. The performance sub-score was diminished by diphenhydramine ($p < .001$; accuracy decreased and reaction time increased in both WM tasks) and sleep loss ($p < .001$; accuracy decreased and reaction time increased in both WM tasks), but was not affected by caffeine (slight increases in accuracy and decreases in reaction time in both WM tasks were not significant) or alcohol (although there was an interaction with task difficulty, $p < .05$, as reaction time decreased in the high load but not the low load WM task). The cortical activation sub-score was not affected by any of the four conditions, suggesting that participants continued to make an effort to perform the tasks. The alertness sub-score was affected by all four conditions, increasing after caffeine ($p < .01$), and decreasing after diphenhydramine ($p < .001$), sleep loss ($p < .01$), and alcohol ($p < .05$). The effects of caffeine and alcohol on EEG power spectra are illustrated in Fig. 2. By contrast, the overall score and sub-scores were not affected by placebo or during the 11 p.m. interval that preceded staying up all night.

4. Discussion

An efficient neurological test of mental acuity that directly measures brain signals of fundamental cognitive functions such as attention and memory could be helpful to researchers, clinicians, and patients alike. Without such an objective measure, assessing treatment of disorders that affect thinking is necessarily imprecise, as a physician can only roughly gauge whether a patient's cognitive brain function is deteriorating or improving with treatment. Here we present a test that is a first step towards addressing this need, and illustrate its use in measuring the effects of several drugs and sleep deprivation.

The Sustained Working Memory Test combines cognitive test performance with simultaneously recorded EEG measures to provide an overall score and three sub-scores which characterize how a treatment affected an individual's cognitive brain function. The between-day, within-day, and total variabilities of the overall score and sub-scores in the absence of a treatment (based on 641 tests administered to 127 healthy adults in seven studies) were all close to zero with standard deviations below one, suggesting that the scores are stable. The measures that comprise the scores are themselves reliable, with EEG power spectra retest reliabilities exceeding .9 (McEvoy et al., 2000). The sensitivity of the test to treatments affecting alertness and working memory was assessed in a study in which 16 healthy adults were tested before and after taking caffeine, diphenhydramine, and alcohol, as well as before and during an overnight sleep deprivation session. The overall score was significantly altered by all four treatments, with diphenhydramine and sleep deprivation having the largest negative impact and caffeine the largest positive effect. By contrast neither placebo nor being tested at 11 p.m. had an effect on the overall score or sub-scores. The overall score was more sensitive than the sub-scores by themselves. The sub-scores

were differentially affected by the treatments: 1) the task performance sub-score was only affected by diphenhydramine and sleep loss; 2) the EEG-based alertness sub-score was affected by all four treatments; and 3) the EEG-based cortical activation sub-score was not affected by any of the treatments, suggesting that participants continued to make about the same degree of effort to perform the tasks across treatments. In applying the SWMT to a new study, estimates of the variability of the scores and sub-scores would be based upon the study particulars and would not necessarily include a heterogeneous mix of experiments and test intervals. Likewise, in future studies, different task performance measures, EEG variables or cognitive tasks may be more effective than the ones reported herein.

The SWMT can be administered by an entry level assistant in under an hour. Automated data analysis is centralized for uniform quality assurance auditing and to facilitate remote examination of raw and processed data in accordance with the American Academy of Neurology and the American Clinical Neurophysiology Society's transparency guidelines for quantitative EEG analysis (Nuwer, 1997). This automation and uniform centralized analysis make the SWMT suitable for large scale clinical trials. The largest trial to date to employ the test is the NIH Apnea Positive Pressure Long-Term Efficacy Study, a randomized, double-blinded, sham-controlled, multicenter, long-term clinical trial to determine whether Continuous Positive Airway Pressure therapy for treatment of obstructive sleep apnea improves neurocognitive function and quality of life (NHLBI -- 5U01HL068060, Kushida et al., 2006). The SWMT is the primary outcome measure of executive frontal-lobe function for the trial.

The SWMT may be among the first of a new generation of tests under development. Although these new tests will not be a substitute for a comprehensive neuropsychological examination, the hope is that by combining cognitive task performance with neurophysiological brain function measures, such tests may detect the effects of diseases and their treatments efficiently and with high sensitivity and specificity.

Acknowledgments

This research was supported by grants from U.S. government agencies including the National Institute of Neurological Diseases and Strokes, The National Institute of Mental Health, The National Heart Lung and Blood Institute, The Air Force Research Laboratory and The Office of Naval Research.

We gratefully acknowledge the essential contributions of the following scientists, engineers and associates in the development reported here, including Enoch Callaway, Behram daCosta, Ritu Chellaramani, Zachary Davis, Robert Du, Jamie Elmasu, Brynn Evans, Moira Fordyce, Mathieu Herbette, Robert Howard, Guogang Hu, Harrison Leong, Min Lu, Emilie Schwager, Emiliania Simon-Thomas, Tim Stearns, Art Sandoval, Patrick Sullivan, and Ivy Tong. This paper is dedicated to Brian Cutillo (1945–2006) who devised the n-back working memory task in 1983.

References

- Baddeley A. Working Memory. *Science* 1992;255:556–559. [PubMed: 1736359]
- Braver TS, Cohen JD, Nystrom LE, Jonides J, Smith EE, Noll DC. A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* 1997;5:49–62. [PubMed: 9038284]
- Callicott JH, Egan MF, Mattay VS, Bertolino A, Bone AD, Verchinski B, et al. Abnormal fMRI response of dorsolateral prefrontal cortex in cognitively intact siblings of patients with schizophrenia. *Am J Psychiatry* 2003;160:709–719. [PubMed: 12668360]
- Carpenter PA, Just MA, Shell P. What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychol Rev* 1990;97:404–431. [PubMed: 2381998]
- Cohen JD, Forman SD, Braver TS, Casey BJ, Servan-Schreiber D, Noll DC. Activation of prefrontal cortex in a non-spatial working memory task with functional MRI. *Hum Brain Mapp* 1994;1:293–304.

- Davis H, Davis PA, Loomis AL, Harvey EN, Hobart G. Human brain potentials during the onset of sleep. *J Neurophysiol* 1937;1:24–37.
- Du, W.; Leong, HM.; Gevins, AS. Ocular artifact minimization by adaptive filtering; Proceedings of the Seventh IEEE SP Workshop on Statistical Signal and Array Processing; Quebec City, Canada. 1994. p. 433-436.
- Engle, RW.; Tuholski, S.; Kane, M. Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In: Miyake, A.; Shah, P., editors. *Models of Working Memory*. Cambridge: Cambridge University Press; 1999. p. 102-134.
- Gevins A, Cuttillo B. Spatiotemporal dynamics of component processes in human working memory. *Electroenceph Clin Neurophysiol* 1993;87:128–143. [PubMed: 7691540]
- Gevins A, Smith ME, McEvoy L, Yu D. High-resolution EEG mapping of cortical activation related to working memory: effects of task difficulty, type of processing, and practice. *Cereb Cortex* 1997;7:374–385. [PubMed: 9177767]
- Gevins A, Smith ME, Leong H, McEvoy L, Whitfield S, Du R, et al. Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Hum Fact* 1998;40:79–91.
- Gevins A, Smith ME. Neurophysiological measures of working memory and individual differences in cognitive ability and cognitive style. *Cereb Cortex* 2000;10:829–839. [PubMed: 10982744]
- Gevins A, Smith ME, McEvoy LK. Tracking the cognitive pharmacodynamics of psychoactive substances with combinations of behavioral and neurophysiological measures. *Neuropsychopharmacology* 2002;26:27–39. [PubMed: 11751030]
- Gevins, A.; Smith, ME. Electroencephalography (EEG) in Neuroergonomics. In: Parasuraman, R.; Rizzo, M., editors. *Neuroergonomics: the brain at work*. Oxford: Oxford University Press; 2008. p. 15-31.
- Gevins AS, Zeitlin GM, Doyle JC, Schaffer RE, Callaway E. EEG patterns during "cognitive" tasks. II. Analysis of controlled tasks. *Electroenceph Clin Neurophysiol* 1979 a;47:704–710. [PubMed: 91499]
- Gevins AS, Zeitlin GM, Doyle JC, Yingling CD, Schaffer RE, Callaway E, et al. Electroencephalogram correlates of higher cortical functions. *Science* 1979 b;203:665–668. [PubMed: 760212]
- Gevins AS, Zeitlin GM, Yingling CD, Doyle JC, Dedon MF, Schaffer RE, et al. EEG patterns during "cognitive" tasks. I. Methodology and analysis of complex behaviors. *Electroenceph Clin Neurophysiol* 1979 c;47:693–703. [PubMed: 91498]
- Gevins AS, Doyle JC, Schaffer RE, Callaway E, Yeager C. Lateralized cognitive processes and the electroencephalogram. *Science* 1980;207:1005–1008. [PubMed: 7352295]
- Gevins AS, Bressler SL, Cuttillo BA, Illes J, Miller JC, Stern J, et al. Effects of prolonged mental work on functional brain topography. *Electroenceph Clin Neurophysiol* 1990;76:339–350. [PubMed: 1699727]
- Gevins AS, Smith ME, Le J, Leong H, Bennett J, Martin N, et al. High resolution evoked potential imaging of the cortical dynamics of human working memory. *Electroenceph Clin Neurophysiol* 1996;98:327–348. [PubMed: 8641154]
- Goldman RI, Stern JM, Engel JJ, Cohen MS. Simultaneous EEG and fMRI of the alpha rhythm. *Neuroreport* 2002;13:2487–2492. [PubMed: 12499854]
- Gundel A, Wilson GF. Topographical changes in the ongoing EEG related to the difficulty of mental tasks. *Brain Topog* 1992;5:17–25.
- Ilan AB, Gevins A. Prolonged neurophysiological effects of cumulative wine drinking. *Alcohol* 2001;25:137–152. [PubMed: 11839457]
- Ilan AB, Smith ME, Gevins A. Effects of marijuana on neurophysiological signals of working and episodic memory. *Psychopharmacology* 2004;176:214–222. [PubMed: 15502936]
- Ilan AB, Gevins A, Coleman M, ElSohly MA, de Wit H. Neurophysiological and subjective profile of marijuana with varying concentrations of cannabinoids. *Behav Pharmacol* 2005;16:487–496. [PubMed: 16148455]

- Ishii R, Shinosaki K, Ukai S, Inouye T, Ishihara T, Yoshimine T, et al. Medial prefrontal cortex generates frontal midline theta rhythm. *Neuroreport* 1999;10:675–679. [PubMed: 10208529]
- Jansma JM, Ramsey NF, Coppola R, Kahn RS. Specific versus nonspecific brain activity in a parametric n-back task. *Neuroimage* 2000;12:688–697. [PubMed: 11112400]
- Jonides J, Smith EE, Koeppel RA, Awh E, Minoshima S, Mintun M. Spatial working memory in humans as revealed by PET. *Nature* 1993;363:623–625. [PubMed: 8510752]
- Kushida CA, Nichols DA, Quan SF, Goodwin JL, White DP, Gottlieb DJ, et al. The Apnea Positive Pressure Long-term Efficacy Study (APPLES): rationale, design, methods, and procedures. *J Clin Sleep Med* 2006;2:288–300. [PubMed: 17561541]
- Kyllonen PC, Christal RE. Reasoning ability is little more than working memory capacity?! *Intell* 1990;14:389–433.
- Larson CL, Davidson RJ, Abercrombie HC, Ward RT, Schaefer SM, Jackson DC, et al. Relations between PET-derived measures of thalamic glucose metabolism and EEG alpha power. *Psychophysiol* 1998;35:162–169.
- Makeig S, Jung TP. Changes in alertness are a principal component of variance in the EEG spectrum. *Neuroreport* 1995;7:213–216. [PubMed: 8742454]
- Matousek M, Petersen I. A method for assessing alertness fluctuations from EEG spectra. *Electroenceph Clin Neurophysiol* 1983;55:108–113. [PubMed: 6185295]
- McCallister TW, Sparling MB, Flashman LA, Gueirn SJ, Mamourian AC, Saykin AJ. Differential working memory load effects after mild traumatic brain injury. *Neuroimage* 2001;14:1004–1012. [PubMed: 11697932]
- McCarthy G, Blamire AM, Puce A, Nobre AC, Bloch G, Hyder F, et al. Functional magnetic resonance imaging of human prefrontal cortex activation during a spatial working memory task. *Proc Natl Acad Sci* 1994;91:8690–8694. [PubMed: 8078943]
- McEvoy LK, Smith ME, Gevins A. Test-retest reliability of cognitive EEG. *Clin Neurophysiol* 2000;111:457–463. [PubMed: 10699407]
- McEvoy, LK.; Chellaramani, R.; Smith, ME.; Gevins, A. Effects of a benzodiazepine on behavioral and neurophysiological measures of working memory. Annual Meeting of the Society for Neuroscience; 2001 November; San Diego, CA. 2001.
- Meador KJ, Gevins A, Loring DW, McEvoy LK, Ray PG, Smith ME, et al. Neuropsychological and neurophysiologic effects of carbamazepine and levetiracetam. *Neurol* 2007;69:2076–2084.
- Miyata Y, Tanaka Y, Hono T. Long term observation on Fm-theta during mental effort. *Neurosci* 1990;16:145–148.
- Mulholland T. Human EEG, behavioral stillness and biofeedback. *Int J Psychol* 1995;19:263–279.
- Nuwer M. Assessment of digital EEG, quantitative EEG, and EEG brain mapping: report of the American Academy of Neurology and the American Clinical Neurophysiology Society. *Neurol* 1997;49:277–292.
- Oken BS, Salinsky M. Alertness and attention: Basic science and electrophysiologic correlates. *J Clin Neurophysiol* 1992;9:480–494. [PubMed: 1361195]
- Paus T, Koski L, Caramanos Z, Westbury C. Regional differences in the effects of task difficulty and motor output on blood flow response in the human anterior cingulate cortex: a review of 107 PET activation studies. *Neuroreport* 1998;9:R37–R47. [PubMed: 9674567]
- Perlstein WM, Dixit NK, C.S.C, Noll DC, Cohen JD. Prefrontal cortex dysfunction mediates deficits in working memory and prepotent responding in schizophrenia. *Biol Psychiatry* 2003;53:25–38. [PubMed: 12513942]
- Pfurtscheller G, Klimesch W. Functional topography during a visuo-verbal judgment task studied with event-related desynchronization mapping. *J Clin Neurophysiol* 1992;9:120–131. [PubMed: 1552000]
- Posner, MI.; Rothbart, MK. Attentional mechanisms and conscious experience. In: Milner, AD.; Rugg, MD., editors. *The Neuropsychology of Consciousness*. San Diego: Academic Press; 1992. p. 91–111.
- Sadato N, Nakamura S, Oohashi T, Nishina E, Fuwamoto Y, Waki A, et al. Neural networks for generation and suppression of alpha rhythm: a PET study. *Neuroreport* 1998;30:893–897. [PubMed: 9579686]

- Salinsky MC, Oken BS, Morehead L. Test-retest reliability in EEG frequency analysis. *Electroenceph Clin Neurophysiol* 1991;79:383–392.
- Shallice T, Marzocchi GM, Coser S, Del Savio M, Meuter RF, Rumiati RI. Executive function profile of children with attention deficit hyperactivity disorder. *Developmental Neuropsychology* 2002;21:43–71. [PubMed: 12058835]
- Smith ME, Gevins A, Brown H, Karnik A, Du R. Monitoring task load with multivariate EEG measures during complex forms of human computer interaction. *Hum Fact* 2001;43:366–380.
- Smith ME, McEvoy LK, Gevins A. The impact of moderate sleep loss on neurophysiologic signals during working memory task performance. *Sleep* 2002;25:784–794. [PubMed: 12405615]
- Smith ME, Gevins A, McEvoy LK, Meador KJ, Ray PG, Gilliam F. Distinct cognitive neurophysiologic profiles for lamotrigine and topiramate. *Epilepsia* 2006;47:695–703. [PubMed: 16650135]
- Vermeulen J, Aldenkamp AP. Cognitive side-effects of chronic antiepileptic drug treatment: a review of 25 years of research. *Epilepsy Research* 1995;22:65–95. [PubMed: 8777903]

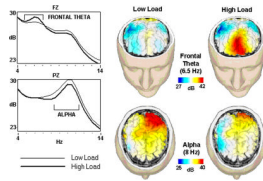


Fig. 1.

As the working memory task gets more difficult, EEG frontal theta power increases and parietal alpha power decreases. (Left): 4–14Hz EEG power at frontal (Fz) and parietal (Pz) midline sites measured during an easier (low load -- light line) and more difficult (high load -- dark line) n-back WM task from 80 participants (adapted from Gevins and Smith, 2000). (Right): 120-electrode WM task EEG images for a single participant, mapped onto a cortical model derived from the participant's MRI (adapted from Gevins et al., 1996). Although such dense electrode arrays improve topographic mapping, practical cognitive neuromonitoring can be accomplished with far fewer electrodes.

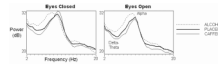


Fig. 2.

EEG power spectra from the midline parietal-occipital electrode recorded from 16 healthy adults in resting eyes closed (left) and eyes open (right) conditions, 1.5 hours after administration of alcohol (dotted line), placebo (dark solid line), or caffeine (light solid line). Across both resting conditions, low frequency EEG power in the delta-theta band was larger after alcohol but smaller after caffeine. The ratio of alpha power in the eyes-closed to eyes-open conditions was smaller after alcohol but larger after caffeine. Such changes contributed to the alertness sub-score decreasing after alcohol ($p < .05$) but increasing after caffeine ($p < .01$).

Table 1

Table 1a. Between-day, within-day and total variability of SWMT overall score, and performance, cortical activation and alertness sub-scores, in the absence of a treatment. Data from 127 participants in seven studies (see text)

		Between-day	Within-day	Total Variability
Overall Score	# tests	279	362	641
	M	-0.15	0.00	-0.07
	sd	0.62	0.54	0.58
	CI (95%)	± 0.07	± 0.06	± 0.04
Performance	M	-0.09	0.07	0.00
	sd	0.75	0.63	0.69
	CI (95%)	± 0.09	± 0.07	± 0.05
	M	-0.01	0.19	0.10
Activation	sd	0.86	0.91	0.90
	CI (95%)	± 0.10	± 0.09	± 0.07
	# tests	245	325	570
	M	-0.28	-0.28	-0.28
Alertness	sd	0.89	0.72	0.79
	CI (95%)	± 0.11	± 0.08	± 0.07

Table 1b. Between-day and within-day variability of the overall score in each of the seven studies

	Between-day variability			Within-day variability		
	Mean	sd	# tests	Mean	sd	# tests
Study1	-0.25	0.74	68	-0.04	0.59	63
Study2	0.03	0.49	29	0.14	0.42	58
Study3	-0.39	0.56	58	-0.06	0.51	86
Study4	0.07	0.55	9	0.06	0.59	72
Study5	-0.02	0.64	47	-0.14	0.58	48
Study6	-0.01	0.46	50	--	--	--
Study7	-0.11	0.63	18	-0.06	0.62	30

M = mean; sd = standard deviation; CI = confidence interval. The mean values are all approximately zero with standard deviations less than one.

M = mean; sd = standard deviation; CI = confidence interval. The mean values are all approximately zero with standard deviations less than one.

Table 2

SWMT overall score and sub-scores following ingestion of caffeine, alcohol, or diphenhydramine, or after staying up all night. Values for placebo at the time of t_{peak} for the drugs, and at 11PM for the sleep deprivation condition, are shown for comparison. Significance was assessed for the drugs by reference to the Within-day variability values in Table 1a, and for sleep deprivation by reference to the Between-day variability values in Table 1a.

		Caffeine	Alcohol	Placebo (1.5 hrs)	Diphenhydramine	Placebo (2.5 hrs)	Sleep Loss	Control (11PM)
Overall Score	M	0.41 ^{***}	-0.34 [*]	0.01	-0.87 ^{***}	-0.23	-0.88 ^{***}	0.00
	sd	0.43	0.62	0.95	0.66	0.95	1.07	0.62
Performance	M	0.36	-0.06	0.20	-1.02 ^{***}	-0.03	-1.16 ^{***}	0.15
	sd	0.57	0.45	1.42	1.04	1.77	1.87	0.76
Activation	M	0.42	-0.09	0.28	-0.04	0.13	-0.31	-0.37
	sd	0.92	1.14	1.13	0.83	1.15	0.86	0.65
Alertness	M	0.33 [*]	-0.74 [*]	-0.10	-1.20 ^{***}	-0.17	-1.07 ^{**}	-0.01
	sd	0.54	0.86	0.86	1.00	0.71	1.02	0.92

* p<.05,

** p<.01,

*** p<.001