Electroencephalogram (EEG) Study of Learning Effects across Addition Problems

James Fielder
US Army Aberdeen Test Center
Aberdeen Proving Ground, MD 21005
james.fielder@us.army.mil

Introduction

Evolving military systems have the potential to inundate Soldiers with complex information and overwhelming tasks. As the design of modern military equipment continues to push the boundary of human mental capacity, more and more pieces of equipment are vying for the cognitive resources of the Soldier. When the Soldier becomes overwhelmed, their task performance suffers and mission effectiveness is compromised.

Testers and evaluators need a way to measure the equipment's effect on the Soldier and their performance. The US Army Aberdeen Test Center (ATC) has begun using electroencephalogram (EEG) headsets to quantitatively measure a Soldier's cognitive performance. Although EEG technology has been used for over 30 years[i] in laboratory experiments to measure mental workload, its use in test and evaluation (T&E) requires several obstacles to be overcome.

First of these obstacles is the motion-filled T&E environment. Very rarely will testing be conducted in a laboratory with a static subject. Routinely, Soldiers are outside a controlled environment; walking, crawling, jumping, running, driving, firing, and talking. All of these conditions introduce motion artifacts, making data collection difficult for most EEG headsets.

Second, test events are always pressed for time. Very rarely will testers have the luxury of taking up to an hour to instrument each Soldier using conventional EEG headsets. Ideally the EEG headset should be a plug-and-play device requiring only minimal external inputs for use. There should be little preparation time required as testers need to deploy many EEG headsets concurrently, limiting the amount of time available for adjusting each EEG headset.

Lastly, testers need to rely on technicians to operate the EEG headsets and evaluate the data. Laboratory and research experiments rely on well-trained professional neuroscientists and cognitive psychologists; however, testers need to rely on technicians for the widest deployment of the EEG headset.

In 2004, ATC began work with Quantum Applied Science and Research (QUASAR) to develop an EEG headset[ii] (Figure 1) for use in a rugged T&E environment. The headset developed consists of nine dry contact electrodes which minimize preparation time as there is no need for skin preparation and the application of conductive gels. It uses a common mode follower electrode to actively measure common mode noise, such as static electricity. The common mode noise is then extracted from the raw data and filters are applied remove other sources of noise such as interfering signals from power lines. Sophisticated algorithms are then applied to the processed EEG data to provide measures of mental workload, engagement and fatigue. The software also requires little training to operate the headset, (allowing for technician use) and the data generated is in a form comprehensible by a technician-level analyst with minimal training.

Figure 1: The QUASAR C2 Headset

The first objective of this investigation was to validate the approach proposed herein for follow-on use in a study to determine the quality of the mental workload data obtained from the EEG headset as compared to data obtained from National Aeronautics and Space Association (NASA) Task Load Index (TLX). The second objective was to investigate the use of the EEG headset as a training quality evaluation instrument.

Method

Three subjects were chosen at random to participate in the study. They were civilian engineers from ATC. All three were males and ranged in age from 21-27 years. All three were invited to an isolated test chamber for the study. The subjects were told of their rights as a subject and asked if they wanted to participate.

First the the QUASAR EEG headset was installed according to the manufacturer's procedures.[iii] The subjects were then given a brief tutorial of the addition tasks used to manipulate workload. The one, two, or three column addition tasks were provided via the Psychology Experiment Building Language (PEBL) [http://pebl.sourceforge.net/]. For each trial, the subjects were given an opportunity to complete 10 sample questions before proceeding to recorded questions. For each trial, the subjects had to complete 45 questions or complete questions for 7 minutes, which ever happened first.

Four trials were collected for each of the three levels of addition problems. See table 1 below for the order and duration of testing procedures.

Table 1: Order and duration of test events

Review of Subject's rights	3 minutes
Donning of the EEG headset	15 minutes
Overview of NASA-TLX and a baseline	5 minutes
Overview of PEBL: Addition Problems	2 minutes
1 column: Trial 1 → NASA-TLX → break	5 minutes
1 column: Trial 2 → NASA-TLX → break	5 minutes
1 column: Trial 3 → NASA-TLX → break	5 minutes
1 column: Trial 4 → NASA-TLX → break	5 minutes
Break	15 minutes
2 column: Trial 1 → NASA-TLX → break	11 minutes
2 column: Trial 2 → NASA-TLX → break	11 minutes

2 column: Trial 3 → NASA-TLX → break	11 minutes
2 column: Trial 4 → NASA-TLX → break	11 minutes
Break	15 minutes
3 column: Trial 1 → NASA-TLX → break	11 minutes
3 column: Trial 2 → NASA-TLX → break	11 minutes
3 column: Trial 3 → NASA-TLX → break	11 minutes
3 column: Trial 4 → NASA-TLX → break	11 minutes
Doffing the EEG headset	1 minute
Total Elapsed Time	164 minutes

All four of the one column add trials were administered before the two column add trials. The same procedure was followed for the two and three column add trials. A performance metric was calculated by dividing the number of answers correct by the number of questions answered for each trial. The average time to complete an addition problem was treated as a time metric. EEG data was collected during each trial.

At the end of each trial, the subject was asked to complete the 6 question NASA-TLX[iv] questionnaire using ATC's eQuestionnaire. The eQuestionnaire is a computer-based survey administration tool, and is effective and reliable means for administering multiple surveys. A brief tutorial of the eQuestionnaire system was given and a baseline NASA-TLX was collected before data collection began.

After the last trial, the EEG headset was removed, 164 minutes from when the headset was first donned. The EEG data was then processed using QUASAR's QStates software.[v] QStates software provides workload data using two different statistical models: multivariate normal probability density function (MVNPDF), and a linear model. The workload model consists of an algorithm that identifies characteristics and features of an EEG that differ between high and low mental workload states. The use of the model automates the data reduction from raw EEG (240 Hz) to workload values (0.5 Hz). A workload model was created[vi] for each subject using the fourth trial of the one column addition task as the low workload state and the first trial of the three column addition task as the high workload state. The individual subject's workload model was then applied to all 12 trials.

A repeated measures analysis was used on the resulting data. Parameter tests were conducted across subject, trial, and workload condition for each of the five response variables (NASA-TLX, MNVPDF, Linear, Performance, and Time),

Results

Table 2 presents the significant effects using repeated measures analysis* across the five dependent variables.

Table 2: Summary of the main effects and two-way interactions

Dependent Variable	Independent Variable	P value
NASA-TLX	Subject	0.010
	Workload	0.004
MVNPDF	Workload	0.001
	Trial*Workload	0.008
Linear	Workload	0.004
Performance	Workload	0.022
	Trial	0.022
Time	Workload	<0.001
	Trial	0.003
	Trial*Workload	0.012

*NOTE: "Linear" response uses "Univar G-G Epsilon" for parameter test, whereas all other responses use "Univar unadjusted Epsilon"

The subject variable for NASA-TLX was significant based on a 95% confidence level, whereas neither EEG-based workload measurement (MVNPDF and Linear) was significant. Therefore, for the 3 test subjects, NASA-TLX is a more subjective response than EEG-based workload. This is also clearly illustrated in the below graphs, with there being a tighter grouping of the subject lines.

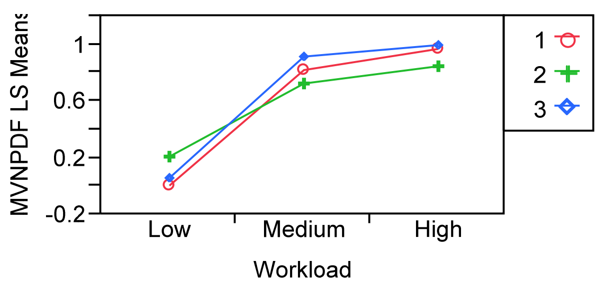

Figure 2: EEG-based workload scores (MVNPDF) across workload conditions.

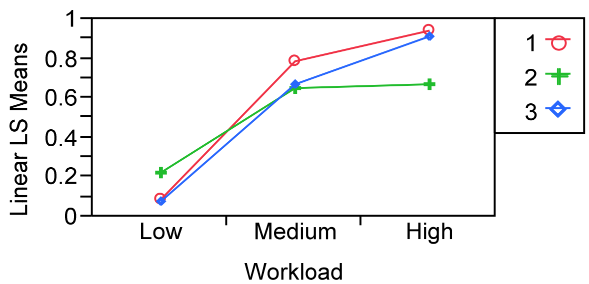

Figure 3: EEG-based workload scores (Linear) across workload conditions

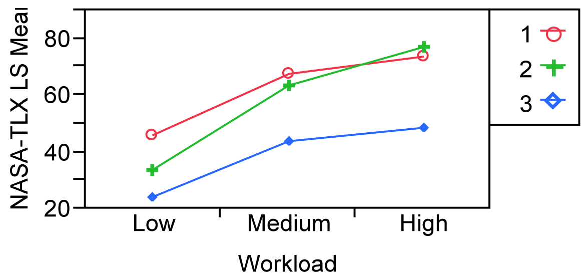

Figure 4: The NASA-TLX scores across subjects and workload conditions.

There was a significant effect across trials for both percentage of problems correct (Figure 5) and average completion time (Figure 6). Note the slight upward trend of percentage of problems correct and a more pronounced downward trend of average completion time, evidences the presence of training effects. There was a two-way interaction of trial and workload for MVNPDF, indicating that MVNPDF may show training effects across trials.

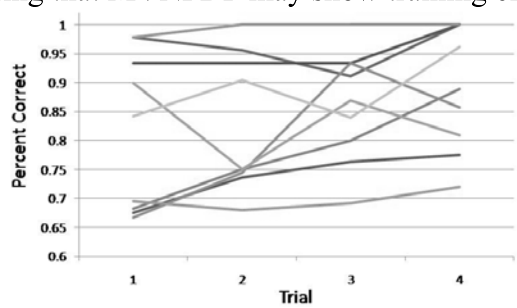

Figure 5: Percentage of Problems Correct Across Trials

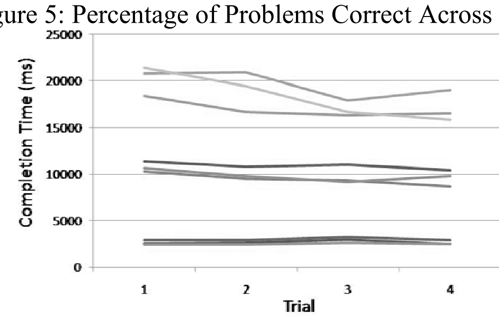

Figure 6: Completion Time Across Trials.

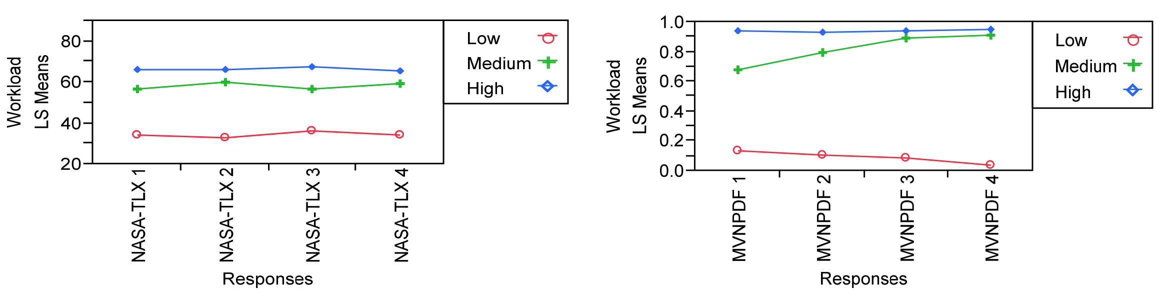

Figure 7: Effect of training as measured by NASA-TLX and MVNPDF across the four trials.

There is a greater change in MVNPDF response over trials than NASA-TLX.

Discussion

Both the NASA-TLX and the EEG-based measures showed the ability to discriminate between the workload conditions. The EEG-based measures were better discriminators than the NASA-TLX. Learning effects were observed with the subjects getting a higher percentage of problems right over the trials and decreasing their completion time.

EEG-based workload measures showed superior discriminatory power over the NASA-TLX. The NASA-TLX is hampered by biases injected from the subject's internal rating system. The EEG-based measures removes the biases and applies everyone's rating using the same scaling markers.

The training effects were present but mainly among the performance metrics (percentage of problems right and completion time). From the observed results, it follows that learning effects did take place and that the subjects were not given enough practice time to achieve the desired results. In this investigation, the subjects had at most 30 minutes under each workload condition. This appears to be insufficient time to demonstrate that the subjects reached training proficiency. It is postulated that if training trials were longer and spread across multiple days, the results would be more definitive.

ATC intends to continue to investigate the potential of this EEG headset as a candidate instrument for T&E programs. A larger sample size (≥15) is needed to more accurately depict the discriminatory ability of the EEG-based measures of workload. Training effects will also continue to be examined as a way to evaluate the effectiveness of military training programs. The next investigation will designed as a statistically rigorous study that will incorporate longer trials and more time between trials to better discern and quantify training effects. Before ATC can use this EEG headset in T&E applications, it must demonstrate it can accurately collect workload data while the subject is on-the-move. This demonstration will involve the subject completing simple mental tasks while sitting and walking, then comparing the workload data against each other.

Acknowledgements

The author would like to give special thanks to the following people and their organizations for making this investigation possible. Pamela Burton of the US Army Research Laboratory for the use of her laboratory space. Paul Roche of the US Army Evaluation Center for his help in analyzing the data. Lastly, and most importantly, Stephen Casey of the US Army Aberdeen Test Center for his help in instrumenting the three subjects.

---

[i] Rolfe, J.M., & Lindsey, S.J.E.. (1973). Flight deck environment and pilot workload: Biological measures of workload. *Applied Ergonomics*, *4*(4), 199-206.

Gevins, A., & Smith, M. E. (2003). Neurophysiological measures of cognitive workload during human-computer interaction. *Theoretical Issues in Ergonomics*, *4*(1-2), 133-131.

Berka, C., Levendowski, D. J., Cvetinovic, M. M., Davis, G., Lumicao, M. N., Zivkovic, V. T., Popovic, M. V., & Olmstead, R. (2004). Real-time analysis of EEG indexes of alertness, cognition, and memory acquired with a wireless EEG headset. *International Journal of Human-Computer Interaction. 17*(2), 151-170.

[ii] Matthews, R., Turner, P. J., McDonald, N. J., Ermolaev, K., McManus, T., Shelby, R. A., & Steindorf, M. (2008). Real time workload classification from an ambulatory wireless EEG system using hybrid EEG electrodes. Proceedings from the *30th Annual International IEEE EMBS Conference*. Vancouver, Canada.

[iii] QUASAR. (2009). *EEG C2 Headset System: User Manual.* San Diego, CA: QUASAR.

[iv] National Aeronautics and Space Administration. (1986). *Task Load Index (NASA-TLX).* Moffett Field, CA: NASA.

[v] QUASAR. (2009). *QSTATES, Data Classification Module: Technical Manual.* San Diego, CA: QUASAR.

[vi] QUASAR. (2009). *QSTATES, Data Classification Module: User Manual.* San Diego, CA: QUASAR.