REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)* 12-04-2011	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)*

4. TITLE AND SUBTITLE Test Operations Procedure (TOP) 01-1-016 Electroencephalogram-Based Measurement of Workload, Engagement, and Fatigue	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHORS	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Soldier Systems Division (TEDT-AT-WFS) U.S. Army Aberdeen Test Center 400 Colleran Road Aberdeen Proving Ground, MD 21005-5059	**8. PERFORMING ORGANIZATION REPORT NUMBER** TOP 01-1-016
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Test Business Management Division (TEDT-TMB) U.S. Army Developmental Test Command 314 Longs Corner Road Aberdeen Proving Ground, MD 21005-5055	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) Same as item 8

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A. Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
Defense Technical Information Center (DTIC), AD No.: ADA540923

14. ABSTRACT
This Test Operations Procedure (TOP) describes the procedures for using electroencephalogram (EEG) instrumentation for measuring cognitive workload, cognitive engagement, and mental fatigue and also the capabilities of the instrumentation.

15. SUBJECT TERMS

EEG	workload
cognitive	mental fatigue

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**B. ABSTRACT**	**C. THIS PAGE**			**19b. TELEPHONE NUMBER** *(include area code)*
Unclassified	Unclassified	Unclassified	SAR	20	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18

U.S. ARMY DEVELOPMENTAL TEST COMMAND
TEST OPERATIONS PROCEDURE

*Test Operations Procedure 01-1-016 12 April 2011
DTIC AD No. ADA540923

ELECTROENCEPHALOGRAM-BASED MEASUREMENT OF WORKLOAD, ENGAGEMENT, AND FATIGUE

1. SCOPE.

This Test Operations Procedure (TOP) describes the procedures for using electroencephalogram (EEG) instrumentation for measuring cognitive workload, cognitive engagement, and mental fatigue, and also the capabilities of the instrumentation.

1.1 Purpose.

The use of this type of instrumentation in Test and Evaluation (T&E), as of this writing, is still in its infancy. Many of the analysis procedures and processes are currently being developed; however, the procedures to collect the data described in this document have not changed for many years. Also stated is the current hypothesis of how to analyze and display the data in a manner relevant to the evaluator.

1.2 Background.

a. Cognitive workload, engagement, and fatigue measurements have been used in T&E for several years using questionnaires. These measurements can be defined as the following:

(1) Cognitive workload is the amount of mental resources a person exerts during a specific task.

(2) Cognitive engagement is the amount of attention a person exerts during a specific task.

(3) Mental fatigue is the cumulative sum of mental workload over time.

b. The limitation of this method stemmed from the fact that the T&E community requires continuous measurements in near-real time. Questionnaires do not meet this requirement. Only recently have advancements in instrumentation technology allowed measurements of objective workload, engagement, and fatigue to be made in a military testing environment. By using electroencephalography; objective, continuous, and near real-time data can be collected in an unobtrusive manner. Modern configuration of EEG instrumentation and associated software allows for technicians, rather than neuroscientists, to use the instrumentation and interpret the results.

1.3 Limitations.

a. The use of EEG is one tool in a tester’s toolbag. It provides global answers to questions about workload, performance bottlenecks, and mental fatigue. By using EEG data, testers can identify periods of high workload, signs of mental fatigue, and symptoms of boredom. Testers can then trace back, using Soldier task recordings, to determine what tasks affect the test participant's (TP's) workload and/or fatigue.

b. Translating EEG recordings of workload, engagement, and fatigue into task performance and mission performance requires the use of Soldier task recordings. Without the

Soldier task recordings, the usefulness of EEG recording in T&E is severely limited because causal factors remain unknown.

2. FACILITIES AND INSTRUMENTATION.

2.1 Facilities.

Use an appropriate testing/operational environment for the military equipment being tested.

2.2 Instrumentation.

a. The following options are recommended when choosing an EEG system:

(1) Dry electrodes. When using dry electrodes, the metal of the electrode make contact directly to the skin, as opposed to wet electrodes which uses a saline-based gel to conduct the signal to the electrode. Dry electrodes do not require significant setup time and there is no gel that needs to be washed out after use.

(2) Wireless communications. Wireless communication from the EEG headset to the data collection base station results in the least obtrusive data collection; however, due to Information Assurance (IA) and security measures a wired solution is sometimes necessary.

(3) Worn underneath a helmet. Not all tests will be conducted in a stationary environment. For added realism and to answer safety concerns, a safety helmet or combat helmet has to be worn over the EEG headset when used by a TP on a test range.

(4) Automated software. Highly-trained neuroscientists are usually needed for removing noise artifacts and analyzing data. Automated software can allow for a technician to complete the same process.

(5) Common mode follower. Useful in a noisy or motion-filled environment, the common mode follower is a system using an isolated electrode that collects the noise signals and then subtracts them from the rest of the electrode signals.

(6) Programmable algorithms. Algorithms take the EEG waveforms from the individual electrodes, filter them and transform them into workload, engagement, and fatigue data. Each TP, task, and test environment combination requires a unique algorithm for peak performance and the most reliable, valid data. Since there are many different TPs, tasks, and test environments, the resulting combination of each is an incalculable number of specific algorithms. The software should allow the algorithms to be easily programmable to meet the specific testing conditions of a test.

b. EEG Specifications.

Item	Requirement
Data epoch	≤ 2 seconds/data point
Data loss	≤ 1 percent
Sample rate of raw EEG	10x higher than phenomena of interest (usually 200 hertz ((Hz))
Data interface	Common interface to a computer
Signal to noise ratio	≥ -6 decibels (dB)
Algorithm reliability	≥ 90 percent
Algorithm validity	Correlation ≥ 5

3. REQUIRED TEST CONDITIONS.

a. Data Requirements.

The first step of test planning is to determine what needs to be tested and what data need to collected. These requirements are usually found in the System Evaluation Plan (SEP) or the Test and Evaluation Master Plan (TEMP). Additionally, working with the evaluators from the US Army Evaluation Center (AEC), the data requirements matrix and data requirements defining the test requirements can be refined.

b. After the test requirements are agreed upon, test planning can begin. During this stage, the testing environment, tasks to be completed, test instruments, TPs, and data analysis will need to be determined. All of this needs to be outlined in a Detailed Test Plan (DTP).

(1) Testing environment. Accurate workload, engagement and fatigue measurements rely on providing the TP with an operationally realistic testing environment and tasks. The environment should reflect the intended environment the system-under-test (SUT) will be used in. For example, if the workload required for operating a wrist-mounted computer is being tested, then the TP can be placed in a Military Operations on Urban Terrain (MOUT) environment or out in the woods instead of at a desk in a comfortable building.

(2) Tasks to be completed. Like the testing environments, the tasks the TP will complete have to be realistic and representative of how the end user will use the SUT. For example, if the test is determining the workload of a small unmanned ground vehicle (SUGV) operator, then the operator should be controlling the SUGV in a realistic environment while sending pictures to higher headquarters (HQ), providing local security, and monitoring team communications.

(3) Test instruments. Besides the EEG instrumentation and the associated data collector, the test instruments refer to instrumentation to capture data related to the Soldier's tasks. For example, if testing Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance (C4ISR) equipment, a network sniffer is appropriate for recording message traffic and times. If testing in a dismounted environment, Global Positioning System (GPS) instrumentation and real-time casualty assessment (RTCA) instrumentation will help determine what the TP is doing.

(4) Test participants. As stated earlier, realism is essential to accurate workload, engagement, and fatigue measurements. Besides selection of environment and tasks, proper selection of TPs needs to be considered. The Military Occupation Specialty (MOS) of the Soldier that will be the end user of the SUT needs to also be the same as the TP. Also important, is a statistically significant number of Soldiers to be TPs in the test. While many statisticians will advocate for sample sizes of 30 or more (since it reduces measurement error), limitations of scheduling and cost place realistic sample sizes at 4 to 6 participants. This reduction in the number of TPs may lead to nonsignificant results even if there is in fact a significant differential effect. This in turn results in increased measurement error.

(5) Data analysis. The amount/kind of Soldier task data collected will determine the analysis conducted. The workload, engagement, and fatigue measurements should be compared to performance metrics taken from the Soldier's task data. EEG data should be reduced to 1- to 2-second epochs and directly compared to performance measures.

c. All of the above considerations for the test should be documented in the DTP. This test plan will be the basis for test conduct and associated test paperwork (i.e., safety, human use, certification, environmental, etc.). Several of these are explained below.

d. Safety, human use, and instrumentation certification documentation all share the same purpose: to ensure that the TP and the test itself are safe and properly conducted. The main difference for these three documents is the approval authority.

(1) Safety documentation, usually generated by the test center, determines if standard testing procedures are being followed and documented and planned for all safety concerns. A Risk Assessment and Job Hazard Analysis must be generated and approved by the Hazard Analysis Working Group (HAWG), and by the Test Center’s Commander or Technical Director.

(2) Human Use documentation determines that the TP does not suffer any adverse effects from the test. The US Army Developmental Test Command (DTC) coordinates the Human Use documentation with the US Army Medical Research and Materiel Command (MRMC), who are the approving authority due to their medical background. DTC will provide guidance and oversight in accordance with DTC memorandum, Human Research Protections Program (HRPP)[1].

*Superscript numbers correspond to Appendix B, References.

(3) Instrumentation certification documents the functionality and use of the instrumentation and its noninterference with the SUT. Approval authority lies at the US Army Test and Evaluation Command (ATEC) level, pending concurrence from DTC, MRMC, and AEC. Part of the process of instrumentation certification is documenting the noninterference with other electrical signals present at the test site. This is carried out with electromagnetic interference (EMI) testing in an isolating chamber.

e. Determining the Soldier task measurements to be collected at the start of the test planning phase will aid in the process of test planning and data analysis. Absence of Soldier task data and associated performance data will severely limit data analysis options and the conclusions that could be made from them. Very rarely will the collection of workload, engagement, and fatigue be conducted in an isolated test; it will most likely be collected as part of a bigger test. When part of a larger test, the importance of Soldier task measurements seems to slip resulting in reduced collection and lower fidelity. To keep Soldier task measurements relevant and important when part of a larger test, testers must, very early in the test, work with the test planners to make sure the Soldier task data are collected and to the fidelity wanted. The following is a list of measurement variables that need to be determined when measuring Soldier task measurements.

(1) Time resolution. This refers to the smallest increment of time at which data are collected. For example, if the EEG system presents data at a 2-second resolution, the Soldier task measurements should be at least a 1-second resolution. It is not very useful if the Soldier task measurements have the accuracy of $\pm$ 10 minutes if the EEG data have a 2-second resolution.

(2) Time synchronized. Since data will be collected by multiple computers and data tend to be coded to the computer's time clock, the clocks should all be synchronized. Usually, if all the computers are on a common network, the clocks can be synched by having the computers use network time.

(3) Time Ordered Events List (TOEL) logs. Many tests involving human performance, which will include measurement of workload, engagement, and fatigue are scripted events. This implies that the stimuli are presented to the TPs at known points, and correct TP responses are known beforehand. A log of when all the events are presented is useful for determining which tasks the TP is completing at which time.

(4) Log file format and location. Log files from simulations, network sniffers, and keystroke loggers are difficult to understand in their native format. Usually trained programmers need to run the log file through a data parser to extract the relevant data. This process needs to be finalized before the test is conducted and a determination of who will be doing the parsing must be made.

(5) The last step before the test is to make sure the EEG system is calibrated. This is usually done by passing a known signal through the electrodes and matching the known signal to the recorded signal. This can be accomplished by a metal shaped head and a frequency generator. The frequency generator touching the head will cause the EEG electrodes touching the head to receive that frequency. EEG signals are in the µV range, if the correct scale frequency generator is not used, the electronics in the instrumentation will become damaged.

4. TEST PROCEDURES.

This section will provide recommended test procedures for collecting EEG-based cognitive workload, engagement, and fatigue measurements. Other tests and tasks, not originally planned for, or conceived at the time of this writing, may change the order or completion of the following procedures.

a. Greet the TPs and provide them with a volunteer agreement, which states that they were not coerced into participating and they have the right to withdraw from testing at any time. The volunteer agreement is usually a major part of the human use approval. After the TPs sign the volunteer agreement, they will be collected.

b. After the volunteer agreements are signed, collect demographic and background information on the TPs. This can include, but is not limited to the amount of SUT training, use of the legacy version of the SUT, MOS, gender, amount of combat, and years in service. This type of data tends to explain the differences of workload, engagement, and fatigue across the TPs. For example, a highly trained Soldier will experience lower workload and/or same workload with better performance than a Soldier that is less trained when performing a command and control task.

c. It is recommended that an EEG communication check be completed before the EEG headset is installed. This includes the complete setup. Even without the EEG headset on a TP, the EEG headset should be picking up ambient noise that can be used to determine if the EEG headset is functioning properly.

d. Fit the EEG headset to the TP's head. The headset should be securely fastened about the head so it does not move as the TP moves, but it should not be so tight that it causes discomfort after a few minutes. If the EEG headset does cause discomfort, either loosen the headset or remove it - the symptoms should disappear almost immediately. A mechanism should ensure the EEG headset and corresponding sensors are held down on the head.

e. Once the headset is in place, the associated sensors can be placed on the body. The sensors usually consist of electrocardiograms, electroculograms, accelerometers, galvanic skin response sensors, thermal sensors and respiratory straps. Each sensor has a different position on the body and configurations between EEG headset manufacturers are set up differently. Place the sensor according to the headset manual and follow recommended procedures.

f. After the headset and associated sensors are in place, a communication test between headset and data collector should be conducted. This ensures the communication link is established and that there are no lost packets. At this point it needs to be decided if the communication between the headset and the data collector will be wired or wireless. Wireless communication allows for unencumbered movement and greater distance between the two. Wired setups would be required if IA has disallowed use of the wireless function. Use of wired or wireless communication is also dependent on the choice of EEG headset available.

g. Begin the impedance test of the headset. This makes sure the resistance between the electrodes and the TP's scalp is at a minimum. Every EEG headset is different. The manual should be checked for the proper impedance values.

(1) Impedance values are decreased by several methods;

(a) Apply more pressure to the sensor to make contact to the scalp.

(b) Part the hair away from where the electrode will be contacting the scalp.

(c) Apply a saline-based gel or manipulate the electrodes to make better contact with the skin.

(d) Wait while impedance values drop over the first 10 to 20 minutes.

(2) Several EEG headset models require abrading the skin for electrode contact. This requires using a needle to scrape the outermost levels of skin cells (these cells are the least conductive). This step is usually common with highly sensitive medical research - grade EEG systems. These types of EEGs are not commonly used for military operational testing, since they are extremely sensitive to noise, are most effective in stationary laboratory environments, and usually contain many sensors and wires resulting in a long preparation time.

h. After the impedance has reached an acceptable level, it is time to generate a task-specific, individual-specific baseline. This will train the algorithm. The baseline should be conducted according to manufacturer's instructions. Several models have prescribed baseline procedures; others will have the software train on known files. For example, if a C4ISR evaluation is being completed, a 1- to 2-minute high workload file could be captured when the TP is bombarded with tasks and a 1- to 2-minute low workload file when the TP is not busy or even waiting for tasks. Use these two files to train the algorithm. This results in a highly specific algorithm for that individual TP and that individual task.

i. After baseline procedures, the EEG headset system is ready to be used. Collect the data for each test run. Make sure to note the start and stop times for each data file along with TP and event information (Soldier tasks recordings). Several of the EEG headsets allow for a real-time monitoring feature which allows for the tester to monitor the signals and determine if a sensor comes loose (or if a TP falls asleep). Each raw EEG data file should be processed through the associated algorithms and software to determine workload, engagement, and fatigue across the run.

j. After data collection, the headset is ready to be removed and cleaned for later use. If the headset uses saline-based gel in the electrodes, the gel has to be cleaned soon after use with soap and water; be careful not to scratch the electrode coating. Also remove any gel from elsewhere on the headset. When using dry electrodes, they need to be cleaned with a brushing of isopropyl alcohol. Remove any wet-adhesive electrodes from the body and any tape holding wires down. Then, make sure the headset is stored properly and batteries are charged for next time.

k. Data reduction and preliminary analysis should be conducted using software that comes with the headset. However if not available, after-market programs are available for use (i.e., EEGLAB). The software should remove any motion artifacts and noise from the measurements before any further analysis is completed. The software should use the associated task/subject-specific baseline (paragraph 4.h) to create the EEG-based cognitive workload, engagement, and fatigue measurements from EEG data. Usually EEG data can only be interpreted by trained neuroscientists and cognitive psychologists. The software allows technicians or nonspecific engineers to interpret the data. The EEG-based cognitive workload, engagement and fatigue measurements are the starting point for the data analysis described in this TOP.

5. DATA REQUIRED.

There are two types of data that are needed for data analysis of the EEG based workload, engagement, and fatigue data and the task performance data. They are explained below:

a. The EEG based workload, engagement, and fatigue data should have the data from all the sensors reduced to data points every 1 to 2 seconds. The data should be on a linear scale, with high and low states being the extremes of the scale.

b. Task performance data should be concurrent with the EEG data. Each of the task events should be tagged with the appropriate event timestamp, for comparison against the EEG data's timestamp.

6. PRESENTATION OF DATA.

This section will provide a starting place for the analysis and presentation of EEG-based cognitive workload, engagement, and fatigue measurements. Other tests and tasks, not originally planned for or conceived at the time of this writing, may change the analysis and presentation of EEG-based measurements as well as collection of the measurements.

6.1 Cognitive Gauge.

The EEG-based measurements can be used independently to form a Cognitive Gauge. This gauge will have various colors for the different stages of load and overload across the tasks measured. This type of data presentation would have the most use for real-time data collection and analysis. For real-time measurements, the gauge can fluctuate with each data point collected. For post processing, averaging of the measurements across tasks will produce colors for individual tasks. The EEG-based measurements can be averaged across all tasks of a given SUT to determine the overall effect of the system on workload, engagement, and fatigue. This will show the impact of the task on workload and fatigue.

a. An example for cognitive workload: green would signify normal workload (up to 80 percent), yellow would signify above-average workload (from 80 to 95 percent), and red would represent overload (over 95 percent). The colors would be determined by the workload measured across the entire task. This will identify the periods of time (or tasks) that the subject was in overload and/or underload, and by relation, the task of the system that caused this state. Underload is caused when the TP has so little to do; they lose interest and disengage from the task. Often, workload may increase as they force themselves to engage in the task. Other times they may fall asleep from boredom.

b. An example for cognitive fatigue: green would signify little or no (up to 80 percent), yellow would signify above an average level of fatigue (from 80 to 95 percent), and red would represent severe fatigue (over 95 percent). The colors would be determined by the fatigue measured across the entire task.

c. A continuous status monitor based on real-time data is shown in Figure 1, and an average status across each task of a system test is shown in Figure 2.

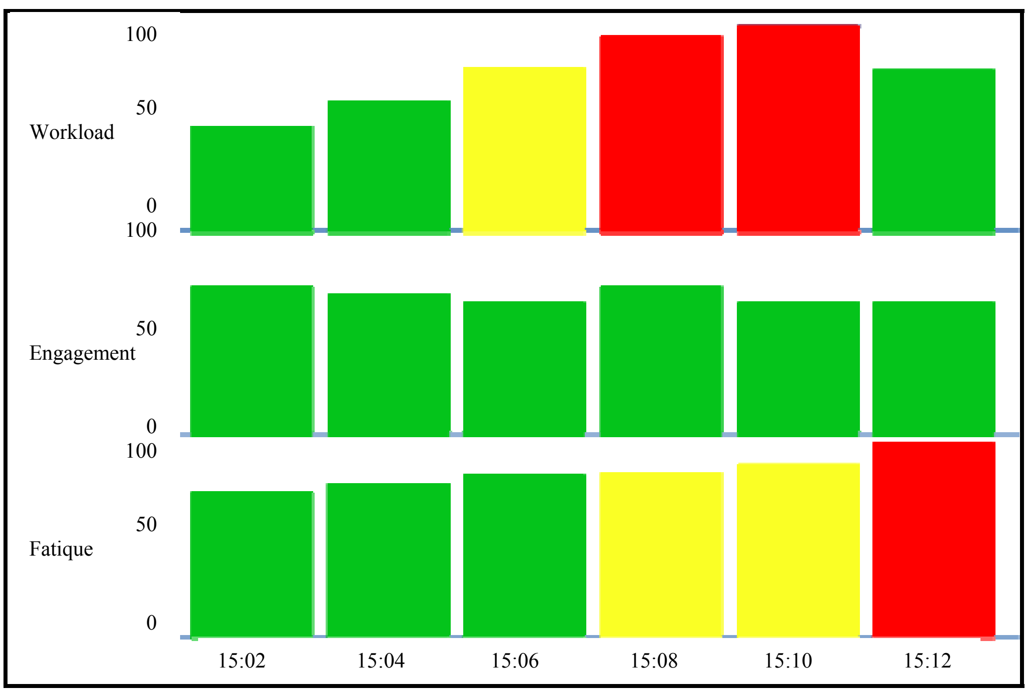

Figure 1. A continuous status monitor based on real-time data.

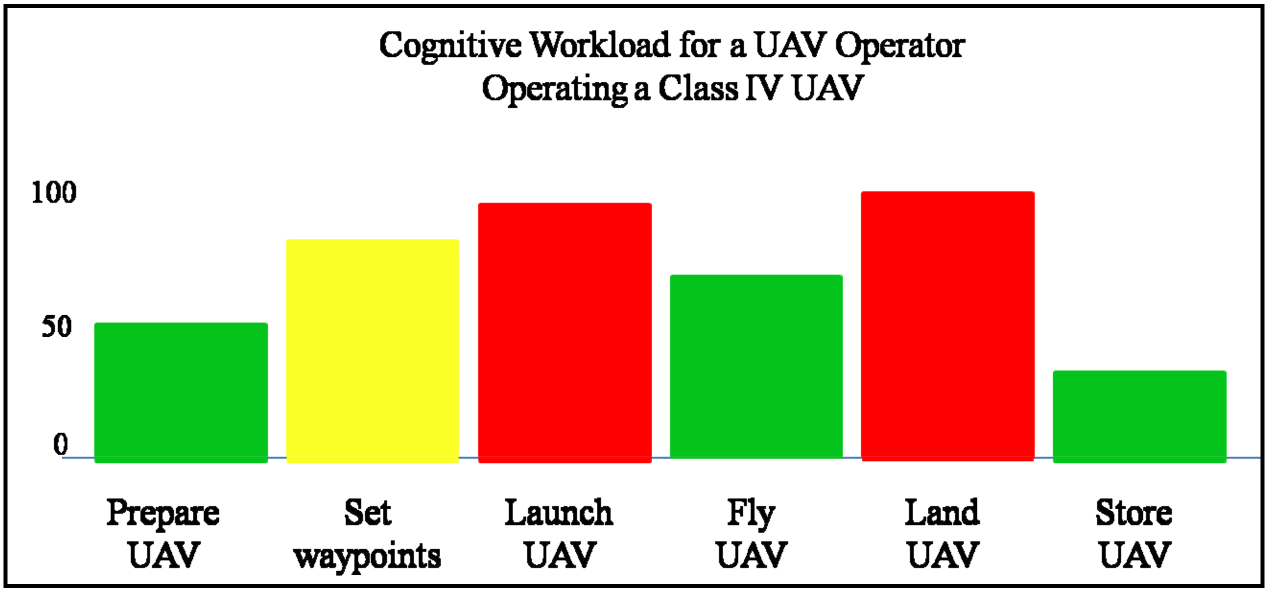

Figure 2. Average status across each task of a system test.

6.2 Comparative Analysis.

Comparing the EEG-based measurements from one SUT to another is another type of analysis. Having TPs complete similar tasks with both systems would allow for direct comparison of workload, engagement, and fatigue. Graphing the resulting measurements from both systems for each task would show the resulting difference between them (Figure 3). The data can be compared to find a percentage increase from one system to the next.

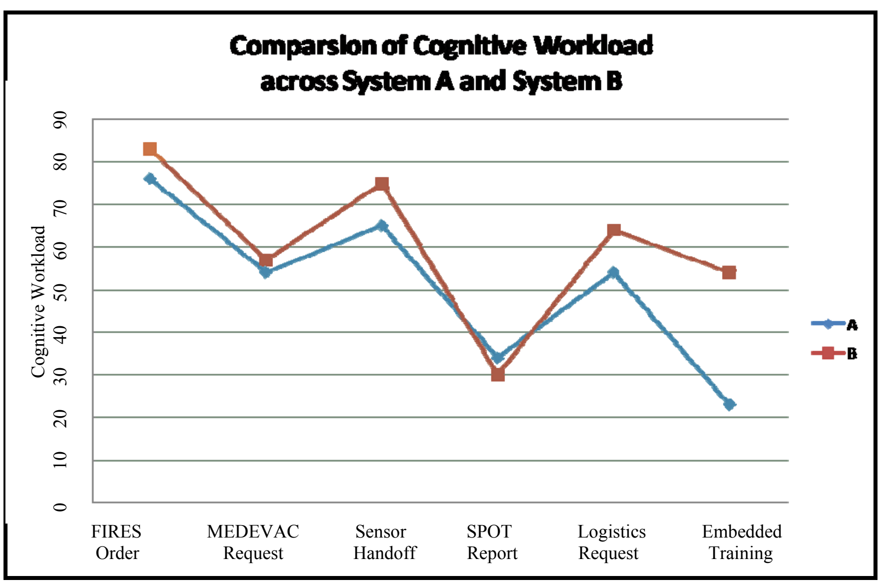

Figure 3. Using EEG measurements for system comparison.

6.3 Task Analysis, Bottleneck.

For determining direct relationships between task and EEG-based measurements, a task analysis needs to be performed to determine which tasks are being performed with the SUT. The task analysis would provide a breakdown of the tasks into their smallest observable part and calculate the associated workload, engagement, and fatigue. The task analysis would then identify bottlenecks and problem areas that could be examined using EEG-based measurements. The process could also be reversed using the task analysis to analyze specific areas identified using EEG-based measurements (Figure 4).

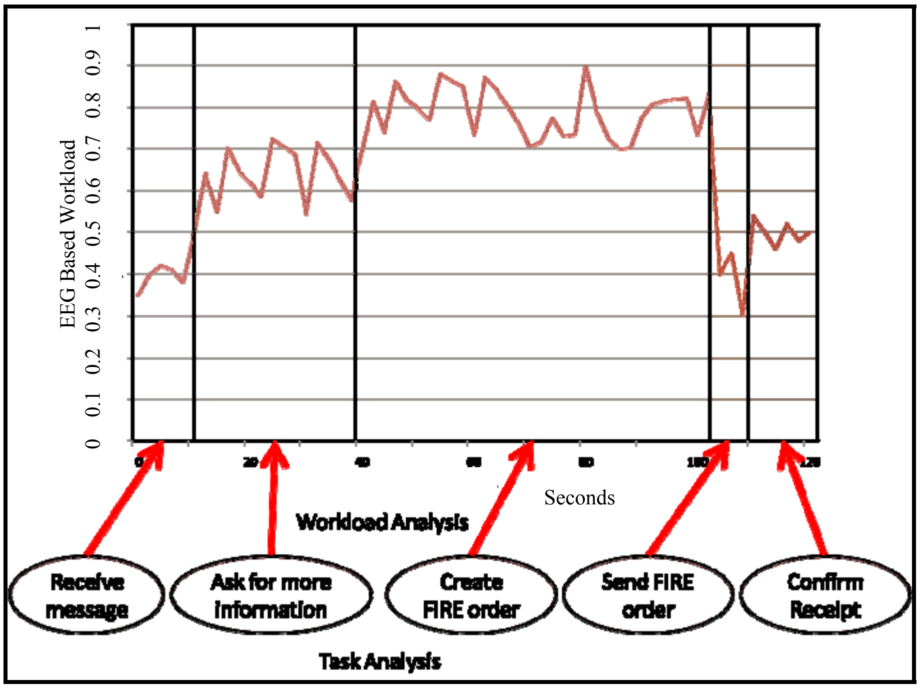

Figure 4. EEG-base task analysis for identifying bottlenecks and problem areas.

6.4 Performance Decrease, Index.

There is a documented link between cognitive workload and performance. As workload increases, performance should decrease. It might not be a measurable difference until the increase in workload reaches the overload state. However, the performance should be monitored to see when workload is impacting performance and mission effectiveness. A plot of performance and workload over time will show this. It will also show (when the test is long enough) a significant drop off in performance due to fatigue (Figure 5). It is at this point that the operator/subject should have been replaced with a new operator. For instances when performance metrics are not available, EEG-based workload measurements can be used as a predictor of performance. Citing the previous link with performance, workload values tend to correlate with performance and should be used as a predictor of performance when performance data cannot be collected.

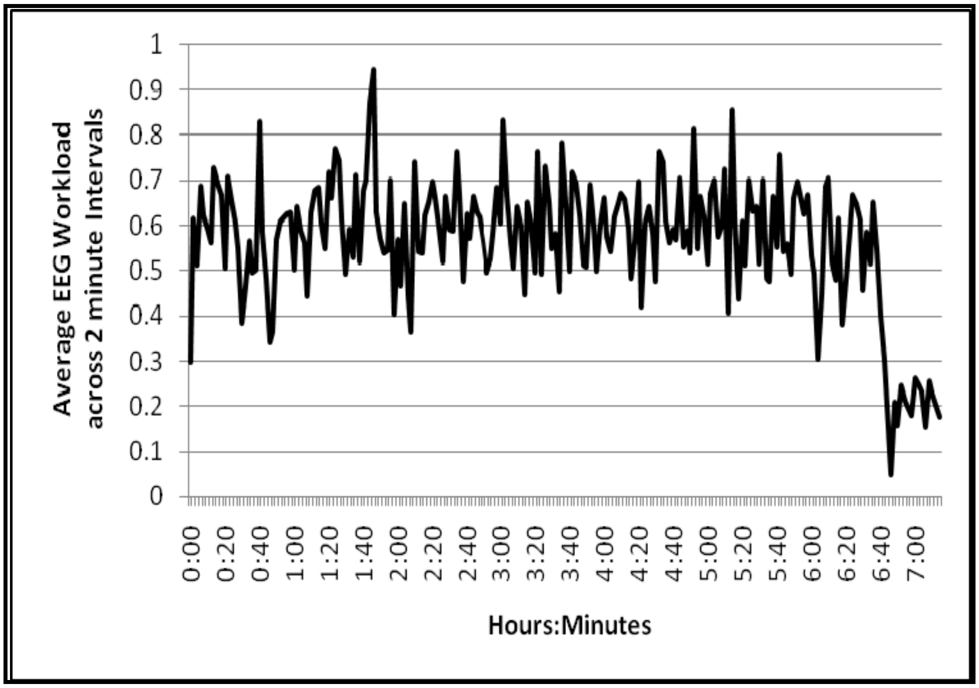

Figure 5. Using EEG-based measurements for fatigue studies.

APPENDIX A. ABBREVIATIONS.

AEC	US Army Evaluation Center
ATEC	US Army Test and Evaluation Command
C4ISR	Command, Control, Communications, Computers, Intelligence, Surveillance and Reconnaissance
dB	decibel
DTC	US Army Developmental Test Command
DTP	detailed test plan
EEG	electroencephalogram
EMI	electromagnetic interference
GPS	Global Positioning System
HAWG	Hazard Analysis Working Group
HRPP	Human Research Protections Program
HQ	headquarters
Hz	hertz
IA	Information Assurance
MOS	military occupation specialty
MOUT	Military Operations on Urban Terrain
MRMC	US Army Medical Research and Materiel Command
RTCA	real-time casualty assessment
SEP	System Evaluation Plan
SUGV	small unmanned ground vehicle
SUT	system under test
T&E	Test and Evaluation
TEMP	Test and Evaluation Master Plan
TOEL	Time Ordered Events List
TOP	Test Operations Procedure
TP	test participant

APPENDIX B. REFERENCES.

1. DTC Memorandum, Subject: Human Research Protections Program (HRPP), 22 April 2010.

Forward comments, recommended changes, or any pertinent data which may be of use in improving this publication to the following address: Test Business Management Division (TEDT-TMB), U.S. Army Developmental Test Command, 314 Longs Corner Road Aberdeen Proving Ground, MD 21005-5055. Technical information may be obtained from the preparing activity: Soldier Systems Division (TEDT-AT-WFS), U.S. Army Aberdeen Test Center. Additional copies can be requested through the following website: http://itops.dtc.army.mil/RequestForDocuments.aspx, or through the Defense Technical Information Center, 8725 John J. Kingman Rd., STE 0944, Fort Belvoir, VA 22060-6218. This document is identified by the accession number (AD No.) printed on the first page.