Neuro-optimization for Accelerated Learning Pace and Elevated Comprehension: Military Applications

Dr. JJ Walcutt	Dr. Cort Horton	Dr. Dhiraj Jeyanandarajan	COL Walt Yates (Ret.)
jjvogelwalcutt@yahoo.com	chorton@qneuro.com	djey@qneuro.com	walt.yates@gmail.com

ABSTRACT

The complexities of war will continue to increase exponentially as technology, digital connectivity, and all-domain warfare mature. While it will be necessary to ensure that our technology not only outpaces our adversaries, the increasing reliance on the ability of servicemembers to train faster and within more cognitively demanding battlespaces, will be equally important. Cognitive optimization, however, is hindered by natural human limits to knowledge and skill acquisition. The effects of overloading the brain with data and stress have been extensively studied and repeatedly shown to lead to constrained information intake and reduced focus and understanding. The importance of building programs aimed at accelerating cognitive capabilities within and across every fighting domain, balanced with the challenges and dangers of over-training, create the necessity of conducting training within a human-technology hybrid system, a requirement.

To date, substantial research has been conducted in personalized learning informed by both performance and neuro-physiological data. However, the data extracted from technology has been largely unreliable, inexact, and delayed. As both the validity and reliability have improved over the past 20 years, we are now reaching a point where the data that can be extracted has the potential to not only optimize the way the military designs training but also raise the cognitive advantage across the force. However, to be able to move neuro-data collection capabilities into training environments, portability of these systems must be increased, and their reliability must be established across a range of tasks and users. Accordingly, this paper compares the reliability of cognitive load measurements across different working memory tasks and reduced data channels to minimize cost and increase apparatus portability. Ultimately, the goal is to define an agreed-upon EEG metric for calculating load using optimal locations and the least number of sensors to define efficient and replicable assessment metrics and apparatus design.

ABOUT THE AUTHORS

JJ Walcutt, Ph.D. is the former Director of Innovation for the Advanced Distributed Learning (ADL) Initiative under the Office of the Secretary of Defense. She is also a former Human Innovation Fellow under the Office of Personnel Management focused on using design thinking to promote Federal Government transformation. Dr. Vogel-Walcutt has 20 years of experience in research and development for training and education with specific interests in applying instructional techniques to improve the effectiveness and efficiency of cognition and educational development.

Cort Horton, Ph.D. is the Director of Research and Development for Qneuro, where he uses his expertise in real-time measurement of cognition using EEG to develop neuro-adaptive applications for better education, training, and performance. He previously co-founded the Brain, Behavior, & Clinical Trial Outcomes Core of the UC Irvine Center for Autism Research and Treatment, which developed multi-method assessments of brain function and behavior in patients with Autism and related genetic disorders.

Dhiraj Jeyanandarajan, M.D., M.B.A. is a board certified neurologist with fellowship training in neurophysiology. He was a former clinical instructor of neurology at the University of Southern California Keck School of Medicine. He has 13 years of specialty experience in the application of neuromonitoring to measure the integrity of the nervous system during spine and brain surgery for improved outcomes. He is the founder of Qneuro, which applies similar neuromonitoring technology used in the medical space to applications for education and performance enhancement.

COL Walt Yates (Ret.) is the former Program Manager for Training Systems at Marine Corps Systems Command. He was a Secretary of Defense Executive Fellow and graduate of the Modeling, Virtual Environments, and Simulation

(MOVES) Institute at Naval Postgraduate School. He has 16 years of experience in the development, acquisition, and application of Modeling and Simulation systems for training and systems engineering. His special areas of interest are in designing resource loaded strategies for mission readiness and cost-effective application of cybersecurity policy for the learning enterprise.

DEMAND SIGNAL

Military Manpower is the product of a recruitment process that screens for mental, moral, medical, physical, and academic qualification to enlist individuals for the purpose of producing ready warfighters. However, the current training pipeline is the product of industrial age approaches to learning. It employs equations that include variables for time, cost, and an acceptable attrition rate. These variables represent the resource constraints on producing a trained force possessing the skills and problem-solving abilities necessary to execute the national military strategy and to achieve victory in battle. Yet, these constraints also create a static, rigid system that hinders change forcing military training and education experts to approach the goal of optimizing learning differently. In other words, we cannot just add more training or training time.

Still, almost every theory of skill acquisition emphasizes the important factor of time spent in study and practice of a skill or knowledge domain. Malcolm Gladwell popularized the theory of 10,000 hours of practice to achieve mastery in a skill domain several years ago in his best-selling book titled Outliers: The Story of Success (Gladwell, 2008). The validity of the 10,000 hour rule is widely debated (Johansson, 2012; Miller, 2018), for a variety of reasons, but the underlying principle is that the acquisition of skill and the assimilation of rote knowledge transformed into practical expertise takes time, a commodity we cannot further stretch in military training. Emphasizing the need to rethink learning expectations to include depth of processing, self-regulation, and location of learning, the battlespace is becoming increasing technologically equivalent across adversarial nations, making the necessity to accelerate cognitive capabilities. COVID-19 has also reminded us that distributed training is no longer a nice-to-have but instead a necessity. Accordingly, to ensure consistent readiness levels, the US military branches will need to define not only an effective method for training anywhere, anytime, but will also need to do so in the most efficient way possible.

In response, national strategies (National Defense Strategy, 2018; DoD Digital Modernization Strategy, 2019), as well as policy (DoDD 1322.18 Military Training, 2019; DoDI 1322.31 Common Military Training [CMT], 2020; DoDD 5000.59 DoD Modeling and Simulation [M&S] Management, 2018; DoDI 1322.26 Distributed Learning, 2017) are now requiring more efficient processes for sharing knowledge, increasing readiness and lethality, and ensuring that training and education can occur even under distributed constraints. Accordingly, several changes are taking place across the military branches. Specifically, interoperability has become a major focus area for how training apparatus connect in the future. The measurement of learning has been expanded to include measuring experiences and other related elements (e.g., using experience application programming interface [xAPI]). Data analysis is making possible equity and personalization of learning creating the possibility to alter learning trajectories based on each individual's needs. However, none of these structural changes on their own will lead to improved learning. Rather, they make it possible to access, store, and use data to inform changes to instructional practices. But what data is needed?

Further optimization can occur if we can gather data beyond what is observable by humans. Neurological measurements can help us dive deeper, understand better, and adjust training for operational readiness but also return to civilian life capabilities after theater. Thus, we consider efficiency improvement in addition to the structural changes occurring across the Department of Defense (DoD). For example, Cognitive Load Theory (CLT) attempts to measure the resource demand on the brain during learning and to optimize it through the management of mental resources during the learning process (Sweller, 1988). Through these advances, the ability to personalize learning based on a variety of data inputs is possible. Specifically, the use of neuro-physiological data, combined with performance data, trainee key factors, and content being learned data has the propensity to yield recommendations about learners and training apparatus settings that when combined can lead to enhanced learning speed and depth of understanding. Application of CLT and dynamic monitoring during the learning process may therefore have tremendous potential for the DoD to conserve its two most precious assets of time and human talent. The value proposition is to dramatically increase the efficiency or learning achieved through optimizing students' cognitive load during instruction.

Doing so would equate to significant reductions in the training wash-out rate and acceleration of skill development for personnel in technical, tactical, managerial, and leadership positions. Learning with optimized cognitive loading for students can increase readiness and capability through efficiency rather than on top-line increases to budgets for manpower and personnel. Systems that demonstrate real-time adaptation to cognitive workload have already been demonstrated in operational environments for aerial vehicle tasks and air traffic control environments (Arico et al., 2016; Wilson et al., 2007). However, cost and equipment complexity currently prohibit deployment of this technology on a large scale As such, a definition for the minimal requirements for a system that can reliability assess CL in real-time is needed. Accordingly, this paper demonstrates the minimal viable channel count necessary for an EEG-based training system. This framework can be used to guide equipment design and cost for more accurate return on investment (ROI) and efficiency calculations.

PREVIOUS RESEARCH

The quest to optimize human assets is not a new goal in the military. However, the definition, target, and capability to optimize has changed over time (Vogel-Walcutt, 2019). The definition of force readiness has broadened from initially focusing on the capability of military personnel to follow orders and physically carry out missions to include mental and physical resilience, decision making, mindfulness, neuro-cognitive clarity, stability, and efficiency. Accordingly, significant research has been sponsored by the US DoD across the past two decades focusing on understanding how the brain most efficiently and effectively takes in information, assimilates that information in long term memory, translates it, and applies it to the real world. Simultaneously, similar research dollars have funded the development and testing of various neuro-physiological apparatus with the goal of better understanding what the brain is doing during learning, how we can enhance the experience and accelerate it, and what happens when the brain is not operating in an optimized way.

Two key areas of research include resilience training (Reivich, Seligman, McBride, 2011; Seligman, 2011; Arnetz, et al., 2009) and CLT (Sweller, 1988) as well as how these can be measured in combination to promote stress reduction, learning inputs, and understanding the burden of emotional trauma on training (Coyne, et al., 2009; Proayska-Pomsta, et al., 2008). With improved understanding of the trainee comes the possibility to inform training practices and augment the training experiences in real-time with micro-adaptations only achievable through the use of technology and sophisticated neurological assessments. It allows for learning experienced to be optimized without under or over-taxing the human mind and can lead to improved decision making strategies, education, and instructional frameworks (Snyder, 1989; Eisenhardt & Zbaracki, 1992; Zsambok & Klein, 2014; Crichton & Flin, 2017; Flin, Salas, Straub, & Martin, 2017). Ultimately, the goal is to improve the speed, trajectory, capability, and therefore the readiness of each warfighter and teams of warfighters. And while the military is constantly in a state of improvement, trying to find the most optimized solutions to promote training effectiveness at the least financial cost possible can be challenging at best and reliant on self-report at worst. However, before the advancements from the science of learning can be properly implemented, we must first have access to valid and reliable neuro-cognitive data to inform the training interventions framework. These data can include performance, test results, neurological, physiological, observation, or other inputs. Particular to this paper are the affordances made by having access to neurological data from electroencephalography (EEG).

Neuro-Cognitive Measurement

Over the past two decades, the DoD has sponsored several research projects aimed at understanding the interplay between measurable cognitive activity and training interventions. In the early 2000s, the Augmented Cognition program substantively progressed the research in measuring and tracking real-time changes in individual's cognition during training (Schmorrow & Kruse, 2002). The goal was to determine how multiple neuro-physiological sensors could be combined to improve understanding and reliability of learning state, workload, and cognitive changes. A significant number of research projects were sponsored under this work and culminated into initial recommendations for combining the sensors but noted two important issues: 1) the reliability of the apparatus were still in their infancy and 2) accompanying instructional recommendations were needed (Schmorrow, Nicholson, Drexler, & Reeves, 2007; Stanney, et al., 2009). In response, the Office of Naval Research sponsored the Virtual Technologies and Environments (VIRTE) project that focused on using existing instructional interventions varied based on data from the sensors to create the Adaptive Instructional Architecture (AIA; Nicholson, Fidopiastis, Davis, Schmorrow, & Stanney, 2007). Initial findings were positive but as learning science was progressing, it became clear that new, more personalized changes to training needed to be investigated leading to the development of the Algorithms Physiologically derived to Promote Learning Efficiency (APPLE) project. The objective was to inform the selection of instructional strategies toward improved learning effectiveness and efficiency across a wide range of domains and to test the utilization of learner state data to inform the application of strategies (Vogel-Walcutt, et al., 2010). Findings led to the development of the State-based Information Processing (SIP) Model that combines a deep understanding of the optimal pathway of information through the mind, the points along that pathway where information is frequently lost, interventions that can be applied to retain knowledge, and the neuro-physiological data needed to inform the decision regarding which strategy to use and when (Vogel-Walcutt, et al., 2013).

The major issue that arose repeatedly was the level of "noise," or variability in the data. In short, while advances had been made, the reliability of the incoming data for individuals was still too low to inform changes to training. However, over the past decade, technological improvements have continued, resulting in significantly improved data to inform training alterations and improve human cognitive performance. Still plaguing the full-scale implementation of these advances, however, is the challenge of transitioning from the lab to the field. There remains a need to identify a measurement tool that is accurate, portable and deployable in real world environments. But first, we must understand how these apparatus work before the reliability and operational use can be judged.

Measuring Cognition: EEG

Electroencephalography (EEG) is a popular brain-imaging technique for both basic research on brain function as well as for building applied systems such as brain-computer interfaces (BCIs). EEG uses a set of sensors on the scalp to passively measure the extracellular current flow produced by the simultaneous activity of large groups of neurons in the cortex (Nunez and Srinivasan, 2006). This current flow produces oscillations in the EEG signal (sometimes referred to as brain waves) that can vary in frequency, amplitude, and phase depending on the type of brain activity that is generating them. Over the past 100 years, EEG has played a pivotal role in developing an understanding of the brain and it continues to be irreplaceable due to its excellent temporal precision (on the order of milliseconds), portability, and low cost relative to other brain imaging techniques such as functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), or positron emission tomography (PET) scans.

A common use for EEG in recent years is to measure neurophysiological correlates of cognition in real-time, including attention, alertness, frustration, and, the focus of this paper, cognitive load (CL). Traditional measurement of cognitive load is through subjective rating scales. The Paas scale from the early 1990s is the most popular of these and is performed after the task is completed (Paas et al., 2003). Alternatively tracking eye gaze is another method used to track split attention which contributes to increased cognitive load. Other widely used self-report measures include the NASA Task Load Index (TLX) (1986) and the Multiple Resources Questionnaire (MRQ) (Boles & Adair, 2001). Both measures assess subjective workload but differ in length and focus. While these measures do provide insight into CL, EEG provides a far more robust measure of CL in real time while the user is engaged in a learning or performance-based task. However, CL, as measured by EEG, relies on an analysis of recorded signals. Computation of power spectral densities (PSD) within classically defined frequency bands or ratios between frequency bands is required as a first step (Berka et al., 2007). CL, attention, and engagement are then derived from classifier models using these PSD bands. Several studies have shown EEG to be an effective tool for measuring CL, with a high correlation between EEG metrics and more traditional measurement tools (Berka et al., 2007; Knoll, et al., 2011; Mathewson, et al., 2012;; Wang, Gwizdka, & Chaovalitwongse, 2016). Previous studies have also shown high accuracy of artificial neural network and stepwise discriminant analysis to precisely classify cognitive workload states with an accuracy between 85-90% (Russell et al., 2006; Wilson et al., 2010). However, significant variability across time makes it difficult to designate one classifier as the optimal choice.

The primary purpose of having an accurate neuro-cognitive measurement system is to: (1) Determine which EEG patterns are associated with broad, complex and relevant measures of learning a specific set of skills or knowledge; (2) Utilize the EEG patterns to modulate the learning or performance environment to optimize factors that will augment the desired outcome; and (3) Examine whether individual differences in EEG patterns predict, alone or in combination, learning success above and beyond other measures of aptitude. Yet, the lack of portability and high cost of these systems has, to date, restricted their use in applied training environments.

Neuro-cognitive Enhancements

There are many active and passive methods to accelerate the acquisition of knowledge and skills. Neuro-cognitive enhancement includes pharmacologic, behavioral and technology-based interventions. Enhancing brain function to augment training is a careful and iterative process that is best served by evidence-based innovations in brain training technologies. However, despite very promising outcomes, EEG is still not a ubiquitous tool in education and training in the workforce or military. One reason for this is that only in the last decade have systems advanced enough in chipset, wireless, data and sensor technology to allow for reliable, relatively low-cost, portable systems. Additionally, significant advances in distributed training capabilities and interventions have also recently occurred that support the development of training content that can directly implement behavioral based changes through the introduction of worked examples, difficulty adjustment, audio and visual complexity alterations, narration and focused attention tools to adapt the lessons to the user's CL in real-time (Mayer, 2005; Walcutt & Schatz, 2019). Dynamic adjustment of the presentation of content enables a high level of personalization for learning. Using real-time neuromonitoring can also adapt to the user's mental state within seconds and allow effective management of complex, immersive training environments. Further, when accumulated data is run through machine learning classifiers the ability to design training programs that rapidly iterate to both populations and individual data can elevate the specificity of personalized training. However, advances in technology efficiency and portability are still needed to allow for deployment of EEG technology into more real-world training and operational environments.

CURRENT STUDY

Participants

Nine adult volunteers (4 male, 5 female) with ages ranging from 18 to 45 years old participated in this study. Volunteers were recruited from advertisements on local social media groups and were paid for their participation. All reported no history of cognitive impairment or neurological disorder. After giving informed consent, subjects completed the study in two sessions that took place within a week. Two subjects did not complete their final session and are each missing data from one of the three tasks.

Equipment

Participants were seated comfortably in a dim room in front of a computer monitor and small speakers that were used to present the experimental tasks. Experiments were controlled using custom scripts running within MATLAB (MathWorks, Natick, MA) using functions from the Psychophysics Toolbox (Brainard, 1997). To give behavioral responses, subjects held a response pad that connected to both the stimulus computer as well as the system recording EEG, ensuring precise timing accuracy between their responses and EEG data.

Tasks

Participants completed three different tasks, all of which are commonly used in the cognitive science literature to elicit varying levels of cognitive load (see figure 1). All tasks were practiced until the participants could perform them comfortably before the experiment began. The order of the tasks was randomized for each subject.

Task 1 was based on the Sternberg Memory Scanning Task (Sternberg, 1969). Participants were shown a series of 3, 5, or 7 unique single-digit numbers, then asked to determine if a probe number was present in the preceding list. Each number in the list remained on screen for 500ms, with a 250ms gap between numbers. Between the digit series and the probe, participants also had to complete a secondary task wherein they were asked to indicate which side of the screen contained a red square. This secondary task served to make rehearsal of the memorized list more difficult. The secondary task and probe both waited for responses from the participants. The probability of the probe being present in the preceding list was 50%. Participant received audio feedback after each trial to let them know if they answered correctly. Participants completed 5 blocks containing 40 trials per block, with trials randomly divided between low (3 digits), medium (5 digits), or high (7 digits) cognitive load.

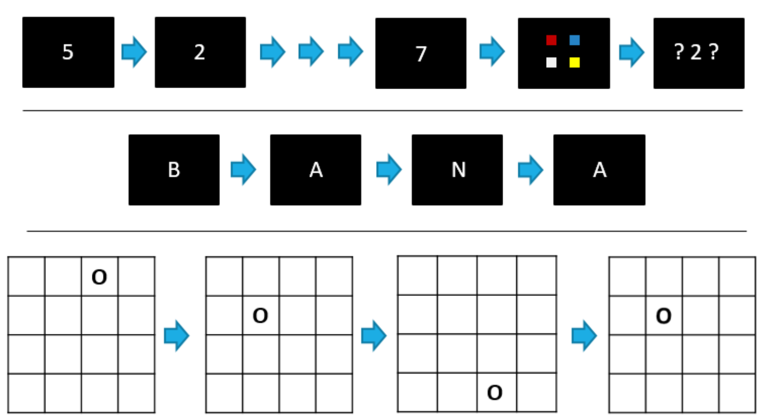

Figure 1: Working memory tasks. Top: Task 1, subjects view a series of 3 to 7 digits, then after an interference task respond if a prompted digit was in the preceding series. Middle: Task 2, *n*-back task. Subjects press a button when a letter appears that was seen *n* steps ago. Bottom: Spatial *n*-back task using positions on a grid instead of letters. The last frame of both *n*-back tasks shows targets for an *n* of 2.

Tasks 2 and 3 were both variations of the *n*-back task used in previous EEG studies of cognitive load (Baldwin and Penaranda, 2012), one using letters and another using spatial locations on a 4 by 4 grid. In an *n*-back task, participants are shown a series of stimuli and are asked to respond by pressing a button whenever a stimulus matches one from *n* steps back. When *n* is 1, participants only need to remember the most recent stimulus and respond if the next stimulus matches. As *n* grows larger, participants need to maintain a larger list of recent stimuli in memory. In practice, an *n* of 1 is trivially easy to perform, an *n* of 2 requires effort, and an *n* of three is very challenging. Both the letter and spatial versions of the *n*-back task had identical timing and target probabilities. Each stimulus appeared on the screen for 1.5 seconds, with a 250ms to 500ms variable inter-stimulus interval. Any given stimulus had a 25% chance to be a target *(i.e.* match the letter or grid position from *n* steps ago) and thus required a response from the subject. The experiment did not wait for participants to respond, thus the response button had to be pressed before the next stimulus appeared on the screen. No feedback was given during blocks as it could interfere with maintaining items in memory. Trials were grouped into blocks of 80 stimuli, with the *n* (and thus difficulty) constant across the block. For each version of the *n*-back task, participants completed 6 blocks each of the easy, medium, and hard conditions, with the ordering of the blocks randomized.

EEG collection and pre-processing

During all tasks, EEG was recorded using a 64-channel HydroCell Geodesic Sensor Net made by Electrical Geodesics Inc. (now Phillips Neuro, Eugene, OR). EEG was collected at 1000 samples per second with a vertex reference. Using EGI's built-in tools, the EEG data was band-pass filtered using zero phase-shift Butterworth filters from 1 to 100 Hz, notch filtered at 60 Hz to remove electrical line noise, then exported to MATLAB for all further analyses. Two seconds (2000 samples) of EEG data were extracted for analysis from each trial. For the digit series task, this was the 2 seconds immediately preceding the appearance of the secondary task, during which the participant was attempting to maintain the preceding series in their memory. For the n-back tasks, the 2 seconds of data from each trial aligned with the onset of each stimulus. Given a minimum of 2 seconds between stimuli ensured no data appeared in more than one trial. Trial data were visually screened by an experimenter to ensure that no large, uncorrectable artifacts were present in the data, such as large subject movements or channels with poor connections to the scalp. In all subjects, this screening resulted in the exclusion of less than 5% of the total data. Independent Component Analysis was then used to identify and remove blinks and eye movements from the data (Jung et al, 2000).

EEG Power Comparison

The spectral power of the EEG for each trial was calculated using the discrete Fourier transform in MATLAB. Power was summed across each of the characteristic frequency bands of EEG (delta, theta, alpha, beta, and gamma) to produce 5 power values per EEG channel, per trial. These power values were then binned by subject, task, and by condition (LowCL, MedCL, HighCL) to create an average for each. Differences in the power values between the HighCL and LowCL conditions were used to identify the scalp locations and frequency bands that varied with working memory load on each task.

Within-Task Classification

In addition to identifying the neural correlates of cognitive load in three tasks, we wanted to determine how reliable those neural correlates could be in determining the cognitive load of individual trials. To do this, a linear model for each subject and task was built that could predict the condition (*i.e.* high or low CL) of each individual trial in a leave one out cross-validation (LOOCV) scheme. To conduct LOOCV, all but one trial is assigned into a testing set to train a linear discriminant classifier, then that model is used to predict the condition of the remaining trial that was not used in training the classifier and record if the classifier guessed correctly. This process is then repeated for every single trial, and the classifier performance is judged by dividing the number of trials in which the classifier correctly identified the condition of the trial by the total number of trials. To perform these classification tests, we used the linear discriminant classifier built into MATLAB (classify.m).

Additionally, we wanted to estimate how reducing the number of EEG channels recorded would impact classification performance. Neighboring EEG channels have largely correlated signals due to volume conduction effects and smearing by the skull (Nunez and Srinivasan, 2006). Thus, information gained by adding more channels typically shows diminishing returns. To estimate the ability to measure cognitive load changes using EEG systems with fewer channels, the within-task classification tests were repeated using subsets of 32, 16, or 8 of the 64 EEG channels that were recorded. For each subset of channels, an even coverage of the scalp was maintained while reducing the density of the channels, similar to the differences between the International 10/20, 10/10, and 10/5 electrode spacing standards (Oostenveld and Praamstra, 2001).

Cross-Task Classification

Previous studies have shown conflicting results regarding the extent to which EEG measurements of cognitive load can be generalized across different tasks, with some studies using a generalized metric across a range of tasks (Berka et al., 2007) while others found poor performance for cross-task classification (Baldwin and Penaranda, 2012). We tested this by training linear classifiers for each subject and each task, then using those classifiers to predict the condition (HighCL vs LowCL) of individual trials from the other two tasks. For example, for a given subject all trials from the Digit Series task were used to train a linear discriminant classifier, then that classifier was used to predict the condition of each trial in the *n*-back tasks. As above, these tests were repeated using subsets of the EEG channels to determine how reducing the number of channels impacted classification performance.

RESULTS

Behavior

Participants performed well on all three tasks, with accuracy falling as the tasks became more demanding. Table 1 contains a full list of behavioral performance for all subjects separated by task and cognitive load.

Behavioral Performance (Accuracy %)									
Digit Series									
	S1	S2	S3	S4	S5	S6	S7	S8	S9
LowCL	98.3	100	98.0	100	100	100	95.2	100	100
MedCL	98.7	98.5	98.0	94.6	94.0	98.5	94.9	96.0	94.0
HighCL	93.2	69.4	81.7	78.7	92.5	91.7	86.2	88.2	92.5
N-back Letters									
	S1	S2	S3	S4	S5	S6	S7	S8	S9
LowCL	98.8	99.7	100	96.9	99.8	99.4	100	n/a	98.4
MedCL	97.5	94.4	97.8	87.5	97.9	97.2	98.4	n/a	90.3
HighCL	85.3	86.3	86.6	81.3	88.8	94.1	90.6	n/a	85.9
N-back Grid									
	S1	S2	S3	S4	S5	S6	S7	S8	S9
LowCL	97.8	99.4	94.1	85.0	100	99.7	99.7	100	n/a
MedCL	95.9	90.6	98.8	85.6	98.3	96.9	96.9	98.1	n/a
HighCL	85.6	77.5	90.0	84.1	93.0	96.3	87.2	89.4	n/a

Table 1: Behavioral Task Performance. Accuracy is reported for each subject as a percentage of all trials answered correctly, separated by the task and by the levels of induced cognitive load.

EEG Power: Task Variability

When comparing the differences in EEG power between the high and low cognitive load conditions averaged over all subjects, noticeable changes only in the theta and alpha frequency bands were observed suggesting that increased workload is primarily affecting the low-frequency cognitive and attention networks that operate in those frequency ranges, as opposed to increasing gain across the board which would be visible in all frequency bands. Furthermore, those theta and alpha power effects were substantially different between the three tasks (see figure 2). This means that an interaction is occurring between the task-specific demands and changing cognitive load, rather than a simple main effect of workload. For the digit series task, increases in cognitive load showed effects commonly noted in previous studies: elevated frontal theta power paired with a reduction in posterior alpha power which is typically interpreted as frontal memory networks increasing their activity while posterior alpha networks are focused on the task. Next, the spatial version of the *n*-back task was the most similar to the digit series pattern in scalp distribution, but the relative size of the theta and alpha effects were quite different. Higher cognitive load in the spatial task produced much more intense frontal theta activity, but a less intense difference in the posterior alpha activity. Finally, despite being structurally similar to the spatial *n*-back task and visually similar to the digit series task, the letter *n*-back task showed an inverted effect of cognitive load on theta band power when compared to the other tasks, with increases in cognitive load corresponding to a decrease in frontal theta. Interestingly, the reduction in posterior alpha remains similar to the other tasks, so not all the power effects are inverted demonstrating that there is a complex interaction between the specific task demands and endogenous memory and attention networks.

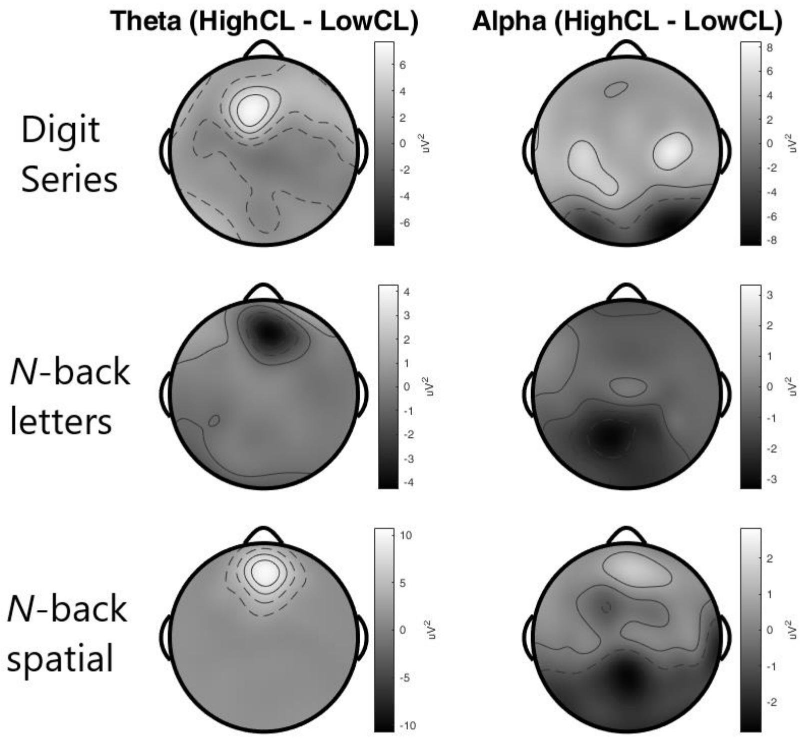

Figure 2: EEG Power - Task Variability. Top-down topographic plots of the head showing areas where EEG power is sensitive to cognitive load. Measurements are averaged over all subjects. Plots show differences in power between high and low cognitive load conditions: bright areas have elevated power with high cognitive load, while dark areas show reduced power with high cognitive load.

EEG Power: Subject Variability

When examining the EEG power differences within individual participants, very few closely followed the patterns present in the subject averages. This means that cognitive load metrics derived from group means would perform poorly for most subjects, as very few individuals exhibit that exact pattern. Figure 3 illustrates the differences between individual subject maps and the average across all subjects for the Digit Series task. Whereas the subject average displays an increase in frontal theta power that appears focused over central parts of the frontal cortex, within individual subjects the location of that theta peak can be maximal anywhere from occipital areas, to prefrontal channels, or even over left temporal regions. The alpha band effects are similarly varied, with only half of the subjects exhibiting the posterior reduction in alpha that dominates the subject average. Moreover, three of the nine subjects showed no decrease in alpha power anywhere on the scalp, which is usually the most consistent effect across EEG studies of cognitive load. Taken together, these findings suggest that subjects are responding to increasing cognitive load demands in a variety of functionally distinct ways that are reflected in different patterns of changes in their low-frequency cognitive networks.

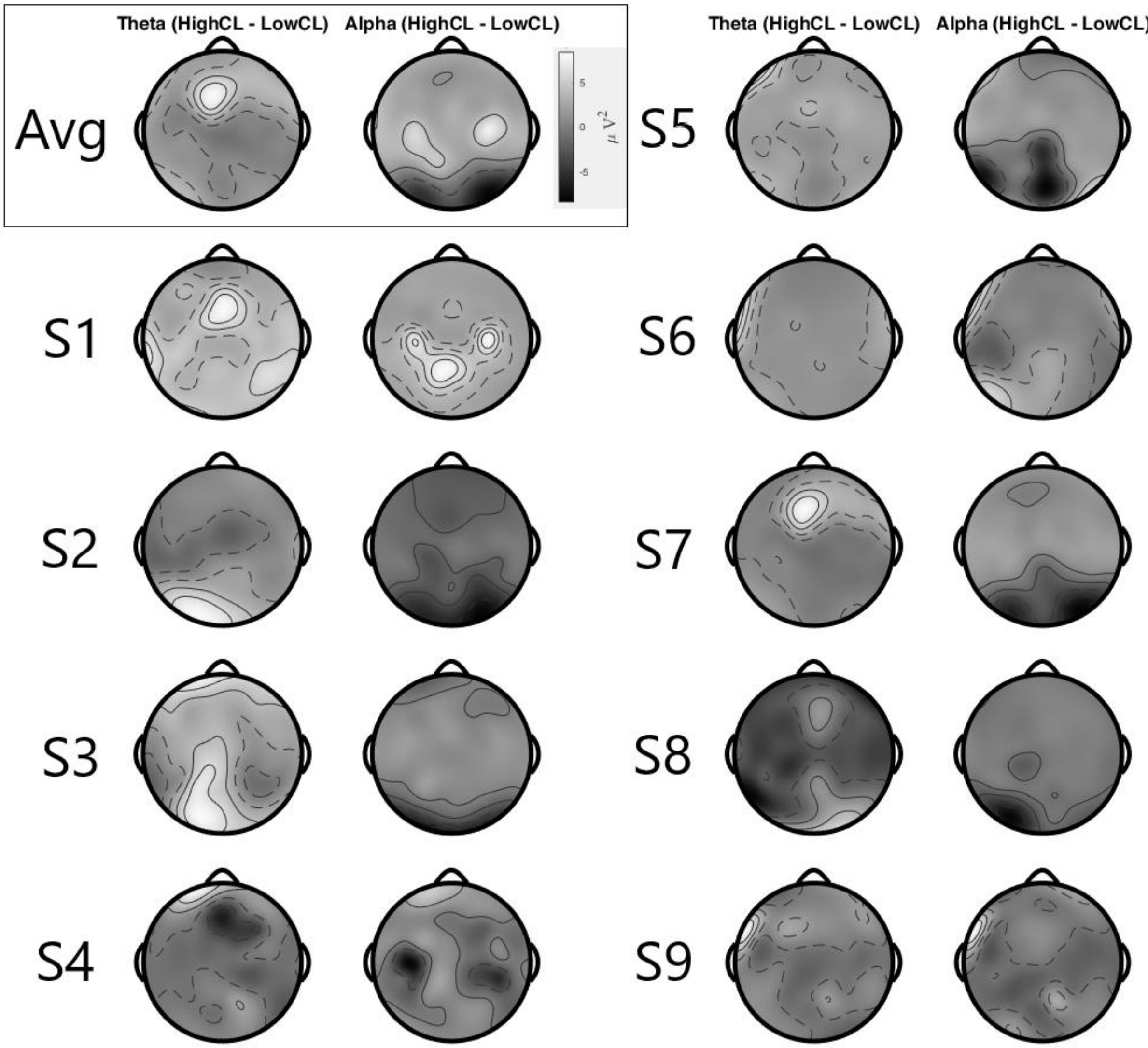

Figure 3: EEG Power – Subject Variability. Top-down topographic plots of the head showing areas where EEG power is sensitive to cognitive load in the Digit Series task. The upper left shows the average map across all subjects, while the rest show the individual subjects. Bright areas have elevated power with high cognitive load, while dark areas show reduced power with high cognitive load.

Within-Task Classification

This analysis demonstrated that linear classifiers trained on EEG power measures were able to discriminate between high and low cognitive load trials well above chance (50%) for the majority of subjects (see Table 2). Overall accuracy was moderate, given the small amount of EEG data used for each judgement (2 seconds). Classifier performance varied across tasks, with the digit series classifiers performing the worst on average and the spatial *n*-back classifiers performing the best, demonstrating that the better-performing task seems to elicit more consistent (and thus more classifiable) changes in EEG power. The success of the spatial *n*-back classifier is particularly noteworthy in that it was able to accurately make predictions even in the subjects for whom the other two task's classifiers performed poorly, with improvements of 15-25% classification accuracy over those other tasks. This finding suggests that increased spatial memory demands elicit particularly robust changes in the EEG. Within-task classification performance was surprisingly resilient when the number of EEG channels used was reduced. The vast majority of classifiers suffered little to no decrease in performance when the number of EEG channels being used was dropped down from 64 to 8 suggesting that accurate readings can be obtained with more portable and less costly 8-channel systems.

Within-Task Classifier Accuracy									
Digit Series									
# Chans	S1	S2	S3	S4	S5	S6	S7	S8	S9
64	70.18	67.24	66.67	52.17	67.72	49.60	67.83	70.33	55.04
32	70.18	68.10	69.97	51.09	66.14	48.24	66.96	70.73	55.14
16	66.67	68.10	69.79	50.00	65.35	48.82	60.00	69.92	57.75
8	66.67	68.97	67.71	58.70	66.93	48.03	59.13	69.10	55.55
N-back Letters									
# Chans	S1	S2	S3	S4	S5	S6	S7	S8	S9
64	67.91	65.72	77.10	59.97	70.66	58.69	71.50	n/a	60.07
32	67.14	66.72	76.61	58.18	70.77	58.36	69.73	n/a	61.41
16	68.17	64.23	76.44	59.81	71.43	59.18	69.40	n/a	60.40
8	66.88	64.89	76.11	56.40	70.55	58.52	68.12	n/a	59.73
N-back Grid									
# Chans	S1	S2	S3	S4	S5	S6	S7	S8	S9
64	79.38	71.06	72.27	72.32	84.13	75.70	69.51	88.16	n/a
32	80.52	71.06	73.90	72.49	78.97	76.36	69.84	88.16	n/a
16	81.17	72.20	74.71	72.65	83.61	74.88	71.80	86.56	n/a
8	80.36	68.62	71.45	71.00	80.77	77.85	67.38	88.32	n/a

Table 2: Within-Task Classification. Classifier accuracy is reported using leave one out cross-validation and represents the likelihood of correctly identifying the cognitive load of any given trial using just that 2 seconds of EEG data. Classifier accuracy is reported separated by task, subject, and as a function of the number of EEG channels retained from the original recording.

Cross-Task Classification

Classifier performance never rose above chance for any subject when attempting to use a classifier trained on one task to predict the data from another task suggesting again that the EEG effects interact with task demands, and thus models for cognitive load need to be fit to both an individual user and the specific task.

DISCUSSION

Data Review

In the present study, we attempted to clarify some lingering questions regarding how to build a reliable and generalizable EEG instrument to measure cognitive load so that better neuro-adaptive learning platforms can be implemented in the near future. Results suggest that varying participants' cognitive load in different tasks produced remarkably different changes in the resultant EEG measures, even in tasks that are structurally very similar. Furthermore, individual differences between subjects were substantial, with the peak location and sometimes even the direction of EEG effects varying widely between individuals. Nevertheless, when using a subject's own data fit to a specific task, it was possible to determine the participant's current CL with very small segments of data, and this ability was preserved even when drastically minimizing the number of EEG channels used in the assessment.

The inter-task variability observed helps to elucidate why most previous studies have been unsuccessful in measuring CL outside of the task that was used to construct the EEG metric (*e.g.* Baldwin and Penaranda, 2012). One notable exception is the Berka et al. (2007) study in which a single workload EEG metric was developed for a range of 5 tasks and appeared to correlate moderately well with traditional subjective measures of cognitive load when tested on new subjects. One possible reason why that study was successful could be due to the large number of subjects used to develop the metric – perhaps the classifier picked up on subtle effects in the EEG that are consistent across subjects but less obvious than the theta and alpha band effects that are usually reported. Alternatively, their classifier puts a majority of its weight on high-frequency activity that is more commonly associated with muscle activity, so it is possible that the metric was measuring muscle tension or frustration that just happened to be correlated with CL.

The present study also reinforces the need pointed out by Wilson et al. (2010) to have metrics that are specifically tailored to an individual, and likely also adjusted due to day-to-day variability in EEG data. While some of the task means appeared to match the patterns reported in earlier studies, any individual participant's data is likely to be maximal in a different scalp location, or even differ entirely in the direction of the effect. A potentially fruitful line of future work could be to develop a more comprehensive model for measuring CL which includes factors for individual differences in working memory capacity, current physiological state (such as alertness), task demands, and more.

Finally, the resilience of the within-task classifiers to a large reduction in the number of available EEG channels is encouraging for potential military and applied settings. Greater numbers of channels would require additional space, weight, power, transmission bandwidth, cost, and would compete for those resources with other potentially-useful monitoring tools such as accelerometers, heart rate and respiration monitors, and other biological sensors that could be critical for monitoring operator efficiency. Thus, being able to make actionable measurements with a minimal number of channels increases the likelihood of being able to quickly translate lab-based research tools to trainee-augmented field systems.

Modernizing Military Learning

Modernizing the way training and education, or learning, is developed and personalized for servicemembers is no small endeavor. Annually, demands for "readiness" are made and with it come expectations for changes in how we measure, teach, drill, and engage our military personnel. Yet one surprising fact is that data collected rarely represents trainees beyond pass fail es. As a result, the ability to determine if an individual, team, or larger unit is "ready" is functioning barely beyond an educated guess. As technology has advanced, however, data science can now help inform everything from training to team formation to predicting operational success. These capabilities have driven senior leadership to embrace the need to consider technology assessment techniques for improving measurement precision. Two decades of research suggest EEG could significantly enhance conclusions drawn about trainees and aid decision makers in assessing readiness. The current study demonstrates that CL data can be collected efficiently and reliably and apparatus are ready for hardening to be used at scale.

Application

In entry level training for almost all military occupational specialties the students are groups with other students in classes that proceed together through a course of instruction. They begin together and graduate together as a class. The program of instruction is designed to challenge but not overwhelm the recruits with the lowest learning aptitude as measured by the Armed Forces Vocational Aptitude Battery (ASVAB). Class based instruction has been the standard approach in public education as well because of the methods of instruction that were traditionally governed by the teacher to student ratio and the limitation on one to one instruction. The maturity and proliferation of network delivered instruction and exploratory learning environments has removed the limitation of a human-to-humans ratio for instructor and students. Self-paced learning becomes a more constant and realistic opportunity for imparting basic knowledge and exploring new concepts through research and experimentation.

The design of the military's training pipelines is derived from a trade-space analysis of time, cost, and training attrition. The analysis attempts to assign a value to the development, maintenance, and replacement of a quantity of soldiers with various specialties that constitute the nation's "force in readiness". The direct cost of training and the time allowed to complete training to a specific standard are fairly straight forward budgetary matters. Skills and wisdom that are expensive and take a long time to develop usually incur a longer obligation of service in the all-volunteer force. Application of dynamic cognitive load optimization may provide a means for avoiding a major portion of training attrition that is the result of boredom at one end of the spectrum and cognitive overload at the other end of the spectrum. For example, military aviators must commit to at least six years or longer of active duty service because the aviation training pipeline is much longer that other military occupational specialties.

Dynamic measurement of cognitive load combined with a mechanism to adjust the presentation of learning material provides the opportunity to move beyond educational programs based on class cohort model for instruction and learning. Specific ranges of cognitive load measured by electro encephalograms (EEG) during the learning process has been shown to strongly correlate with the assimilation of knowledge into practical application of procedural skills and problem-solving ability. Instruction that insufficiently stimulates cognitive load results in boredom and lack of challenge the impedes the assimilation of knowledge into skill. At the other end of the spectrum, presentation of learning material that overstimulates the cognitive load of the student may overwhelm an individual's learning capacity.

Individual differences between individuals also means that the same presentation of learning material will result in significantly different levels of cognitive load for each individual (Pass, Kalyuga, Leutner, 2010). That implies that the presentation of learning material and instruction can be optimized at the individual level rather than an approximation median learning ability of a class cohort. Individual optimization of learning presents a problem for presenting instruction to the traditional class cohort in a collective manner. Learning experiences tailored to individuals to allow progression through a program of instruction under optimized cognitive load is now a possibility.

Implementing cognitive load management during the learning process will make it possible to preserve human capital and the investment of training and education for individuals that would otherwise would have been considered learning attrition for not completing training with their class cohort. These data, combined with previous findings may lower the cost and size of EEG sensors while maintaining a level of precision measurement that makes it possible to bring cognitive load management out of the laboratory and into the Department of Defense's learning enterprise.

REFERENCES

Arico, P., Borghini, G., Flumeri, G., Colosimo, A., Bonelli, S. Golfetti, A. Pozzi, S., Imbert, J., Granger, G., Benhacene, R., Babiloni, F. 2016. Adaptive Automation Triggered by EEG-Based Mental Workload Index: A Passive Brain -Computer Interface Application in Realistic Air Traffic Control Environment. Frontiers in Human Neuroscience. 10: Article 539.

Arnetz, B. B., Nevedal, D. C., Lumley, M. A., Backman, L., & Lublin, A. (2009). Trauma resilience training for police: Psychophysiological and performance effects. *Journal of Police and Criminal Psychology*, *24*(1), 1-9.

Baldwin, C.L, Pnaranda, B.N. 2012. Adaptive training using an artificial neural network and EEG metrics for withing and cross task workload classification. NeuroImage 59: 48-56.

Berka, C., Levendowski, D., Lumicao, M., Yau, A., Davis, G., Zivkovic, V., Olmstead, R., Tremoulet, P., Craven, P. 2007. *EEG Correlates of Task Engagement and Mental Workload in Vigilance, Learning, and Memory Tasks.* Aviat Space Environ Med 78: B231-44.

Boles, D. B., & Adair, L. P. (2001, October). The multiple resources questionnaire (MRQ). In Proceedings of the human factors and ergonomics society annual meeting (Vol. 45, No. 25, pp. 1790-1794). Sage CA: Los Angeles, CA: SAGE Publications.

Brainard, D. H. (1997). The Psychophysics Toolbox. Spat Vis 10: 433–436.

Coyne, J. T., Baldwin, C., Cole, A., Sibley, C., & Roberts, D. M. (2009, July). Applying real time physiological measures of cognitive load to improve training. In International Conference on Foundations of Augmented Cognition (pp. 469-478). Springer, Berlin, Heidelberg.

Crichton, M., & Flin, R. (2017). Command decision making. In *Incident command: Tales from the hot seat* (pp. 201-238). Routledge.

DoD Digital Modernization Strategy. (2019). https://media.defense.gov/2019/Jul/12/2002156622/-1/-1/1/DOD-DIGITAL-MODERNIZATION-STRATEGY-2019.PDF

DoDD 1322.18 Military Training. (2019). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/132218p.pdf

DoDD 5000.59 DoD Modeling and Simulation (M&S) Management. (2018). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/500059p.pdf

DoDI 1322.26 Distributed Learning. (2017). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/132226_dodi_2017.pdf?ver=2017-10-05-073235-400/home_new

DoDI 1322.31Common Military Training (CMT). (2020). https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodi/132231p.pdf?ver=2020-02-20-114040-493

Eisenhardt, K. M., & Zbaracki, M. J. (1992). Strategic decision making. *Strategic management journal*, *13*(S2), 17-37.

Flin, R., Salas, E., Straub, M., & Martin, L. (2017). *Decision-making under stress: Emerging themes and applications*. Routledge.

Gladwell, M. (2008). *Outliers: The Story of Success.* Little Brown and Company.

Hart, S. G. (1986). NASA Task load Index (TLX). Volume 1.0; Paper and pencil package.

Johansson, F. (2012). *The Click Moment: Seizing Opportunity in an Unpredictable World*. Portfolio

Knoll, A., Wang, Y., Chen, F., Xu, J., Ruiz, N., Epps, J., & Zarjam, P. (2011). Measuring cognitive workload with low-cost electroencephalograph. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6949 LNCS*, 568–571.

Mathewson, K. E., Basak, C., Maclin, E. L., Low, K. A., Boot, W. R., Kramer, A. F., … Gratton, G. (2012). Different slopes for different folks: Alpha and delta EEG power predict subsequent video game learning rate and improvements in cognitive control tasks. *Psychophysiology*, *49*(12), 1558–1570.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. The Cambridge handbook of multimedia learning, 41, 31-48.

Miller, M. (2018). The Great Practice Myth: Debunking the 10,000 Hour Rule. Retrieved from: https://www.6seconds.org/2018/02/09/the-great-practice-myth-debunking-the-10000-hour-rule-and-what-you-actually-need-to-know-about-practice/

National Defense Strategy Summary. (2018). https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf

Nicholson D.M., Fidopiastis C.M., Davis L.D., Schmorrow D.D., Stanney K.M. (2007) An Adaptive Instructional Architecture for Training and Education. In: Schmorrow D.D., Reeves L.M. (eds) Foundations of Augmented Cognition. FAC 2007. Lecture Notes in Computer Science, vol 4565. Springer, Berlin, Heidelberg

Nunez PL, Srinivasan R. (2006). Electric Fields of the Brain: The Neurophysics of EEG. New York: Oxford Univ. Press.

Paas, F., Tuovinen, J. E., Tabbers, H., Van Gerven, P. W. M., & Gerven, P. W. M. Van. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, *38*(1), 43–52.

Pass, J., Kalyuga, S., & Leutner, D. (2010). Individual Differences and Cognitive Load Theory. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive Load Theory* (pp. 65-88). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511844744.006 Retrieved from https://doi.org/10.1017/CBO9780511844744.006

Porayska-Pomsta, Kaska, Mavrikis, Manolis, and Pain, Helen. (2008). Diagnosing and acting on student affect: the tutor's perspective. *User Model User-Adap Inter, 18,* 125-173.

Reivich, K. J., Seligman, M. E., & McBride, S. (2011). Master resilience training in the US Army. *American Psychologist*, *66*(1), 25.

Russell, CA., Wilson GF., Riski MM., 2006. Comparing classifiers for real time estimation of cognitive workload. Human Factors and Ergonomics Society 49th Annual Meeting, Orlando, FL.

Schmorrow, D. D., Nicholson, D. M., Drexler, J. M., & Reeves, L. M. (Eds.). (2007). Foundations of augmented cognition (4th Ed.). Arlington, VA: Strategic Analysis, Inc.

Schmorrow, D.D, & Kruse, A.A. (2002). "DARPA's Augmented Cognition Program-tomorrow's human computer interaction from vision to reality: building cognitively aware computational systems," *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*, Scottsdale, AZ.

Seligman, M. E. (2011). Building resilience. *Harvard business review*, *89*(4), 100-6.

Snyder, J. (1989). *The ideology of the offensive: Military decision making and the disasters of 1914* (Vol. 2). Cornell University Press.

Sternberg, S., 1969. The discovery of processing stages: extensions of Donders' method. Acta Psychol. 30, 276–315.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12, 257-285.

Vogel-Walcutt, J., Fiorella, L., Malone, N., 2013. Instructional strategies framework for military training systems. *Computers in Human Behavior* 29: 1490-1496.

Vogel-Walcutt, J., Gebrim, J., Bowers, C., Carper, T., Nicholson, D. 2011. Cognitive load theory vs. constructivist approaches: which best leads to efficient, deep learning? *Journal of Computer Assisted Learning* 27: 133-145.

Vogel-Walcutt, J.J. (2019). Cognitive Weaponry: Optimizing the Mind. Paper presented at the I/ITSEC Conference, December. Orlando, FL.

Vogel-Walcutt, J.J., Bowers, C.A., Marino-Carper, T., & Nicholson, D. (2010). Increasing Learning efficiency in military learning: Combining efficiency and deep learning theories. *Military Psychology*, 22(3).

Walcutt, J.J. & Schatz, S. (2019). Modernizing Learning: Developing the Future Learning Ecosystem. Government Publishing Office. Washington, D.C.

Wang, S., Gwizdka, J., & Chaovalitwongse, W. A. (2016). Using Wireless EEG Signals to Assess Memory Workload in the n-Back Task. *IEEE Transactions on Human-Machine Systems*, *46*(3), 424–435.

Wilson, G., Russell, C., 2007. Perfromance Enhancement in an Uninhabited Air Vehicle Task Usint Psychophysiologically Determined Adaptive Aiding. Human Factors: The Journal of the Human Factors and Ergonomics Society 49: 1005.

Wilson, G.F., Russell, C.A., Monnin, J.W., Estepp, J.R., Christensen, J.C. 2010. How does Day to Day Variability in Psychophysiological Data Affect Classifier Accuracy? Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting.

Zsambok, C. E., & Klein, G. (Eds.). (2014). *Naturalistic decision making*. Psychology Press.