

USAARL Report No. 2019-12

Real-time Operator State Monitoring during Periods of High and Low Workload across Four Workload Domains

By Kathryn Feltman¹, Amanda M.
Kelley¹, Kyle Bernhardt^{1,2}, Thomas
Britt^{1,3}, Colby Mathews^{1,2}

¹U.S. Army Aeromedical Research Laboratory
²Oak Ridge Institute for Science and Education
³Clemson University



United States Army Aeromedical Research Laboratory

Warfighter Performance Group

August 2019

Approved for public release; Distribution unlimited.

Notice

Qualified Requesters

Qualified requesters may obtain copies from the Defense Technical Information Center (DTIC), Fort Belvoir, Virginia 22060. Orders will be expedited if placed through the librarian or other person designated to request documents from DTIC.

Change of Address

Organizations receiving reports from the U.S. Army Aeromedical Research Laboratory on automatic mailing lists should confirm correct address when corresponding about laboratory reports.

Disposition

Destroy this document when it is no longer needed. Do not return it to the originator.

Disclaimer

The views, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation. Citation of trade names in this report does not constitute an official Department of the Army endorsement or approval of the use of such commercial items.

REPORT DOCUMENTATION PAGE

*Form Approved
OMB No. 0704-0188*

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 12-08-2019	2. REPORT TYPE Final	3. DATES COVERED (From - To) Nov 16 to Mar 18
--	--------------------------------	---

4. TITLE AND SUBTITLE A Systematic Review of Cognitive Enhancement Interventions for Use in Military Operations	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S) Kelley, Amanda; Feltman, Kathryn A.; Nwala, Emmanuel; Bernhardt, Kyle; Hayes, Amanda; Basso, Jared; Mathews, Colby	5d. PROJECT NUMBER 22900
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Aeromedical Research Laboratory P.O. Box 620577 Fort Rucker, AL 36362	8. PERFORMING ORGANIZATION REPORT NUMBER USAARL 2019-12
--	---

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

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

13. SUPPLEMENTARY NOTES
Oak Ridge Institute for Science and Education; Clemson University

14. ABSTRACT
Future battlefield environments will require Soldiers to process large amounts of mission critical information within short periods and to interact with adaptive autonomous systems. Enhancing cognitive abilities such as sustaining attention over longer periods, more efficient processing of perceptual information, and quicker decision-making processes in these environments will promote mission success and operational effectiveness. The literature suggests numerous interventions for cognitive performance enhancement, including the use of stimulant agents, transcranial direct current stimulation, as well as others, some of which may be more appropriate for military settings. The purpose of this review was to summarize findings regarding the enhancement interventions from randomized, placebo-controlled, experimental research studies. While evidence supports the efficacy of some of the identified methods, further evaluations of not only the effectiveness and safety of these interventions but also the reliability and duration of effects are needed prior to implementation in military settings.

15. SUBJECT TERMS
cognition, attention, memory, enhancement

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 54	19a. NAME OF RESPONSIBLE PERSON Loraine St. Onge, PhD
a. REPORT UNCLAS	b. ABSTRACT UNCLAS	c. THIS PAGE UNCLAS			19b. TELEPHONE NUMBER (Include area code) 334-255-6906

This page is intentionally blank.

Acknowledgements

The authors would like to express their gratitude to the research team at the U.S. Army Aeromedical Research Laboratory for their contributions to this project. Special gratitude is extended to Mr. Walter Waltz who developed the electrooculogram algorithm for data processing.

This research was supported in part by an appointment to the Research Program at the U.S. Army Aeromedical Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the U.S. Army Medical Research and Materiel Command.

This page is intentionally blank.

Table of Contents

	Page
Introduction	1
Quantifying Workload	3
Study 1	6
Method	6
Participants.....	6
Materials.	6
Questionnaires.....	7
Psychophysiological recording.....	8
Cognitive Tasks	9
Auditory task.....	9
Visual task.....	10
Physical task.....	11
Cognitive task	12
Procedure	12
Statistical analysis and quality control.....	13
Results.....	13
Assessment of the Workload Manipulations: Performance.....	13
Assessment of the Workload Manipulations: NASA-TLX Ratings.....	14
Effect of Workload on Psychophysiological Recordings	15
Correlations between performance, psychophysiological recordings, and perceived workload	16
Phase 1 Discussion.....	16
Study 2	16
Method	16
Participants.....	16
Materials	17
Questionnaires.....	17
Workload Profile.....	17
Psychomotor Vigilance Task	17
Psychophysiological recording	17
Black Hawk Flight Simulator	17
Flight scenarios	17
Procedure	20
Statistical analysis and quality control.....	20
Results.....	20
Assessment of the Workload Manipulations: Performance Data	20
Visual Domain	21
Cognitive Domain.....	21
Auditory Domain	22
Physical Domain.....	22
Assessment of the Workload Manipulations: NASA-TLX Ratings.....	22
Assessment of the Workload Manipulations: Workload Profile Ratings	23
Study 2 Discussion.....	30
General Discussion	30
Conclusions	32

Table of Contents (continued)

	Page
References	33
Appendix	40

List of Tables

1. Descriptive summary of questionnaires and survey instruments.....	7
2. Performance on the Different Tasks as a Function of Workload.....	14
3. NASA-TLX Ratings on the Different Tasks as a Function of Workload.....	15
4. Workload Manipulation Descriptions of Flight Scenarios	18
5. Descriptions of Variables of Interest for Flight Scenarios.....	19
6. Performance Descriptive Statistics Visual Domain.....	21
7. Performance Descriptive Statistics Cognitive Domain.....	21
8. Performance Descriptive Statistics Auditory Domain.....	22
9. Performance Descriptive Statistics Physical Domain.....	22
10. NASA-TLX Results.....	23
11. NASA-TLX Descriptive Statistics.....	23
12. Workload Profile Results.....	23
13. Workload Profile Descriptive Statistics.....	24
14. Psychophysiological Descriptive Statistics Visual Domain	24
15. Psychophysiological Descriptive Statistics Cognitive Domain.....	25
16. Psychophysiological Descriptive Statistics Auditory Domain	26
17. Psychophysiological Descriptive Statistics Physical Domain	27

List of Figures

1. Example of a trial in the visual task: A) Low workload, and B) High workload.	11
2. Example of a trial in the MAT-B task.	12

Introduction

Current doctrine outlines the Army's goal of developing a more mobile, integrated Warfighter that is capable of quickly adapting to an ever-changing operational environment (Army Vision, 2014). This vision entails developing increased capabilities with reduced personnel and equipment. Two tenants of this doctrine include optimization of human performance and interoperability within Army units along with other services. As the military begins to look towards the challenges of the next 30 to 40 years, it is apparent that an increased reliance on technology will play a pivotal role in meeting this goal. As the nature of missions change from conflicts between nation states to isolated, independent non-state actors shifting from large-scale ground campaigns to smaller, tactical units, the Warfighter of the future will need to be capable of handling multiple, complex tasks on an independent basis.

In the civilian aviation sector, human-computer automation capabilities are ubiquitous. Nearly every function in the planning and execution of air travel can be accomplished through sophisticated computer systems. Aviation in the U.S. Army, which predominantly uses rotary-wing aircraft, has not yet fully incorporated the same level of automation, partially due to the unique nature of military operations. However, the Army is moving toward developing a leaner, more capable and adaptive force that increasingly requires multifunctional capabilities to accomplish more goals with fewer personnel or resources. For aviators, this will entail new and additional tasks to meet operational objectives that will subsequently increase cognitive demands. The current fleet of Army aircraft has been in service for over 30 years. Advances in engineering and technology have led to the development of a new generation of rotary-wing aircraft under a program called Future Vertical Lift (FVL) (Tadjeh, 2015). In addition to the increased capabilities from these innovative designs, the incorporation of human-computer interface technologies is a near certainty. Automation has not been widely used in the Army's aircraft, thus its implementation will require a substantial amount of research to ensure these airframes meet military standards for use in an operational environment, and meet usability expectations for this population.

Today's Army rotary-wing pilot experiences high levels of cognitive demand. Throughout the duration of a flight, a pilot must maintain orientation of the aircraft's location in space, survey the functioning of various systems within the aircraft, analyze weather conditions, communicate with others on board and on the ground, engage in airspace surveillance for other aircraft and terrain, and keep track of time-to-mission objective. The successful completion of a mission, ranging from performing reconnaissance to providing medical evacuation, requires the pilot to maintain constant awareness of these aspects throughout the flight. The variety of information the pilot keeps within memory and awareness throughout the flight is used to ensure adequate situation awareness (SA) (Wickens, 2002). The most widely accepted definition of SA is Endsley's (2000) definition which states that "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (p. 5). The maintenance of SA throughout a flight is a key aspect for ensuring the pilot is able to meet mission objectives, as well as make timely and accurate responses to malfunctions or unexpected events. However, engaging in these tasks during flight can increase the *workload* experienced by the pilot, particularly during key portions of the flight, such as takeoff and landing.

Workload is a construct that describes the dynamic relationship between the amount of cognitive resources demanded by a task and the operator's ability to supply those resources (Wickens & Tsang, 2015). For this study, cognitive resources may be thought of as pools of energy that can be utilized for receiving, interpreting, and/or processing information for the purpose of completing a task (Wickens, 2002). When resource demands exceed the operator's capacity for supplying resources, a state of overload occurs and sub-optimal performance may result (Wickens & Tsang, 2015).

The conceptualization of workload is understood within the context of attentional resource theory. Early resource theory assumed a single, limited pool of attentional-resources that could be allocated towards a task (Kahneman, 1973). If adequate resources are not available to direct towards a task, workload increases and subsequent task performance suffers. However, Wickens (1980, 1984, 2002) proposed a multiple resource model in which there are several, qualitatively different resource pools. According to the multiple resource model, performance on a set of co-occurring tasks is dependent upon the extent to which the individual tasks elicit the use of different resources across four dimensions of resources. The dimensions include: stages of processing, codes of processing, perceptual modalities, and visual channels. The multiple resource model asserts that performance is dictated by the overlap of resources nested within these dimensions, such that tasks with little overlap will yield little to no harmful effects on performance, but tasks with greater overlap will have greater deleterious effects on performance. In other words, more drastic increases in workload result when two tasks draw resources from the same qualitative resource pool. For example, the stages modalities dimension predicts that two tasks requiring cognitive processing (i.e., mental arithmetic and working memory) will likely result in worse performance than if one task required cognitive processing and the other was response driven, i.e., visual tracking (Wickens, 2008; Wickens, Sandry, & Vidulich, 1983).

The multiple resource model provides a conceptual framework for understanding the complexities of workload. However, because many of the tasks associated with operating an aircraft are complex and could engage multiple dimensions within the Multiple Resource Model, it is necessary to reconfigure the model into discrete domains that allow for easier quantification of workload for individual aviation tasks. For this study, we identified the domains as visual, auditory, cognitive, and physical. The aforementioned domains were selected in part based on a previous study that identified subjective workload levels for a variety of aviation tasks from U.S. Army pilots (Webb, Gaydos, Estrada, & Milam, 2010).

The tasks performed in the auditory and visual domains can be defined as the perception and processing of information through the senses of hearing and sight, respectively. Studies have shown significant increases in subjective workload ratings and decreases in performance metrics when subjects are given a secondary task that requires the same domain (visual or auditory) as the primary task (Haas, Shankle, Murray, Travers, & Wheeler, 2000; Phillips & Madhavan, 2011; Taylor, Reinerman-Jones, Szalma, Mouloua, & Hancock, 2013). This can easily be envisioned in a cockpit where a pilot could experience high radio traffic mixed with audio alarms or a pilot monitoring multiple instrument panels and inputs from symbology while flying the aircraft.

The cognitive domain can be defined as understanding information, evaluating situations, and decision making. Increased cognitive demands (i.e., larger amount of information to process, increased complexity of task, or time constraint) typically lead to primary task performance decrements (Gonzalez, 2005; Morrison, Burnham, & Morrison, 2015). Examples of the cognitive domain in a cockpit include route planning or fuel calculations that would need to be performed during flight.

Lastly, the physical domain is defined as physical stress and coordination of gross or fine motor movements. Performance has been shown to be negatively correlated with increased physical task demands or increasing complexity of cognitive tasks paired with physical tasks (DiDomenico & Nussbaum, 2008; Raisbeck & Diekfuss, 2015). Rotary-wing pilots can experience increased workload through adjusting instrument inputs while simultaneously utilizing both hands to operate the cyclic and collective to maneuver the aircraft. Many tasks performed during aviation operations are complex and rely on the interaction of multiple domains. While previous research on the model focused mainly on visual and auditory perceptual modalities, increased attention has been given to cognitive and physical modalities, as indicated in Webb and colleagues' study (2010). Webb and colleagues found a number of tasks were rated as moderate to high in workload across multiple domains, including the aforementioned cognitive and physical domains. Furthermore, while each of these domains typically utilize different resources, an overall depletion of resources can result in an inability to reallocate resources to different tasks as demands change (Wickens, 2002). Any attempt to develop models for measuring workload in an aviation setting likely require studies measuring workload within each of these domains during aviation specific tasks. Previous research has largely focused on measuring workload in simplified tasks within a specific domain. For this reason, the overall aim of the current study is to determine if differences in workload can be measured for individual domains, thereby increasing the predictability of performance breakdowns.

Quantifying Workload

Workload can be measured in a variety of ways including subjective questionnaires, task performance, and psychophysiological measures. In particular, psychophysiological devices are a promising method for monitoring operator workload in real time. Compared to subjective and performance measures, psychophysiological measures have the advantages of continuously monitoring the operator's cognitive state, not interrupting operational flow, and can potentially be useful for driving adaptive aiding systems (i.e., adaptive automation) (Kramer & Weber, 2000; Parasuraman, 2015). If the operator's level of workload can be continuously monitored, this information could potentially be used by brain-computer interfaces and adaptive aiding systems to enhance operator performance (Parasuraman, 2015). Some of the commonly used psychophysiological measurements of workload include electroencephalogram (EEG) for measuring brain activity (Borghini, Astolfi, Vecchiato, Mattia, & Babiloni, 2014), electrocardiogram (ECG) for cardiovascular measures [e.g., heart rate, heart rate variability, blood pressure (Stuiver, Brookhuis, de Waard, & Mulder, 2014)], electrooculography (EOG) to record eye movements (Wickens & Hollands, 1999), and to a lesser extent, respiration (Veltman & Gaillard, 1996).

Changes in task demands have been shown to modulate certain EEG bandwidths (i.e., alpha, beta, theta) (Borghini et al., 2014; Fairclough, Venables, & Tattersall, 2005; Wilson, Fullenkamp, & Davis, 1994). Borghini et al. (2014) reviewed several studies from the fields of aviation and driving that used EEG to quantify workload. Across these studies, Borghini and colleagues found that high workload situations typically resulted in increased frontal midline theta activity and decreased parietal alpha activity. Moreover, researchers have also constructed multivariate measures of workload derived from EEG power spectral densities (PSD) across multiple frequencies and scalp locations. Berka et al. (2007) developed such a metric using linear discriminant function analysis and found this metric to correlate with overt performance decrements during high workload conditions ($r = -.68$) as well as subjective workload ratings ($r = .79$). Historically, activity in the alpha bands have been found to increase during relaxed, wakeful conditions (Okogbaa, Shell, & Filipusi, 1994), while decreasing during complex, cognitively demanding tasks (Slobounov 2000; Fairclough 2005). An increase of power in beta bands has been found with increased alertness and arousal (Okogbaa et al., 1994). Similar to beta bands, theta band activity has been found to increase in the frontal region during focused attention (Doppelmayer, Finkenzeller, & Sauseng, 2008), as well as during a task with time pressure (Slobounov, Fukada, Simon, Rearick, & Ray, 2000). Additionally, a ratio of frontal beta activity to frontal alpha plus frontal theta activity has been suggested as an index of task engagement, as well as a method of simplifying interpretation of the multiple bands examined (Freeman, Mikulka, Prinzel, & Scerbo, 1999; Hockey, Nickel, Roberts, & Roberts, 2009).

ECG activity has also been shown to vary with workload demands. In general, high workload corresponds to increased heart rate (Lenneman & Backs, 2009; Reimer, Mehler, Coughlin, Roy, & Dusek, 2011; Stuver et al., 2014). Past studies have shown that ECG measures correlate with ratings on the National Aeronautics and Space Administration Task Load Index (NASA-TLX), a well validated self-report measure of workload, in air traffic controllers ($r = .95$; Collet, Averty, & Dittmar, 2009) and commercial airline pilots flying a Boeing 747-400 simulation ($r = .81$; Lee & Liu, 2003). Distracted driving studies reveal similar patterns, with increased heart rate observed when drivers perform secondary cognitive tasks in addition to driving (Collet, Clarion, Morel, Chapon, & Petit, 2009; Reimer & Mehler, 2011). Heart rate variability (HRV) is another commonly used measure derived from ECG recordings to measure workload. HRV is the variability between successive R-R peaks in the Q-R-S ECG complex (Thayer, Hansen, Saus-Rose, & Johnsen, 2009). Mandrick, Peysakhovich, Rémy, Lepron, and Causse (2016) found that HRV decreased with increased difficulty on a working memory task. Moreover, higher ratios of low frequency variability (sympathetic) to high frequency variability (parasympathetic) HRV frequency components typically arise in higher workload situations (Durantin, Gagnon, Tremblay, & Dehais, 2014). Mansikka and colleagues (2016) also reported similar low frequency to high frequency ratio findings in F-18 pilots completing various flight maneuvers. Specifically, pilots exhibited the highest low to high frequency ratios during approach maneuvers, which were hypothesized to have the highest workload amongst the maneuvers.

EOG has been used extensively to measure operator workload. Previous research has demonstrated that blink rate and duration are indicative of mental workload and attention, with longer durations of blink indicating reduced alertness (Caffier, Erdmann, & Ullsperger, 2003), and a lower blink rate indicative of increased workload (Bauer, Stroock, Goldstein, Stern, &

Walrath, 1985). Ahlstrom and Friedman-Berg (2006) found that air traffic controller blink rate decreased and subjective ratings of workload increased with higher air traffic volumes. In a review of eye-related measures to quantify driver workload, Marquart, Cabrall, and de Winter (2015) concluded that increased blink latency and decreased blink duration were characteristic of drivers experiencing higher workloads.

Lastly, changes in respiration have been shown to be sensitive to changes in operator workload. Mehler, Reimer, Coughlin, & Dusek (2009) found that while individuals performed a working memory task during a driving simulation, respiration rate significantly increased compared to a driving only condition. Additionally, Karavidas et al. (2010) reported higher respiration rates corresponded to worse performance ($R^2 = .16$) and higher NASA-TLX ratings ($R^2 = .06$) in Boeing 373-800 pilots performing a variable workload simulation. Similar changes in respiration rates and performance decrements have been observed in other aviation studies where workload was manipulated through experience of subjects (Yao et al., 2008) or introduction of secondary tasks (Veltman & Gaillard, 1998).

Overall, the above reviewed literature on psychophysiological measures of workload generally show characteristic patterns in response to changes in task demands. While each of the previously mentioned measures is known to provide a relatively reliable assessment of an operator's cognitive state, there are still several problems associated with their use. One current problem is that the use of multiple psychophysiological measurements occasionally produce divergent results. For example, Matthews, Reinerman-Jones, Barber, and Abich (2015) compared the sensitivity of various psychophysiological indices (HRV, EEG, cerebral hemodynamics, eye tracking) to differing cognitive demands. They found that while the majority of the measures used demonstrated sensitivity to task demands, the measures tended to be sensitive to either the task type or dual tasking workload manipulation but not to both. Moreover, the physiological measures demonstrated poor intercorrelations with each other. The authors asserted that these findings may suggest differential diagnosticity of the various psychophysiological indices used. It may be the case that different psychophysiological indices better reflect workload for different types of tasks or domains, which could assist in determining how to employ automation to reduce the overload by aiding in diagnosing the source of workload. Other studies have also found similar effects. Ryu and Myung (2005) measured alpha oscillation suppression, eye blink interval, and HRV from participants while they performed a tracking and mental arithmetic dual task. Results indicated that alpha suppression was sensitive to arithmetic task demands, while eye blink intervals and HRV were sensitive to tracking task demands. Relatedly, other studies have found both eye movements and increases in heart rate to be associated with higher visual demands, where EEG measures were linked to increased demand during mental calculations (Fallahi, Motamedzade, Heidarimoghdam, Soltanian, & Miakye, 2016; Hankins & Wilson, 1998). Together, these studies indicate a potential connection between psychophysiological measures and domain specificity for workload that warrants further systematic investigation.

Currently, no empirical studies have systematically investigated how multiple physiological measures vary according to different types of task domains hypothesized in the multiple resource model. Therefore, the purpose of this study was to conduct a preliminary validation study of multiple psychophysiological measurements with respect to workload and

specific domains using the multiple resource model as a guide. The current study consisted of two phases. In Study 1, participants from the general population performed laboratory tasks of varying difficulty corresponding to the domains previously mentioned. This phase was conducted to maximize internal validity. Recent studies exploring the sensitivity of these measures have been limited to low-fidelity simulators or basic laboratory tasks with university students. Therefore, a second study employed a high-fidelity JUH-60 Black Hawk simulator and rated aviators to maximize ecological validity. Rated pilots flew several different scenarios corresponding to the domains presented above.

Ultimately, these two studies will assist in setting the foundation for developing a model for measuring workload in real time for Army rotary-wing pilots. Identification of which individual or combined psychophysiological indices are able to identify increased workload demands respective of domain will be critical to building models for real-time monitoring of a pilot's cognitive state. This model will be useful in the future design of adaptive automation systems that would require continuous monitoring of the pilot's workload. If the aviator workload can be accurately measured across multiple domains with the fewest number of physiological sensors, physiological data from these sensors may be useful in advancing adaptive aiding systems that can augment aviator performance and enhance Warfighter lethality. Moreover, the evaluation of new cockpit systems in terms of the workload imposed on the pilot may be improved by identifying physiological measures that can better characterize sources of increased workload.

Study 1

The objective of Study 1 was to establish relationships between psychophysiological recordings and performance as well as perceived workload using validated laboratory tasks taxing each of the domains: physical, visual, auditory, and cognitive. The results of Study 1 serve as a means of comparison for Study 2 employing multi-domain aviation tasks in the simulator.

Method

Participants. Sixteen healthy adults (13 males, 3 females) participated. The mean age was 31.33 ($SD = 9.72$) years. All participants completed the Beck's Depression Inventory and the State-Trait Anxiety Index and scored below the published cut-off criterion suggesting all participants were free of depression and anxiety symptoms, respectively (Beck, Steer, & Brown, 1996; Ercan, Hafizoglu, Ozkaya, & Kirli, 2015). The mean Shipley Institute of Living Scale total score was 57.5 ($SD = 9.94$) suggesting slightly above average intelligence in the sample with two participants scoring below average. With respect to education, 8 participants reported having some college education, 4 associate's degrees, 2 graduate degrees, 1 bachelor's degree, and 1 high school diploma. Fourteen participants were right-handed, one left-handed, and one reported being ambidextrous. Three participants reported regular tobacco use and 10 regular caffeine use. All participants had either uncorrected or corrected-to 20/20 vision, no other visual impairments, and no hearing impairments. The average body mass index was 28.65 ($SD = 4.35$).

Materials. Participative and objective instruments were used to measure performance, individual differences, workload, and psychophysiological data.

Questionnaires. All questionnaires and survey instruments are presented in Table 1.

Table 1. Descriptive Summary of Questionnaires and Survey Instruments

Variable	Description
Demographics and personal history questionnaire	Brief questionnaire to collect basic demographic information including age, sex, rank, ethnicity, and education history as well as personal health and wellness information including tobacco use, current health, any current medications taken, attention-related diagnoses, handedness, vision, and body mass index.
Beck's Depression Inventory – II	The BDI-II is a commonly used 21-item, multiple-choice self-report that captures affect, cognitions, and physical symptoms of depression over the previous two weeks before completion. On this measure, higher scores indicate greater endorsement of depression symptoms (Beck, Steer, & Brown, 1996).
Sleep Wake Questionnaire	The Sleep-Wake Questionnaire (SWQ) is a simple self-report survey that was developed in-house to collect information on the participant's prior night of sleep (e.g., time in bed the night prior to the study, wake time on the day of the study, if any naps were taken, and judging the quality of their sleep and current level of sleepiness on a 5-point Likert scale).
NASA Task Load	The NASA Task Load Index (NASA-TLX) (Hart & Staveland, 1988) is a questionnaire that measures participative workload. Participants rate each iteration of each task on the following categories, using a 100-point scale: mental demand, physical demand, temporal demand, performance, effort, and frustration. The NASA-TLX then provides a total workload score and scores for the six subscales.
Shipley Institute of Living Scale	The Shipley Institute of Living Scale (SILS) (Shipley & Burlingame, 1941) assesses general intellectual functioning in adults and adolescents and serves as a screening tool for cognitive impairment in individuals with normal original intelligence (Matthews, Orzech, & Lassiter, 2011). The SILS takes a maximum of 20 minutes to complete, and yields three major summary scores: Vocabulary, Abstraction, and combined Total scores. The Vocabulary sub-scale consists of 40 multiple-choice verbal reasoning questions and primarily taps crystallized intelligence. The Abstraction subscale includes 20 series completion items of inductive reasoning that tap fluid ability.
State Trait Anxiety	The State-Trait Anxiety Inventory (STAI) (Spielberger & Gorsuch,

Index

1983) is a widely used 40-item, self-report anxiety inventory employing a 4-point Likert-type scale that captures two types of anxiety: state, or event-dependent anxiety, and trait, or persistent demonstrations of anxiety as a personal characteristic. Anxiety scores on the STAI are calculated by reverse-coding select responses and then summing the total point values of the items, with higher scores indicating higher levels of anxiety for both the state and trait subscales.

Psychophysiological recording.

Electrocardiogram (ECG). The Biopac BioNomadix system was used to collect the electrocardiogram and respiratory data. The BioNodadix system consists of wireless, wearable data collection devices and is equipped with AcqKnowledge software. The AcqKnowledge software was used to display, control, analyze, and export data collected from the BioNomadix system.

Data were collected through the BioNodamix electrocardiogram amplifier module (ECG100). The ECG100 is a single channel, high gain, differential input, biopotential amplifier designed specifically for monitoring the heart's electrical activity. Single-lead electrodes were placed on each of the participant's clavicles and one below the right pectoral area, and sampled at a rate of 1,000 Hz. The ECG100 is equipped with an R-wave detector function, which was used to provide an output signal of smooth positive peaks every time the R-wave was detected. The R-wave data was used to calculate beats per minute (BPM) and inter-beat intervals (IBI). Following data collection, the ECG data was filtered using a high pass filter with a Hamming window at 1 Herz (Hz), a low pass filter with a Hamming window at 35 Hz, and a band stop filter at 60 Hz to reduce interference from electrical lines. After the data were filtered, each participant's ECG data was visually inspected for artifacts. Identified artifacts were corrected using linear interpolation. AcqKnowledge's (AcqKnowledge 4.4.1) automated HRV analysis algorithm was applied to each participant's ECG data to derive the very low (VLF; 0—0.04 Hz), low (LF; 0.04—0.15 Hz), and high frequency (HF; 0.15—0.4 Hz) values. The low-frequency and high-frequency values were used to derive the low/high frequency ratio. The AcqKnowledge software derived these values through the following three stages: 1) The respiration rate (RR) intervals were extracted from the ECG signal using a modified Pan-Tompkins QRS detector (Pan & Tompkins, 1985); 2) the RR intervals were re-sampled at a continuous sampling rate to extract frequency information using cubic-spline interpolation; 3) the frequency information was extracted from the RR intervals and analyzed to produce standard ratios, with power sums reported in units of sec^2 . Finally, a Welch periodogram generated the power spectral density value of VLF, LF, and HF. The IBIs were calculated as a mean of the RR intervals, and was used as a measure of heartrate. Overall averages of the LF/HF ratio was used as an index of HRV and IBI was used as index for HR for each flight scenario for data analyses.

Respiration Data. Respiration data was collected using the BioNomadix respiration pneumogram amplifier module (RSP100C), which is a single channel, differential amplifier designed specifically for recording respiration effort. Respiration was collected through a

transducer strap placed on the chest or abdomen of the participant and sampled at a rate of 50 Hz. After data collection the respiration data was filtered using a high-band pass filter at 0.05 Hz, and then visually inspected for artifacts, corrected using linear interpolation. The AcqKnowledge (AcqKnowledge 4.4.1) software extracted respiration rate in breaths per minute using the peak-to-peak values between respiration waves. The average respiration rate (RR) per flight scenario were used in data analyses.

Electroencephalogram (EEG). The B-Alert X-24 wireless wet electrode system was used to record EEG activity. The X-24 incorporates 20 channels corresponding to scalp locations according to the International 10-20 system (frontal channels: Fp1, Fp2, F7, F3, Fz, F4, F8; central channels: C3, Cz, C4, T3, T4; parietal and occipital channels: P3, POz, Pz, P4, T5, T6, O1, O2). Power spectral density (PSD) values were computed using the automated algorithms provided through the B-Alert Live Software (Advanced Brain Monitoring, 2009). Prior to computing PSD values, artifacts were identified and removed using the ABM algorithms for artifacts associated with electromyography (EMG), eye blinks, excursions, saturations, and spikes (B-Alert Live, 2009). The software was used to perform Fast Fourier Transform (FFT) on the data and calculate the amplitudes of the sinusoidal components for designated frequency bins. Frequency domain variables are based on the PSD derived after application of a 50% overlapping window, and a FFT with and without application of a Kaiser window. Each window size was defined as 1 epoch containing 1 s of data (256 decontaminated EEG samples). The software then provided PSD values ranging from 1 to 40 Hz for each EEG channel that were logged to obtain a Gaussian distribution. Selected 1—Hz bins were averaged, then logged to create the EEG bands used in analyses (theta: 4—8 Hz; alpha: 9—13 Hz; beta: 14—30 Hz). Frontal theta, alpha, and beta PSD values were examined from the following frontal channels: Fp1, Fp2, F7, F3, Fz, F4, and F8, with values averaged across each task as well as the ratio of beta values to alpha plus theta values were used in data analyses.

Electrooculogram (EOG). Data was collected through the EEG device. EOG was used to assess vertical eye blink activity by placing Ag/AgCl electrodes above and below the left eye. EOG data were processed within MATLAB using an algorithm developed in-house. The algorithm was modeled after the blink detection algorithm described by Epling et al. (2015). Blink rate and blink duration were derived through a blink detection algorithm based on the vertical electrooculogram (VEOG). A blink within the VEOG signal is characterized by a waveform with a sharp rise immediately followed by sharp fall, which is short in duration and consists of an overshoot before the signal returns to zero. To remove high-frequency noise from the data, the data were filtered using a low-pass filter, smoothing algorithms were also applied. Blinks were then detected by identifying the beginning and end of blink based on a sharp rise in rate of change (beginning of blink), which was determined by a threshold of 10, followed by a sharp decrease to negative rate of change (end of blink), which was determined by a threshold of -7. Eye blinks were detected and scored based on when the VEOG signal went above and below the threshold, and from this data blink rate and duration metrics were derived, and averaged across each task for data analyses.

Cognitive Tasks

Auditory task. The auditory domain was represented with an oddball task. Participants

listened to pure sine-wave tones presented binaurally through headphones. Two different tones were presented. One tone was a lower frequency (1,000 Hz) and the other tone was a higher frequency (2,000 Hz). The low frequency tone was considered common and occurred on 80% of trials, while the high frequency tone was considered deviant or infrequent and occurred on 20% of trials. Each tone was presented for 100 ms with a 10-ms rise and fall time. Inter-stimulus intervals randomly varied between 800 and 1200 ms. In order to minimize any muscle movements or artifacts in the data, participants remained silent and still with the exception of their right index finger in accordance with the task. Three different workload conditions of the task were completed, each presenting 200 sequential stimuli. Condition 1 was a no-workload or control condition; a simple odd ball detection task (Kramer & Weber, 2000; Trejo, Kramer, & Arnold, 1995), which required participants to respond as quickly as possible when a target was detected (infrequent 2,000-Hz tones) by pressing the spacebar with their dominant hand and inhibit the responses on non-target trials (frequent 1,000-Hz tones). Condition 2 was a low-workload condition; a counting forward task where participants responded by pressing a button when a target was heard and also to further count the number of target occurrences and report this total at the end of the task. Finally, Condition 3 was a high-workload condition; a counting backward task which required participants to respond to the target and then to subtract one day from the current date for each target detected (i.e., counting in reverse order) and finally reported the date calculated at the end of the session. At the start of the task, the volume was adjusted to a comfortable, participative volume for each participant (typically ranging from 80 to 90 dB). Between each workload condition, a 3-minute (min) break was given. Reaction times (RTs) and accuracy were recorded. Completion of all three versions required approximately 30 min. The versions of the task have been validated for purposes of studying workload (Horat et al., 2016).

Visual task. For the visual domain, participants performed a visual search task at two workload levels: low and high. For the low-workload condition, each trial consisted of four letters were presented in a vertical column (display duration = 2000 ms), and participants determined whether the target letter (“X”) was present (see Figure 1 right panel). If the letter X was present, participants were instructed to press the spacebar as quickly as possible. If the letter X was not present, participants made no response. Between trials, a blank screen was presented for 150 ms. In the high workload condition, four columns with four letters each were presented. Participants had to identify if the letter X was present by pressing a key that was spatially mapped to the column in which the X was displayed (i.e., the leftmost display column used the leftmost response key and the rightmost display column used the rightmost response key with intermediate keys correspondingly mapped). The parameters of this task for high and low workload conditions were previously validated (Colle, 2010; Cassenti & Kelley, 2006, 2008). Completion of the task (both high and low-workload conditions) required approximately 30 minutes.

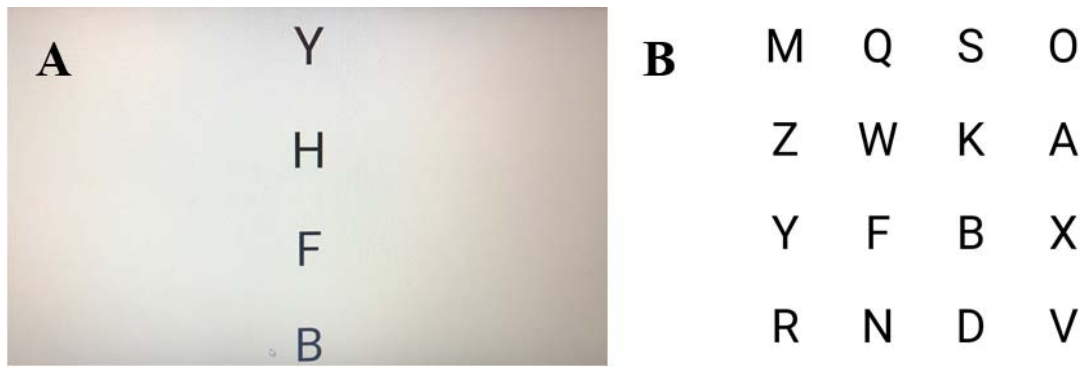


Figure 1. Example of a trial in the visual task: A) Low workload and B) High workload.

Physical task. The Multi-Attribute Task Battery-II (MATB-II) (Santiago-Espada, Myer, Latorella, & Comstock, 2011; Santiago-Espada, 2014), an updated version of the Multi-Attribute Task Battery (Comstock & Arnegard, 1992), simultaneously presents four tasks that are generalizations of piloting tasks. Only the tracking task was completed in this study. The tracking task (see Figure 2) required participants to keep a circular reticle within a box centered between the intersection of two crosshairs with the use of a joystick (positioned in the dominant hand). The MATB-II interface introduced random movement of the circular reticle with the use of a 4:3 horizontal-vertical sine wave function. Participants performed two workload conditions: low and high. In the low-workload condition, the random reticle movement was set to a low update rate according to MATB-II settings, resulting in less reticle movement. In the high-workload condition, the random reticle movement was set to a high update rate, resulting in high reticle movement. Each condition was 10-min long and were separated by a 3-min break. Before the testing conditions, participants performed a 5-min practice session. These settings were previously validated as methods to manipulate workload (Fairclough et al., 2005). Task completion required approximately 20-30 minutes.

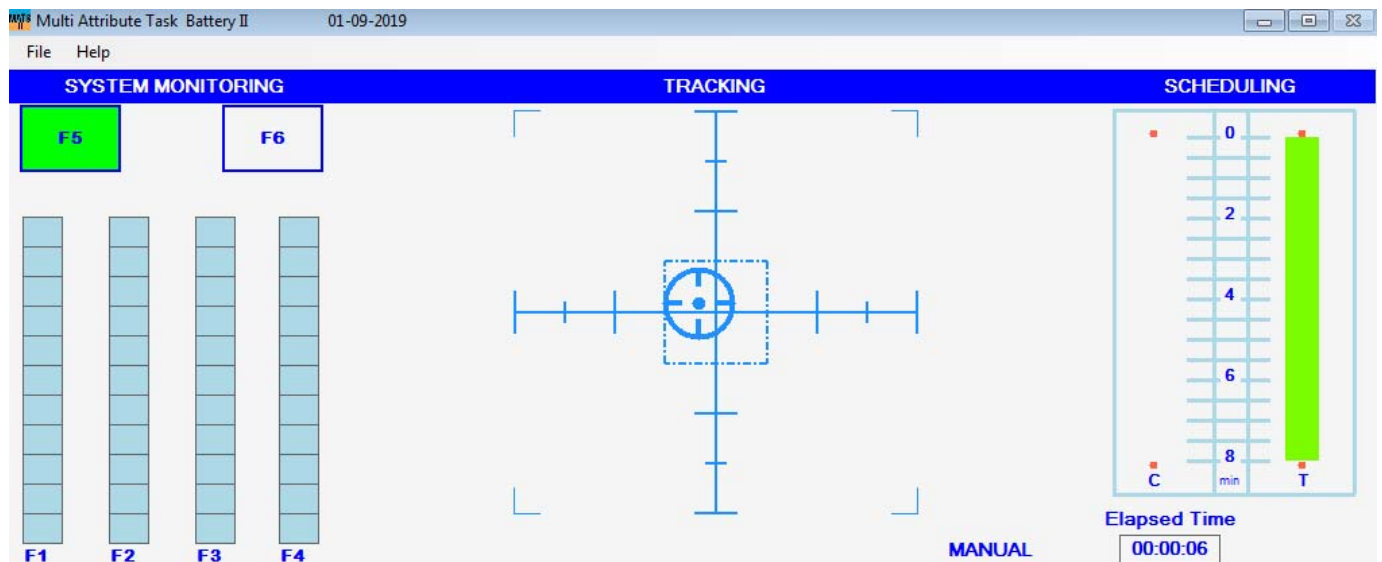


Figure 2. Example of a trial in the MATB-II task.

Cognitive task. A mental arithmetic task was used to elicit the cognitive domain. Participants completed 50 two-digit multiplication problems. The first five trials were used as a short training and were excluded from data analyses. The remaining 45 trials were presented in three blocks composed of 15 trials each. The three blocks corresponded to low, medium, and high workload conditions. Block 1 contained the 15 easiest multiplications (outcomes ranging between 80 and 108), Block 2 contained 15 multiplications of intermediate difficulty (outcomes between 126 and 192), and Block 3 contained the 15 hardest multiplications (outcomes between 221 and 324). Each trial began with two fixation xx's in the center of the screen for 4,000 ms. Then, multiplicand was presented for 1,000 ms, followed by two fixation xx's presented for 1,500 ms, and finally the multiplier presented for 1,000 ms. The two fixation xx's appear once again for 10 s, and the participant computed the product. After the 10 s period, participants were prompted to enter their answers with the keyboard. After striking the 'enter' key to submit their answer, the next trial began. This task has been validated for manipulating cognitive workload (Hwang, Kim, Hawngbo, & Oh, 2014; Marquart & de Winter, 2015; Siegenthaler, et al., 2014). Task completion required approximately 20-30 min.

Procedure. The U.S. Army Medical Research and Development Command Office of Research Protections Institutional Review Board reviewed and approved the protocol for this study. All procedures were conducted according to institutional ethical standards.

Prior to participation, all participants provided written informed consent. Following study enrollment, participants completed the demographics and personal history questionnaire, SWQ, BDI, STAI, and SILS. Participants were required to have 8 hours of sleep the night before the study, with a minimum of 6 accepted for participation (confirmed by self-report), refrain from taking over the counter medications that may induce drowsiness, refrain from any alcohol or sedatives 24 hours prior to testing, and refrain from any use of nicotine 2 hours and caffeine 16

hours prior to testing. These criteria were enforced to reduce variability in physiological recordings (e.g., Gilbert, Dibb, Plath, & Hiyane, 2000 [caffeine, nicotine]; Kähkönen, Wilenius, Nikulin, Ollikainen, & Ilmoniemi, 2003 [alcohol]) to have good baseline psychophysiological recordings. Then, participants were introduced to the NASA-TLX and how to complete it.

Next, participants were fitted with the psychophysiological recording devices. A 5-min baseline for ECG and respiration was collected by having the participant sit and look at a blank wall with instructions to not talk and relax. Participants then completed three tasks associated with the EEG system that were not included in analyses in this report. Finally, participants completed the scanning/visual search tasks (two administrations – low and high workload), auditory discrimination tasks (three administrations – control, low, and high workload), mental arithmetic tasks (three administrations – low, moderate, and high workload), and a manual tracking task (two administrations – low and high workload). Each task began with participants completing a practice session for the respective tasks. Workload was manipulated for each task using previously validated parameters and tasks for each the workload conditions (e.g., high, moderate, low) and the domains (i.e., visual, auditory, cognitive, and physical) were counterbalanced amongst participants. Following each task workload condition, participants completed the NASA-TLX.

Statistical analysis and quality control. All hand-entered data were double-checked for accuracy using a 10% validation check. Prior to analyses, all electronically records were inspected for any impossible values or output errors. Distributions of all performance and TLX ratings were evaluated for normality, and any outliers exceeding three standard deviations from the mean were excluded from analyses. Psychophysiological data were inspected for any impossible values. To determine the efficacy of the workload manipulations, paired-samples *t*-tests were used to evaluate outcomes from the tracking and visual tasks (given two levels of workload) and repeated-measures analyses of variance (ANOVAs) to evaluate outcomes from the auditory and cognitive tasks (given three levels of workload). Similarly, repeated-measures ANOVAs and paired-samples *t*-tests were used to estimate effects of workload on psychophysiological outcomes. Finally, correlational analyses were used to evaluate the concurrent validity of the psychophysiological measurements as indices of workload (relative to the NASA-TLX scores). Statistical significance was set at $\alpha = .05$ and η_p^2 is reported as the effect size for omnibus ANOVAs.

Results

The results for Study 1 are divided into assessments of the workload manipulations in terms of performance and NASA-TLX, effects of the workload manipulations on the psychophysiological indices, and correlations between the psychophysiological indices, performance, and NASA-TLX.

Assessment of the Workload Manipulations: Performance. Repeated-measure ANOVAs were conducted for performance on the different tasks as a function of workload. For those tasks with more than two levels of workload, follow-up comparisons were conducted following significant ANOVAs to isolate the sources of the effect. The means and standard deviations for the different tasks under different levels of workload are provided in Table 2. For the auditory discrimination task, there was no significant effect of the workload manipulation on

the number of correct responses, $F(2, 7) = 0.19, p = 0.85$. For the visual search task, there was a significant effect of the workload manipulation on the total number of correct responses, $F(1, 13) = 13.30, p = 0.003, \eta p^2 = 0.51$. As seen in Table 2, the total number of correct responses was lower in the high workload condition in the low workload condition.

For the mental arithmetic task, there was a significant effect of workload condition on the percentage of correct responses, $F(2, 13) = 11.93, p = 0.001, \eta p^2 = 0.65$. As seen in Table 2, follow-up comparisons revealed significant differences between all three levels of low, medium, and high workload. For the manual tracking task, there was a significant effect of the workload manipulation on the accuracy of maintaining the target on the center point, $F(1, 14) = 364.94, p < 0.001, \eta p^2 = 0.96$. As seen in Table 2, accuracy was much higher in the low-workload condition than the high-workload condition.

Table 2. Performance on the Different Tasks as a Function of Workload

Task	Workload					
	Low		Medium		High	
	Mean	SD	Mean	SD	Mean	SD
Auditory Discrimination Task	197.00 ^a	0.27	197.33 ^a	2.83	196.44 ^a	4.33
Visual Search Task	39.93 ^a	2.92	N/A		38.57 ^b	1.45
Mental Arithmetic Task	66.67 ^a	33.24	46.67 ^b	30.55	22.67 ^c	31.15
Manual Tracking Task	18.44 ^a	7.04	N/A		31.43 ^b	7.63

Note. Means with a different superscript differ significantly at $p < .05$.

Assessment of the Workload Manipulations: NASA-TLX Ratings. Repeated-measure ANOVAs with appropriate follow-up comparisons were also conducted for NASA-TLX ratings of the different tasks as a function of workload. The means and standard deviations for the different tasks under different levels of workload are provided in Table 3. For the auditory discrimination task, although there was no significant effect of the workload manipulation on performance, there was a significant effect on the NASA-TLX ratings, $F(2, 13) = 6.44, p = 0.001$. As seen in Table 3, follow-up comparisons revealed significant differences between the TLX ratings for all three conditions of workload. For the visual search task, there was a significant effect of the workload manipulation on the NASA-TLX ratings, $F(1, 14) = 22.27, p < 0.001, \eta p^2 = 0.61$. Ratings were higher in the high workload condition than in the low workload condition (see Table 3).

Table 3. NASA-TLX Ratings on the Different Tasks as a Function of Workload

Task	Workload					
	Low		Medium		High	
	Mean	SD	Mean	SD	Mean	SD
Auditory Discrimination Task	23.02 ^a	13.56	35.09 ^b	20.50	44.22 ^c	26.00
Visual Search Task	21.20 ^a	15.74	N/A		40.73 ^b	20.68
Mental Arithmetic Task	51.09 ^a	29.65	56.56 ^{ab}	28.16	61.69 ^b	27.51
Manual Tracking Task	43.77	20.90	N/A		46.83	18.78

Note. Means with a different superscript differ significantly at $p < 0.05$.

For the mental arithmetic task, although there was a significant effect of the workload manipulation on performance, the effect of the manipulation was not significant for the NASA-TLX ratings, $F(2, 13) = 2.44$, $p = 0.126$. However, follow-up comparisons did reveal a difference in the TLX ratings between the low and high conditions of workload. For the manual tracking task, there was no significant effect of the workload manipulation on NASA-TLX ratings, $t(12) = -1.60$, $p = 0.14$, despite the differences seen in performance. These results highlight the dissociations that often occur between performance and ratings of workload (Wickens & Tsang, 2015).

Effect of Workload on Psychophysiological Recordings

ANOVAs did not support an effect of workload on EEG outcomes (total frontal theta, beta, and alpha activity, respectively) for the mental arithmetic task or the manual tracking task. There was an effect of workload on frontal alpha activity during the oddball task such that activity increased with more difficult workload levels ($F(2,22)=5.67$, $p = 0.01$). Similarly, theta activity decreased with higher levels of workload during the visual task ($t(13) = 2.43$, $p = 0.03$).

Heart rate increased on high-workload levels of the visual task compared to low levels ($t(14) = -2.76$, $p = 0.015$). No effects emerged for the other tasks, and HRV was not affected by workload level on any of the tasks.

Respiration rate was not affected by workload level in any of the tasks.

Correlations between performance, psychophysiological recordings, and perceived workload

Analyses suggested that perceived workload and performance were correlated in the mental arithmetic task ($r(32) = -0.62, p < 0.001$) such that performance decreased as perceived workload increased. Performance and perceived workload did not significantly correlate for the other three tasks.

Frontal alpha activity and perceived workload on the visual task were significantly correlated ($r(29) = 0.41, p = 0.024$), such that activity increased as perceived workload increased. Similarly, frontal theta activity correlated with perceived workload on the manual tracking task ($r(28) = 0.44, p = 0.02$). Additionally, differences in perceived workload between the high and low workload levels on the auditory task correlated with changes in frontal alpha activity ($r(11) = 0.65, p = 0.03$) and frontal beta activity ($r(11) = -0.64, p = 0.03$) such that alpha activity increased with workload and beta activity decreased with workload. For the cognitive task, heart rate and RR correlated with differences in perceived workload between high and low levels of workload such that heart rate and RR increased with perceived workload ($r(14) = 0.60, p = 0.02; r(14) = 0.67, p < 0.001$; respectively).

Study 1 Discussion

The results suggest that the workload manipulations showed differences in performance for all tasks and differences in perceived workload for three of the tasks. No differences in perceived workload were apparent in the manual tracking task. Note that informal feedback from the participants suggests that all levels of the cognitive task were challenging. Study 1 suggests differential patterns in psychophysiological recordings and associations with workload levels by workload domain. Specifically, relationships were seen between perceived workload and frontal alpha activity on the visual and auditory tasks, frontal beta activity on the auditory task, and frontal theta activity on the manual tracking task. Relationships between heart rate and respiration rate and perceived workload were seen in the cognitive task. These initial findings align with past workload research with respect to increases and decreases in EEG recordings. Given the complexity and multi-domain nature of the flight tasks completed in Study 2, these patterns were used as a means of comparison for Study 2 results in hypothesizing the workload domains manipulated.

Study 2

Method

Participants. Twenty-three participants completed the second study. All participants were male, aged 30 to 53 years ($M = 36.13, SD = 4.99$). All participants completed the Beck's Depression Inventory and the State-Trait Anxiety Index and scored below the published cut-off criterion suggesting all participants were free of depression and anxiety symptoms, respectively (Beck, Steer, & Brown, 1996; Ercan et al., 2015). Participants were also predominantly right-handed ($n = 18$), three left-handed, and two ambidextrous. Participants' body mass index (BMI) ranged from 23.71 to 35.15 ($M = 27.56, SD = 2.96$). All participants had either uncorrected or corrected-to 20/20 vision, no other visual impairments. One participant reported tinnitus, but was

determined eligible by an in-house audiologist. Five participants reported tobacco use. All participants had deployed at least once, with years in their military aviation career ranging from 5 to 28 ($M = 10.26$ years, $SD = 6.11$ years). Average flight hours in the last year ranged from half an hour up to 25 hours ($M = 7.81$, $SD = 6.14$).

Materials. Self-report and objective instruments were used to measure performance, individual differences, workload, and psychophysiological data.

Questionnaires. The same questionnaires were used in Study 2 as Study 1 (see Table 1), with the addition of the Workload Profile and aviation-specific questions in the demographic questionnaire.

Workload Profile. The Workload Profile (WP) (Tsang & Velazquez, 1996) is a rating technique that asks participants to rate the proportion of attentional resources they used for each task after completing a set of tasks, where participants are asked to rate across each of the following eight dimensions: perceptual/central processing, response processing, spatial, verbal, visual, auditory, manual, and speech. Each dimension is rated using a number ranging from 0 to 100, with 0 equal to no demand and 100 equal to maximum attention. These were summed for an overall workload rating for each task.

Psychomotor Vigilance Task. The Psychomotor Vigilance Task (PVT) (Dinges & Powell, 1985) assesses sustained attention by requiring participants to respond to a visual stimulus. The PVT provides the participant's reaction time to responding to the stimulus. The PVT was administered on a laptop computer using a standard mouse for responses. The version of the PVT used was developed by Khitrov and colleagues (2014), and has been demonstrated to be valid and comparable to the original handheld version of the PVT. The PVT trial time was set for 3 min. The outcome measures assessed included mean RT for correct responses and number of lapses (RTs greater than or equal to 500 ms). The PVT was included as a measure of divergent validity for the psychophysiological metrics, given that it is a sustained attention task that is not necessarily domain specific.

Psychophysiological recording. The same recording devices were used in Study 2 as outlined for Study 1. Baseline tasks and procedures were the same, as well.

Black Hawk Flight Simulator. A full-motion Black Hawk simulator was used to create simulated flight scenarios (described later) designed to induce high and low workload conditions. The simulator consists of a simulator compartment containing a cockpit, instructor/operator station, observer station and a six-degree-of-freedom motion system. A Dell Precision laptop received information concerning changes in the aircraft/simulator state parameters at a 60-Hz capture rate. Settings in the simulator were manipulated to create the experience of a high or low workload by manipulating the level of the clouds, the visibility, and flight system functionality; Table 4 outlines these differences.

Flight scenarios. Eight operationally relevant flight scenarios, lasting approximately eight to 10 minutes each, were developed by in-house research pilots and validated for workload manipulations by three independent pilots. Each scenario consisted of a high and low workload manipulation, resulting in a total of 16 flight scenarios to be performed by participants. The

flights were developed so that there were four scenarios (two high workload, two low workload) for each of the identified workload domain (auditory, physical, mental, visual), thus resulting in 16 (four workload X four domains). Two sets of high and low workload manipulations for each domain were developed given this is a new approach to assessing workload domains using operationally relevant tasks. The scenarios and performance outcome variables of interest for each of the scenarios are identified in Table 4. Examples of flight routes are also depicted in Appendix A, Figures 1 through 3.

Table 4. Workload Manipulation Descriptions of Flight Scenarios

Domain	Task	Low Workload	High Workload
Visual	Terrain Flight*	Day	Night
	Terrain Flight MEDEVAC*	Day	Low visibility and precipitation
Cognitive	Respond to Emergency	#1 Engine failure: Return to airfield	#1 Fuel filter bypass: Continue mission or return to base
	Respond to Inadvertent Instrument Meteorological Conditions (IIMC)	Highly competent Co-Pilot	Non-functioning Co-Pilot
Auditory-Verbal	Perform Radio Communications	Low call volume	High call volume
	Perform Rescue Hoist Operations	Calm winds	High winds
Physical/Fine Motor	Perform Deck Landing and Boost Off Flight	Deck landing	SAS 1/2, trim FPS, and boost off flight
	Perform Non-Precision Approach	Low winds	High, gusting crosswinds and moderate turbulence

Note. *Different flight paths.

Table 5. Descriptions of Variables of Interest for Flight Scenarios

Task	Workload Domain	Workload level	Variables Measured
Terrain Flight	Visual	Low	Time to LZ; speed deviations (120 kn ^a instructed); duration of each Δ Altitude (ALT ^b) (200' AGL ^c instructed)
		High	
Terrain Flight MEDEVAC	Visual	Low	Time to LZ; speed deviations (120 kn instructed); duration of each Δ ALT (150' AGL instructed)
		High	
Respond to Inadvertent Instrument Meteorological Conditions	Cognitive	Low	Land / crash ; speed deviations (120 kn instructed) taken from first three minutes of flight in high condition/total flight in low condition
		High	
Respond to Emergency	Cognitive	Low	Land/crash; speed deviations (120 kn instructed) taken from first three minutes of flight in low condition/total flight in high condition
		High	
Perform Radio Communications	Auditory	Low	Proportion of correct responses, calculated from: no. correct responses, no. missed responses, & no. correct rejections; duration altitude deviations (400' AGL instructed); speed deviations (120 kn instructed)
		High	
Perform Rescue Hoist	Auditory	Low	Altitude deviations during hover (deviations from 50' AGL); Duration of altitude deviations en route (300' AGL instructed)
		High	
Perform Deck Landing and Boost Off Flight	Physical	Low	Airspeed deviations (120 kn instructed); heading deviations (056 instructed)
		High	
Perform Non-Precision Approach	Physical	Low	Airspeed deviations (120 kn instructed)
		High	

Note. ^akn = knots ^bALT = altitude ^cAGL = above ground level

Procedure. The U.S. Army Medical Research and Development Command Office of Research Protections Institutional Review Board reviewed and approved the protocol for this study. All procedures were conducted according to institutional ethical standards.

Participants were required to follow the same sleep instructions and caffeine/nicotine/alcohol consumption guidelines as Study 1. These were all confirmed by self-report when participants arrived at the Laboratory. Participants provided written informed consent prior to engaging in research activities. Once consent was obtained, participants completed the demographics and personal history questionnaire, as well as the remaining questionnaires, except for the NASA-TLX and WP. Upon completion of the questionnaires, participants were instructed on how to complete the NASA-TLX and Workload Profile.

Participants were then fitted with the psychophysiological recording devices and completed baseline measurements, as in Study 1. Following this, the research pilot briefed the participant on the tasks to be completed within the simulator and escorted the participant to the simulator to familiarize. The research pilot ensured all participants were able to adequately perform the tasks required during the scenarios. Participants then completed the 16 flight scenarios in a randomized order. Upon completion of each scenario, the participants completed the PVT while in the cockpit and then the NASA-TLX. At the end of a segment of eight scenarios, participants completed the WP, thus it was completed twice. Breaks were taken as needed.

Statistical analysis and quality control. All hand-entered data were double-checked for accuracy using a 10% validation check. Prior to analyses, all electronically recorded data were inspected for any impossible values or output errors. Distributions of all performance and TLX ratings were evaluated for normality and any outliers exceeding three standard deviations from the mean were excluded from analyses. Psychophysiological data were inspected for any impossible values and are reported below when applicable. To determine the efficacy of the workload manipulations, paired-samples *t*-tests were used to evaluate performance, NASA-TLX and WP, and psychophysiological outcomes (given two levels of workload). Correlational analyses were used to evaluate the concurrent validity of the psychophysiological measurements as indices of workload (relative to the NASA-TLX scores). Lastly, correlational analyses were used to evaluate divergent validity with psychophysiological and PVT data.

Results

The results for Study 2 are divided into assessments of the workload manipulations in terms of NASA-TLX, WP and performance, effects of the workload manipulations on the psychophysiological indices, and correlations between the psychophysiological indices, performance, and NASA-TLX. With PVT correlations reported last.

Assessment of the Workload Manipulations: Performance Data

Technical difficulties in data recording resulted in missing data for one participant on all flights, whereas four subjects had missing data for some scenarios and not others. Sample sizes are reported with descriptive statistics to indicate missing values. There were no significant outliers in the data.

Visual Domain. Workload manipulation was confirmed for the *terrain flight* based on significant differences in both *altitude* and *airspeed deviations*. *Altitude deviations* were greater during the *high workload* compared to the *low workload* condition ($t(19) = 3.46, p = 0.003$). Similarly, *airspeed deviations* were greater during the *high workload* compared to *low workload* ($t(19) = 2.16, p = 0.04$) condition. No differences were found in *time to LZ*. Manipulation of workload was not confirmed for the *MEDEVAC task*.

Table 6. Performance Descriptive Statistics Visual Domain

Task	Metric	n	Low		High		
			M	SD	n	M	SD
Terrain Flight	Altitude Deviations	20	19.85*	8.04	20	25.25*	8.23
	Airspeed Deviations	20	20.55*	5.71	20	24.55*	6.47
	Time to Landing Zone	19	10.10	0.58	19	10.40	1.18
Terrain Flight MEDEVAC	Altitude Deviations	19	8.81	8.46	19	12.16	7.47
	Airspeed Deviations	19	33.21	5.14	19	32.85	5.16
	Time to Landing Zone	19	7.37	0.48	19	7.51	0.53

Note. Means with a superscript differ significantly at $p < 0.05$.

Cognitive Domain. During the *IIMC task*, *airspeed* metrics were examined by comparing the first three minutes, as the *low workload* condition emergency occurred two minutes into the flight and then the research pilot took controls. There was an effect of workload during this task ($t(20) = 12.91, p < 0.001$), with greater deviations during the *low workload* compared to *high workload* condition, which is the opposite of what was to be expected. There were also more crashes during the *high workload* (two crashes) than the *low workload* condition (none), although this difference is not significant.

During the *respond to emergency* task, performance metrics were examined by comparing the first three minutes, as the emergency occurred two minutes into the flight and then the research pilot took controls in the *high workload* condition. There was no effect of workload ($t(19) = -1.38, p = 0.18$), although again *airspeed* deviations were slightly higher in the *low workload* compared to the *high workload* condition. There were no crashes during this task.

Table 7. Performance Descriptive Statistics Cognitive Domain

Task	Metric	Low Workload			High Workload		
		n	M	SD	n	M	SD
IIMC	Airspeed	21	25.73*	6.21	21	3.06*	3.64
	No. of Crashes	23	Frequency 0		23	Frequency 2	
Emergency	Airspeed	20	30.49*	13.94	20	25.23*	8.76
	No. of Crashes	23	Frequency 0		23	Frequency 0	

Note. Means with a superscript differ significantly at $p < 0.05$.

Auditory Domain. There was a significant difference on *airspeed deviations* during the *radio calls task* ($t(19) = 5.68, p < 0.001$), with greater *airspeed deviations* during the *high workload* compared to *low workload* condition. Responses during the *radio calls task* also had an effect of workload ($t(22) = -3.04, p = 0.006$), with a larger proportion of correct responses during the *low workload* compared to *high workload* condition. There was no effect of workload during the *rescue hoist task*.

Table 8. Performance Descriptive Statistics Auditory Domain

Task	Metric	Low Workload			High Workload		
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Radio Communications	Airspeed Deviations	20	11.46*	3.50	20	18.57*	2.96
	Altitude Deviations	20	8.08	9.02	20	8.65	6.43
	Proportion Correct Calls	23	1.00*	0.02	23	0.97*	0.04
Rescue Hoist	Hover Altitude Deviations	20	5.84	4.37	20	9.08	9.79
	Altitude Deviations	19	0.78	3.12	19	0.26	1.14

Note. Means with a superscript differ significantly at $p < 0.05$.

Physical Domain. There was an effect of workload during the *deck landing task*, however, this was in the opposite direction than expected. *Airspeed deviations* ($t(18) = -19.65, p < 0.001$) were significantly lower in the *high workload* as compared to the *low workload* condition. This difference may be due to participants focusing on maintaining performance parameters more closely during the *high workload* condition when their system was not functioning properly, possibly due to compensatory effort (Hockey, 1997, 2011). Similar was found with *heading deviations* during this task ($t(18) = -10.66, p < 0.001$), with greater deviations during *low workload* compared to *high workload*. There was also an effect workload during the *non-precision approach task*. *Airspeed deviations* were greater during the *high workload* as compared to the *low workload* condition ($t(18) = 2.91, p = 0.009$).

Table 9. Performance Descriptive Statistics Physical Domain

Task	Metric	Low Workload			High Workload		
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Deck Landing and Boost Off	Heading Deviations	19	109.11*	17.13	19	28.49*	2.30
	Airspeed Deviations	19	51.87*	7.87	19	28.32*	5.29
Non-Precision Approach	Airspeed Deviations	19	8.58*	3.78	19	12.42*	4.40

Note. Means with a superscript differ significantly at $p < 0.05$.

Assessment of the Workload Manipulations: NASA-TLX Ratings. Using the overall workload rating from the NASA-TLX, all but two tasks, the *second terrain flight* ($p = 0.34$ [visual domain]) and the *non-precision approach* ($p = 0.45$ [physical domain]), had a significant effect of workload. The *t*-test statistics are reported in Table 10. The means and standard deviations for the tasks for each level of workload are provided in Table 11. The ratings were higher in all of the high-workload conditions than those for the low-workload conditions.

Table 10. NASA-TLX Results

Task	Workload Domain	df	<i>t</i>	<i>p</i>
Terrain Flight	Visual	22	-5.93	<0.001
Terrain Flight MEDEVAC	Visual	22	-0.98	0.34
Respond to IIMC	Cognitive	22	-2.14	0.04
Respond to Emergency	Cognitive	22	-4.03	0.001
Perform Radio Comms.	Auditory	22	-7.49	<0.001
Perform Rescue Hoist	Auditory	22	-4.29	<0.001
Deck Land & Boost Off	Physical	22	-9.21	<0.001
Non-precision approach	Physical	22	-0.77	0.45

Table 11. NASA-TLX Descriptive Statistics

Task	Workload Domain	Workload			
		Low Workload		High Workload	
		Mean	SD	Mean	SD
Terrain Flight	Visual	53.01*	17.07	68.58*	14.30
Terrain Flight MEDEVAC	Visual	32.07	15.71	36.56	16.48
Respond to IIMC	Cognitive	43.83*	16.57	53.70*	20.77
Respond to Emergency	Cognitive	29.72*	13.82	48.00*	18.92
Perform Radio Comms	Auditory	18.42*	11.23	41.54*	13.25
Perform Rescue Hoist	Auditory	47.30*	16.24	61.03*	16.75
Deck Land & Boost Off	Physical	47.17*	15.09	75.75*	13.32
Non-precision approach	Physical	40.25	17.81	43.52	16.34

Note. Means with an asterisk differ significantly at $p < 0.05$.

Assessment of the Workload Manipulations: Workload Profile Ratings. Using the overall summed rating across the subscales, only one task resulted in a significance difference in perceived workload on this measure. This was the *perform rescue hoist* ($t(22) = -1.83, p = 0.004$). Results of all of the tasks are in Table 12 and descriptive statistics are in Table 13. Based on reports from participants following completion of the WP rating, as well as the research team members administering the WP rating, it is likely participants did not fully understand the nature of the questionnaire regarding how they were rating the tasks completed. Therefore, the correlation analyses will use NASA-TLX data only.

Table 12. Workload Profile Results

Task	Workload Domain	df	<i>t</i>	<i>p</i>
Terrain Flight	Visual	22	-1.50	0.15
Terrain Flight MEDEVAC	Visual	21	-1.69	0.11
Respond to IIMC	Cognitive	22	-1.61	0.12
Respond to Emergency	Cognitive	22	-0.20	0.84
Perform Radio Comms.	Auditory	22	-3.20	0.004
Perform Rescue Hoist	Auditory	22	-1.83	0.08
Deck Land & Boost Off	Physical	22	-0.37	0.72
Non-precision approach	Physical	22	-1.14	0.27

Table 13. Workload Profile Descriptive Statistics

Task	Workload Domain	Workload			
		Low Workload		High Workload	
		Mean	SD	Mean	SD
Terrain Flight	Visual	46.98	12.21	50.22	9.68
Terrain Flight MEDEVAC	Visual	39.80	12.65	45.00	13.13
Respond to IIMC	Cognitive	46.23	14.28	50.08	9.79
Respond to Emergency	Cognitive	49.39	15.85	49.86	12.17
Perform Radio Comms	Auditory	44.09*	16.51	54.02*	13.55
Perform Rescue Hoist	Auditory	47.09	13.56	51.14	11.83
Deck Land & Boost Off	Physical	45.00	12.20	46.25	11.71
Non-precision approach	Physical	41.71	14.40	44.81	10.80

Note. Means with a superscript differ significantly at $p < 0.05$.

Assessment of the Workload Manipulations: Psychophysiological Recordings.

Three participants' ECG and respiration, and EOG data were excluded due to recording errors across all conditions. Within conditions participants' who had impossible values, based on normative HRV values presented by Shaffer and Ginsberg (2017) were removed. As these differed for each test condition n values are reported later in the descriptive statistics tables.

Visual Domain. There was no effect of workload on the psychophysiological measures during either flight task (see Table B1 in Appendix B for test statistics). There were, however, non-significant trends regarding the mean values for each of the outcome measures. Within both conditions, there were trends for higher *HRV* values in the low conditions than high conditions and lower *HRs* in the low conditions than high conditions. *RRs* during the *visual terrain flight* were lower in the low condition than high condition. These non-significant trends are in line with the expected direction for these outcome measures within each workload condition.

Table 14. Psychophysiological Descriptive Statistics Visual Domain

Task	Metric	n	Low Workload		High Workload		
			M	SD	n	M	SD
Terrain Flight	HR	19	71.85	8.71	19	72.30	9.80
	HRV	19	4.23	3.99	19	3.74	3.98
	RR	19	23.04	3.62	19	24.12	3.96
	EOG-Duration	17	0.13	0.03	19	0.12	0.03
	EOG-Blink Rate	16	5.71	2.40	16	5.36	2.25
	EEG-Frontal Alpha	23	1.58	0.10	23	1.59	0.09
	EEG-Frontal Beta	23	1.76	0.07	23	1.76	0.07
	EEG-Frontal Theta	23	1.37	0.11	23	1.37	0.11
	EEG-Beta Ratio	23	0.60	0.06	23	0.60	0.06
Terrain Flight	HR	17	71.55	9.41	16	74.72	11.37
MEDEVAC	HRV	17	4.42	3.27	16	2.96	2.28

RR	17	22.60	4.89	16	22.60	4.89
EOG-Duration	17	0.13	0.03	17	0.13	0.03
EOG-Blink Rate	13	4.93	1.82	13	5.59	1.86
EEG-Frontal Alpha	22	1.56	0.09	22	1.56	0.09
EEG-Frontal Beta	22	1.79	0.07	22	1.80	0.07
EEG-Frontal Theta	22	1.32	0.09	22	1.32	0.11
EEG-Beta Ratio	22	0.62	0.06	22	0.63	0.05

Note. Values reported here are the raw values collected, while baseline-corrected values were used for HR, HRV, and RR analyses.

Cognitive Domain. There was no effect of workload on either of the flight tasks on the *HR*, *HRV*, *RR* and *EOG* outcome measures, although there were non-significant trends, similar to the visual domain tasks. These non-significant trends were in the expected direction based on workload during the *respond to IIMC* task, but not the *respond to emergency* task for *HR*, *HRV*, and *RR* (see Table B2 in Appendix B for full test statistics).

There was a significant effect of workload amongst the *EEG* outcome measures. During the *respond to IIMC task* several of *EEG* metrics were sensitive to workload. *Frontal beta values* were significantly lower during the *high workload* compared to *low workload* condition ($t(22) = 3.69, p = 0.001$), which was in the opposite direction of what was expected for this metric. *Frontal theta values* were significantly elevated during the *high workload* compared to *low workload* condition ($t(22) = -4.25, p < 0.001$), as expected. *Beta ratio values* were significantly lower during *high workload* than *low workload* condition ($t(22) = 4.09, p < .001$). During the *respond to emergency task* only *frontal beta values* ($t(22) = -2.07, p = 0.05$) were sensitive to workload manipulations, although marginally significant, with higher values during the *high workload* compared to the *low workload* condition. See Table 15 for all descriptive statistics.

Table 15. Psychophysiological Descriptive Statistics Cognitive Domain

Task	Metric	Low Workload			High Workload		
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Respond to IIMC 1	HR	19	71.83	8.28	17	72.97	10.61
	HRV	19	3.77	2.79	17	3.53	2.88
	RR	19	21.26	4.11	17	21.28	3.05
	EOG-Duration	19	0.13	0.03	19	0.13	0.03
	EOG-Blink Rate	15	5.64	1.92	15	5.77	1.88
	EEG-Frontal Alpha	23	1.57	0.10	23	1.58	0.09
	EEG-Frontal Beta	23	1.79*	0.09	23	1.74*	0.05
	EEG-Frontal Theta	23	1.33*	0.13	23	1.41*	0.12
	EEG-Beta Ratio	23	0.62*	0.07	23	0.59*	0.05
Respond to Emergency 2	HR	19	71.70	8.91	19	71.59	9.34
	HRV	19	4.73	4.10	19	4.39	4.43
	RR	19	22.59	4.14	19	23.59	3.59
	EOG-Duration	18	0.12	0.03	18	0.13	0.03
	EOG-Blink Rate	14	5.51	2.20	14	5.21	1.52
	EEG-Frontal Alpha	23	1.56	0.09	23	1.55	0.09
	EEG-Frontal Beta	23	1.77	0.07	23	1.80	0.07

EEG-Frontal Theta	23	1.35	0.12	23	1.32	0.11
EEG-Beta Ratio	23	0.61	0.06	23	0.63	0.05

Note. Values reported here are the raw values collected, while baseline-corrected values were used for HR, HRV, and RR analyses. Means with an asterisk differ significantly at $p < 0.05$.

Auditory Domain. There was no effect of workload for the *HR*, *HRV*, and *RR* outcome measures during either flight task. There was effect of workload on the EEG outcome metrics during the *radio calls* task. *Frontal beta values* were significantly higher values during the *high workload* compared to *low workload* condition ($t(20) = -2.55, p = 0.02$). The *beta-ratio values* were also sensitive to workload ($t(20) = -2.58, p = 0.02$), with higher values during the *high workload* compared to *low workload* condition. None of the EEG metrics were sensitive to workload during the *rescue hoist* task. There was an effect of workload on *blink rates* during the *rescue hoist* task, with a higher *blink rate* during the *high workload* compared to *low workload* condition, which was opposite of the expected outcome ($t(13) = 2.39, p = 0.03$). Table 16 includes all means, and Table B3 in Appendix B includes all test statistics.

Table 16. Psychophysiological Descriptive Statistics Auditory Domain

Task	Metric	Low Workload			High Workload		
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Radio Communications	HR	19	69.85	9.80	19	70.63	9.30
	HRV	19	2.97	2.74	19	3.97	3.34
	RR	19	24.09	3.73	19	23.70	3.79
	EOG-Duration	18	0.13	0.03	18	0.12	0.02
	EOG-Blink Rate	17	6.68	3.92	17	6.48	2.52
	EEG-Frontal Alpha	21	1.57	0.10	21	1.55	0.10
	EEG-Frontal Beta	21	1.77*	0.08	21	1.80*	0.08
	EEG-Frontal Theta	21	1.35	0.11	21	1.31	0.12
	EEG-Beta Ratio	21	0.61*	0.06	21	0.63*	0.06
Rescue Hoist	HR	16	71.64	10.08	16	72.52	11.37
	HRV	16	4.69	4.34	16	3.68	3.57
	RR	16	21.98	3.63	16	22.08	1.77
	EOG-Duration	15	0.13	0.03	15	0.12	0.03
	EOG-Blink Rate	14	5.39*	2.13	14	6.51	2.38*
	EEG-Frontal Alpha	23	1.57	0.10	23	1.57	0.09
	EEG-Frontal Beta	23	1.79	0.07	23	1.79	0.07
	EEG-Frontal Theta	23	1.34	0.15	23	1.33	0.12
	EEG-Beta Ratio	23	0.62	0.06	23	0.62	0.06

Note. Values reported here are the raw values collected, while baseline-corrected values were used for HR, HRV, and RR analyses. Means with an asterisk differ significantly at $p < 0.05$.

Physical Domain. There was no effect of workload during either flight task for the *HR*, *HRV*, *RR*, and *EOG* outcome measures. Non-significant trends in the hypothesized direction can be noted for the *deck landing & boost off* task for *HRV* and *HR* mean values (see Table 17).

During the *deck landing* task, there was an effect workload on *frontal alpha* and *beta-ratio values*, with lower *alpha values* during the low workload condition than the high workload

condition whereas *beta-ratio values* were higher in the high compared to low workload condition ($t(22) = -4.31, p < 0.001$; $t(22) = 2.51, p = 0.02$, respectively). *Frontal theta values* were marginally significant during the *deck landing task*, with higher values during the *high workload* compared to *low workload* condition ($t(22) = -2.08, p = 0.05$). A similar pattern was found during the *non-precision approach task* for *beta-ratio values*, although they were marginally significant ($t(22) = -2.13, p = 0.05$). *Frontal beta values* were sensitive to workload as well, with lower values during *low workload* compared to *high workload* ($t(22) = -2.78, p = 0.01$).

Table 17. Psychophysiological Descriptive Statistics Physical Domain

Task	Metric	Low Workload			High Workload		
		<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>
Deck Landing and Boost Off Flight	HR	18	72.64	7.91	19	76.01	9.17
	HRV	18	5.00	5.62	19	4.45	4.32
	RR	19	22.78	3.64	19	23.51	3.76
	EOG-Duration	17	0.13	0.03	17	0.13	0.03
	EOG-Blink Rate	15	5.78	2.66	15	6.61	4.19
	EEG-Frontal Alpha	23	1.57*	0.10	23	1.61*	0.10
	EEG-Frontal Beta	23	1.77	0.07	23	1.75	0.08
	EEG-Frontal Theta	23	1.34*	0.10	23	1.41*	0.15
	EEG-Beta Ratio	23	0.61*	0.51	23	0.58*	0.07
Perform Non-Precision Approach	HR	16	73.21	10.20	17	74.93	13.72
	HRV	16	3.82	2.76	17	3.98	3.66
	RR	16	20.65	3.94	17	21.98	3.85
	EOG-Duration	17	0.12	0.03	17	0.12	0.03
	EOG-Blink Rate	15	6.54	3.36	15	5.90	1.80
	EEG-Frontal Alpha	23	1.56	0.10	23	1.54	0.09
	EEG-Frontal Beta	23	1.79*	0.08	23	1.81*	0.09
	EEG-Frontal Theta	23	1.32	0.11	23	1.30	0.11
	EEG-Beta Ratio	23	0.62*	0.06	23	0.64*	0.06

Note. Values reported here are the raw values collected, while baseline-corrected values were used for HR, HRV, and RR analyses. Means with a superscript differ significantly at $p < 0.05$.

Correlations between performance, psychophysiological recordings, and perceived workload. Correlation analyses were completed to examine the relationships between each of the variables. Only overall TLX weighted ratings are reported here.

Visual Domain.

Performance and TLX. There were no significant correlations of performance measures and TLX ratings during *visual terrain low*. Significant correlations were found for *visual terrain high* between *airspeed deviations* and *overall TLX ratings* ($r(22) = 0.50, p = 0.02$), such that *airspeed deviations* increased with higher workload ratings. There were no significant correlations between performance measures and TLX ratings during *terrain MEDEVAC low* or *high workload*.

Psychophysiological and TLX. No significant relationships were found between TLX

ratings and psychophysiological metrics during either the *low* or *high workload* conditions for the *visual terrain* task. During the *terrain MEDEVAC low*, *HR* was negatively correlated with *TLX overall rating* ($r(18) = -0.50, p = 0.04$), such that *HR* decreased as workload rating increased. None of the remaining psychophysiological variables correlated with *TLX ratings* during the *low workload* condition. During the *high* condition *TLX ratings* correlated with *beta-ratio* ($r(22) = -0.46, p = 0.03$), such that *beta-ratio values* decreased as workload ratings increased.

Performance and Psychophysiological. None of the psychophysiological measures correlated with performance during *visual terrain low* or *high workload* conditions. During the *terrain MEDEVAC* there was a significant correlation only for the *high workload* condition between *airspeed* and *HRV* ($r(14) = -0.62, p = 0.02$) such that *HRV* values decreased as *airspeed* increased.

Cognitive Domain.

Performance and TLX. None of the performance measures during the *respond to IIMC* nor *respond to emergency* correlated with *TLX ratings*.

Psychophysiological and TLX. During *respond to IIMC low workload* condition *TLX ratings* correlated with both *HRV* ($r(20) = 0.46, p = 0.04$) and *HR* ($r(20) = -0.53, p = 0.02$), such that *HRV* increased with higher workload ratings whereas *HR* decreased. No correlations were found during the *high workload* condition. During *respond to emergency* task there were no significant correlations for either condition.

Performance and Psychophysiological. There were no significant correlations between performance measures and physiological measures during *respond to IIMC* in the *low workload* condition. During the *high workload* condition, there was a significant correlation between *airspeed* and *RR*, $r(18) = -0.59, p = 0.01$, such that as *airspeed* deviations increased, *RR* decreased. There were no significant correlations for the *respond to emergency* task during either condition.

Auditory Domain.

Performance and TLX. There were no significant correlations for either condition between performance measures and *TLX ratings* for the *radio communications* task. There were also no significant correlations between the *TLX ratings* and performance measures for the *rescue hoist* task.

Psychophysiological and TLX. There were no significant correlations during *radio communications* task *low workload* nor *high workload* condition. In the *rescue hoist* task there were also no significant correlations for either condition.

Performance and Psychophysiological. During the *radio communications high workload* condition, *HR* and *airspeed deviations* were significantly correlated ($r(19) = -0.48, p = 0.04$), where *HR* decreased with increased *airspeed deviations*. There were no significant correlations with performance metrics during the *rescue hoist low workload* condition. There was a

significant correlation between *frontal alpha activity* and *hover altitude deviations* ($r(22) = 0.44$, $p = 0.04$) during the *high workload* condition, such that *frontal alpha activity* increased as *altitude deviations* increased.

Physical Domain.

Performance and TLX. During the *deck landing* task *low workload*, *airspeed deviations* correlated with *TLX overall rating* ($r(20) = 0.56$, $p = 0.01$), such that there were greater *airspeed deviations* with higher workload ratings. During the *high workload* condition there were no significant correlations. There is no performance data for the *non-precision approach*, *high* or *low* conditions.

Psychophysiological and TLX. There were no significant correlations for either condition during the *deck landing* task. During the *non-precision approach* task *low workload* condition, *HR* correlated with *TLX ratings* ($r(16) = 0.75$, $p = .001$), where *HR* increased as workload ratings increased. None of the measures correlated during the *high workload* condition.

Performance and Psychophysiological. During the *deck landing*, there was a significant correlation between *blink rate* and *heading deviations* ($r(15) = -0.61$, $p = 0.02$), such that *blink rate* decreased as *heading deviations* increased. None of the physiological measures correlated with the performance measures during the *deck landing* task *high* condition. During the *non-precision approach* task *low workload* there was a significant correlation between *frontal alpha values* and *airspeed deviations* ($r(20) = 0.67$, $p = 0.001$), where *frontal alpha values* increased with increased *airspeed deviations*. None of the psychophysiological metrics correlated with the *high workload* condition.

Correlations between psychomotor vigilance task performance and psychophysiological recordings

PVT outcome variables included *reaction time (RT)* in milliseconds and number of *lapses*. Four participants' data *HRV* data were excluded as outliers; one from *respond to emergency high* (cognitive domain); one from *perform rescue hoist high* (auditory domain); one from both *deck land and boost off high* and *non-precision approach high*; and one from *non-precision approach low* (physical domain).

Significant correlations between psychophysiological metrics and PVT outcome metrics were found for six of the 16 flight tasks. Two ECG metrics correlated with PVT metrics. *HRV* correlated with *RT* ($r(17) = 0.50$, $p = 0.04$) following the *terrain flight MEDEVAC high workload* condition (visual domain). *HR* correlated with *lapses* ($r(23) = 0.46$, $p = 0.05$) such that *HR* increased as *lapses* increased, following the *deck land and boost off low workload* condition (physical domain). *Blink rate* correlated with *RT* ($r(12) = -0.60$, $p = 0.04$), such that *blinks rates* decreased as *RTs* increased, following the *terrain flight MEDEVAC high workload* (visual domain). *Frontal alpha* correlated with *PVT metrics* following three flight tasks: with *RT* ($r(23) = 0.46$, $p = .03$) following *IIMC low workload condition* (cognitive domain), with *RT* ($r(23) = 0.46$, $p = 0.03$) following *perform radio communications high* (auditory domain), and with *lapses* ($r(23) = 0.45$, $p = 0.03$) following *perform rescue hoist low* (auditory domain). For each of the *frontal alpha values* correlations, *alpha values* increased as *RTs* or *lapses* increased.

Study 2 Discussion

Workload manipulations were successful for five of the eight flight tasks, as indicated by differences in performance metrics. These included one task from each domain (visual: *terrain flight*; cognitive: *IIMC*; auditory: *radio calls*) as well as both tasks in the physical domain (*deck landing* and *non-precision approach*). This was further supported with participative workload reported with the NASA-TLX, where there was an effect of workload for six of the eight flight tasks. The two tasks that did have an effect of workload were the *MEDEVAC flight* (visual domain) and the *non-precision approach* (physical domain). Several of psychophysiological metrics were also sensitive to workload during the tasks, with the EEG metrics as the most frequently identified.

Study 2, similar to Study 1, also suggested differential patterns in psychophysiological recordings and associations with workload levels by workload domain. Here, *HR* was associated with each of the domains, but *HRV* was only associated with the visual and cognitive domains. *HR* relationships were found with TLX ratings in the visual domain, cognitive domain, and physical domain, though the directions differed (negative correlations in visual and cognitive domains; positive correlation in physical domain). *HRV* relationships were with performance metrics within the physical domain and TLX ratings in the cognitive domain, where correlations with performance were negative but positive for TLX ratings. Relationships between RR and performance metrics were found within the cognitive domain, although the relationship was negative, where RR decreased as performance metrics worsened. Several EEG metrics were found to have relationships as well. These were the beta-ratio values and TLX ratings within the visual domain (negative correlation), and frontal alpha values with performance metrics within the auditory domain and the physical domain (both positive correlations). Finally, blink rate was associated with performance metrics in the physical domain, where the correlation was negative.

Finally, examination of relationships between the psychophysiological metrics and the PVT outcome metrics suggests some evidence of divergent validity. Significant correlations occurred in six of the 16 PVT administrations, for which *frontal alpha values* were positively with outcome metrics in three instances, which is not unexpected given the PVT is a vigilance task and much of the literature surrounding *frontal alpha activity* is related to attention. The remaining three had positive correlation with *HRV*, a positive correlation with *HR*, and a negative correlation with *blink rate*. It is possible that some of these are due to carry-over effects from the flight scenario that was completed prior to the completion of the PVT.

General Discussion

This study sought to evaluate whether differences in workload can be measured within individual domains using the multiple resource model as a guide. We found several consistencies in domain-specific relationships of psychophysiological measures between Study 1 and Study 2. Specifically, *HR* and *RR* were significantly related to metrics within the cognitive domain for both studies. Frontal alpha activity was associated with the auditory domain in both studies, and in a positive direction for both. There were significant relationships with EEG metrics for both studies within the physical domain; however, these occurred within different bands. During the first study, theta values were related to perceived workload, whereas during the second study, alpha values were related to performance. Finally, beta-ratio values were associated with TLX

ratings within the visual domain in Study 2, while only alpha values were associated in Study 1. However, given that the beta-ratio takes alpha values into account, it is worth noting.

Much of what was found regarding psychophysiological indices in response to the task manipulations were in-line with what is currently known within the literature. We found *HR* associated with metrics within the cognitive domain, similar to Fallahi et al. (2016) and Hankins and Wilson (1998), who found increased HR related to mental calculations. The relationship of EEG metrics within the physical domain has been supported in prior work, although the band activity and cortical regions examined differ from what was found here (for a review, see Mehta & Parasuraman, 2013).

Our findings of an increase in alpha activity within the auditory domain in response to increased workload is similar to what has been found in previous auditory studies (e.g., Parvaz, MacNamara, Goldstein, & Hajcak, 2012). Strauß, Wöstmann, & Oblesser (2014) discussed the role of alpha oscillations in auditory inhibition and suggested a framework for the role of alpha activity in auditory tasks where they postulated alpha activity increases to aid in the processing of relevant auditory information, by reducing the interference of non-relevant information. Our findings that alpha activity was associated with flight performance metrics during the radio communications task, but not during the rescue hoist task suggests support for Strauß and colleagues' hypotheses regarding the role of alpha bands during these types of auditory tasks. Specifically, during the radio communications task, participants were required to listen for their unique call-sign amongst distractor call-signs, thus requiring them to essentially "tune out" task irrelevant information. Whereas the auditory listening requirements of the rescue hoist task were simply to listen for directional information, without distractors to ignore. The findings during the radio communications tasks line well with the auditory oddball task used in Study 1, where participants were also presented with distractor auditory information. This suggests that not only is there a domain specific difference, but there may be also be a task specificity, as suggested by Matthews et al. (2015). Alpha activity associations within the visual domain during Study 1 may serve a similar function as the increase in alpha activity within the auditory domain tasks. Specifically, the increase of alpha activity as a function of increased workload suggests that alpha activity, measured from the frontal regions played an inhibitory role by inhibiting some of the activity within the frontal region, as this region is less relevant to the processing basic visual information (Jensen & Mazaheri, 2010). As the same was not found within the visual domain tasks for the second study, it suggests that the role of this activity is more pronounced when visual processing is isolated. However, it may be worthwhile to examine alpha activity within other regions, such as the parietal or midline regions, to evaluate whether alpha bands have a greater inhibitory role in those locations.

Multiple resource theory can help frame the above findings. According to the multiple resource model, there are several, qualitatively separate resource pools to processes different sources of task demands. Moreover, Wickens (2008) asserted that the multiple resource model should have "neurophysiological plausibility," meaning that distinct cortical networks underlie qualitatively different resource pools. Here, we showed that different physiological metrics are more sensitive to different domains of task demands, generally supporting the multiple resource framework. If there was one pool of undifferentiated resources, it would be expected that all physiological metrics would show the same patterns across different workload domain

manipulations. However, the results here support a multiple resource account of workload. That is, physiological measures showed different relationships with performance and self-report workload across domains. One avenue for future research is to use neuroimaging techniques to spatially map changes in cortical activity during the performance of tasks across domains with concurrent manipulations of task demands. Just and Varma (2007) asserted that distributed cortical networks change their activity to accommodate increased resource demands. Moreover, these networks are different depending on the qualitative structure of the task (i.e., sentence comprehension vs. spatial processing). Thus, portable neuroimaging techniques such as functional near-infrared spectroscopy or high density EEG may be useful for mapping functional workload domains at the level of cortical network activation.

Conclusions

The results of this study lay a foundation for future work examining the use of psychophysiological indices to identify workload in real-time. More specifically, we found indications of domain-specificity within some of the measures employed. This suggests that multiple measures may be able to better diagnose changes in workload, as well as determine which domain is the source of workload. While further work is needed to establish precision of these measures, findings reported here provide an initial indication of which measures are most promising. Furthermore, we found evidence that cortical activity as measured by EEG may be more influenced by the domain in which the task occurs versus the degree of workload presented. Traditionally, the use of EEG metrics in workload research has focused on attention-driven hypotheses. That is, much of the literature postulates changes in bands based on what is known for attention paradigms, such as vigilance tasks (for a review of EEG indices, see Kamzanova, Matthews, & Kustubayeva, 2011). However, here we noted activity in the different bands to occur in a domain-specific pattern that differed from what is known with the same bands and attention theories. As an example, we found alpha activity to increase with workload within the auditory domain. Traditional work within attention suggests that there should rather be an alpha suppression (e.g., Slobounov, 2000; Fairclough et al., 2005) since it is expected that increases in workload requires more attention/cognitive processing from the participant. However, we demonstrated here that is likely less clear-cut. Therefore, here we suggest that rather than identifying a one-size-fits-all approach to monitoring and identifying workload in real-time, the focus should be on identifying which sensors are able to reliably identify domain-specific workload changes.

References

- Advanced Brain Monitoring, Inc. (2009). *B-Alert Live user manual*. Carlsbad, CA.
- Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, 36(7), 623-636.
- Allan, K., White, T., Jones, L., Merlo, J., Haas, E., Zets, G., & Rupert, A. (2010, September). Getting the buzz: What's next for tactile information delivery? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 54(18), 1331-1334.
- Army Vision - Force 2025 White Paper. (2014, January). Retrieved from http://www.arcic.army.mil/app_Documents/USArmy_WhitePaper_Army-Vision-Force-2025_23JAN2014.pdf
- Bauer, L. O., Strock, B. D., Goldstein, R., Stern, J. A., & Walrath, L. C. (1985). Auditory discrimination and the eyeblink. *Psychophysiology*, 22(6), 636-641.
- Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio*, 78(2), 490-8.
- Berka, C., Levendowski, D. J., Lumicao, M. N., Yau, A., Davis, G., Zivkovic, V. T., . . . & Craven, P. L. (2007). EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. *Aviation, Space, and Environmental Medicine*, 78(5), B231-B244.
- Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., & Babiloni, F. (2014). Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. *Neuroscience and Biobehavioral Reviews*, 44, 58-75.
- Caffier, P. P., Erdmann, U., & Ullsperger, P. (2003). Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, 89(3-4), 319-25.
- Cassenti, D. N., & Kelley, T. D. (2006). Towards the shape of mental workload. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting* (pp. 1147-1152). Santa Monica, CA: Human Factors and Ergonomics Society.
- Cassenti, D. N., & Kelley, T. D. (2008). Towards a workload-performance prediction tool. *Proceedings of the 30th Annual Meeting of Cognitive Science Society*, New York.
- Colle, H. A. (2010). Mental workload manipulation using multiple homogeneous tasks: Performance effects. Retrieved from: <https://apps.dtic.mil/dtic/tr/fulltext/u2/a530185.pdf>
- Collet, C., Averty, P., & Dittmar, A. (2009). Autonomic nervous system and subjective ratings of strain in air-traffic control. *Applied Ergonomics*, 40(1), 23-32.
- Collet, C., Clarion, A., Morel, M., Chapon, A., & Petit, C. (2009). Physiological and behavioural

- changes associated to the management of secondary tasks while driving. *Applied Ergonomics*, 40(6), 1041-1046.
- Comstock, J. R., Jr., & Arnegard, R. J. (1992). The Multi-Attribute Task Battery for human operator workload and strategic behavior research (NASA TM-104174). Hampton, Virginia: NASA Langley Research Center.
- DiDomenico, A., & Nussbaum, M. A. (2008). Interactive effects of physical and mental workload on subjective workload assessment. *International Journal of Industrial Ergonomics*, 38(11), 977-983.
- Dinges, D. F., & Powell, J. W. (1985). Microcomputer analyses of performance on a portable, simple visual RT task during sustained operations. *Behavior Research Methods, Instruments and Computers*, 17, 652-655.
- Doppelmayr, M., Finkenzeller, T., & Sauseng, P. (2008). Frontal midline theta in the pre-shot phase of rifle shooting: Differences between experts and novices. *Neuropsychologia*, 46(5), 1463-1467.
- Durant, G., Gagnon, J. F., Tremblay, S., & Dehais, F. (2014). Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural Brain Research*, 259, 16–23. doi:10.1016/j.bbr.2013.10.042
- Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Epling, S. L., Middendorf, M., Hoepf, M., Gruenwald, C., Stork, L., & Galster, S. (2015). *The electrooculogram and a new blink detection algorithm* (Report No. AFRL-RH-WP-TP-2015-0025). Wright Patterson AFB, OH: Air Force Materiel Command.
- Ercan, I., Hafizoglu, S., Ozkaya, G., & Kirli, S. (2015). Examining cut-off values for the state-trait anxiety inventory. *Revista Argentina de Clinica Psicologica*, 24(2), 143-148.
- Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., & Miyake, S. (2016). Effects of mental workload on physiological and subjective responses during traffic density monitoring: A field study. *Applied Ergonomics*, 52, 95-103.
- Fairclough, S. H., Venables, L., & Tattersall, A. (2005). The influence of task demand and learning on the psychophysiological response. *International Journal of Psychophysiology*, 56(2), 171-18
- Freeman, F. G., Mikulka, P. J., Prinzel, L. J., & Scerbo, M. W. (1999). Evaluation of an adaptive automation system using three EEG indices with a visual tracking task. *Biological Psychology*, 50, 61-76.
- Gilbert, D. G., Dibb, W. D., Plath, L. C., & Hiyane, S. G. (2000). Effects of nicotine and

- caffeine, separately and in combination, on EEG topography, mood, heart rate, cortisol, and vigilance. *Psychophysiology*, 37, 583-595.
- Gonzalez, C. (2005). Task workload and cognitive abilities in dynamic decision making. *Human Factors*, 47(1), 92-101.
- Haas, E., Shankle, R., Murray, H., Travers, D., & Wheeler, T. (2000, July). Issues relating to automatic speech recognition and spatial auditory displays in high noise, stressful tank environments. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 44, No. 22, pp. 754-757). SAGE Publications.
- Hankins, T. C., & Wilson, G. F. (1998). A comparison of heart rate, eye activity, EEG and subjective measures of pilot mental workload during flight. *Aviation, Space, and Environmental Medicine*, 69(4), 360-367.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, 52, 139-183.
- Hockey, G. R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, 45, 73-93.
- Hockey, G. R. J. (2011). A motivational control theory of cognitive fatigue. In P. L. Ackerman (Ed.), *Cognitive fatigue: multidisciplinary perspectives on current research and future applications* (pp. 167-188). Washington, DC: American Psychological Association.
- Hockey, G. R. J., Nickel, P., Roberts, A. C., & Roberts, M. H. (2009). Sensitivity of candidate markers of psychophysiological strain to cyclical changes in manual control load during simulated process control. *Applied Ergonomics*, 40, 1011-1018.
- Horat, S. K., Herrmann, F. R., Favre, G., Terzis, J., Debatisse, D., Merlo, M. C., & Missonnier, P. (2016). Assessment of mental workload: A new electrophysiological method based on intra-block averaging of ERP amplitudes. *Neuropsychologia*, 82, 11-17.
- Hwang, T., Kim, M., Hwangbo, M., & Oh, E. (2014, August). Comparative analysis of cognitive tasks for modeling mental workload with electroencephalogram. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE* (pp. 2661-2665). IEEE.
- Jensen, O., & Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: Gating by inhibition. *Frontiers in Human Neuroscience*, 4(186), 1-8.
- Just, M. A., & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, and Behavioral Neuroscience*, 7(3), 153-191.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.

- Karavidas, M. K., Lehrer, P. M., Lu, S. E., Vaschillo, E., Vaschillo, B., & Cheng, A. (2010). The effects of workload on respiratory variables in simulated flight: a preliminary study. *Biological psychology*, *84*(1), 157-160.
- Kähkönen, S., Wilenius, J., Nikulin, V. V., Ollikainen, M., & Ilmoniemi, R. J. (2003). Alcohol reduces prefrontal cortical excitability in humans: A combined TMS and EEG study. *Neuropharmacology*, *28*, 747-754.
- Kamzanova, A. T., Matthews, G., Kustubayeva, A. M., & Jakupov, S. M. (2011). EEG indices to time-on-task effects and to a workload manipulation (cueing). *International Journal of Social, Behavioral, Educational, Economic, Business, and Industrial Engineering*, *5*(8), 928-931.
- Khitrov, M. Y., Laxminarayan, S., Thorsley, D., Ramakrishnan, S., Rajaraman, S., Wesensten, N. J., & Reifman, J. (2014). PC-PVT: A platform for psychomotor vigilance task testing, analysis, and prediction. *Behavioral Research*, *46*, 140-147.
- Kramer, A. F., & Weber, T. (2000). Applications of psychophysiology to human factors. In J. T. Cacioppo & G. G. Bemston (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 794–814). New York, NY: Cambridge University Press.
- Lee, Y. H., & Liu, B. S. (2003). Inflight workload assessment: Comparison of subjective and physiological measurements. *Aviation, Space and Environmental Medicine*, *74*(10), 1078-1084.
- Lenneman, J. K., & Backs, R. W. (2009). Cardiac autonomic control during simulated driving with a concurrent verbal working memory task. *Human Factors*, *51*(3), 404-418.
- Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology*, *121*, 62–73.
- Mansikka, H., Virtanen, K., Harris, D., & Simola, P. (2016). Fighter pilots' heart rate, heart rate variation and performance during an instrument flight rules proficiency test. *Applied Ergonomics*, *56*, 213-219.
- Marquart, G., & de Winter, J. (2015). Workload assessment for mental arithmetic tasks using the task-evoked pupillary response. *PeerJ Computer Science*, *1*, e16.
- Marquart, G., Cabrall, C., & de Winter, J. (2015). Review of eye-related measures of drivers' mental workload. *Procedia manufacturing*, *3*, 2854–2861.
- Matthews, T. D., Orzech, J. A., & Lassiter, K. S. (2011). Does the Shipley Institute of Living Scale measure fluid and crystallized abilities?. *North American Journal of Psychology*, *13*(2).
- Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich IV, J. (2015). The psychometrics

- of mental workload: Multiple measures are sensitive but divergent. *Human Factors*, 57(1), 125-143.
- Mehler, B., Reimer, B., Coughlin, J., & Dusek, J. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record: Journal of the Transportation Research Board*, 2138, 6-12.
- Mehta, R. K., & Parasuraman, R. (2013). Neuroergonomics: A review of applications to physical and cognitive work. *Frontiers in Human Neuroscience*, 23. doi: 10.3389/fnhum.2013.00889
- Morrison, N., Burnham, D., & Morrison, B. W. (2015). Cognitive Load in Cross-Modal Dual-Task Processing. *Applied Cognitive Psychology*, 29(3), 436-444.
- Okogbaa, O.G., Shell, R. L., & Filipusi, D. (1994). On the investigation of the neurophysiological correlates of knowledge worker mental fatigue using the EEG signal. *Applied Ergonomics*, 25(6), 355-365.
- Pan, J., & Tompkins, W. J. (1985). A real-time QRS detection algorithm. *IEEE Transactions on Biomedical Engineering*, 32(3), 230-236.
- Parasuraman, R. (2015). Neuroergonomic perspectives on human systems integration: Mental workload, vigilance, adaptive automation, and training. In D. A. Boehm-Davis, F. T. Durso, & J. D. Lee. (Eds.), *APA handbook of human systems integration* (pp. 163–176). Washington, DC: American Psychological Association.
- Parvaz, M. A., MacNamara, A., Goldstein, R. Z., & Hajcak, G. (2012). Event-related induced frontal alpha as a marker of lateral prefrontal cortex activation during cognitive reappraisal. *Cognitive Affective Behavioral Neuroscience*, 12, 730-740.
- Phillips, R. R., & Madhavan, P. (2011). The effect of distractor modality and processing code on human-automation interaction. *Cognition, Technology & Work*, 13(4), 233-244.
- Raisbeck, L. D., & Diekfuss, J. A. (2015). Fine and gross motor skills: The effects on skill-focused dual-tasks. *Human movement science*, 43, 146-154.
- Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: A field study and simulation validation. *Ergonomics*, 54(10), 932-942.
- Reimer, B., Mehler, B., Coughlin, J. F., Roy, N., & Dusek, J. A. (2011). The impact of a naturalistic hands-free cellular phone task on heart rate and simulated driving performance in two age groups. *Transportation Research Part F: Traffic Psychology and Behaviour*, 14(1), 13-25.
- Ryu, K., & Myung, R. (2005). Evaluation of mental workload with a combined measure based

- on physiological indices during a dual task of tracking and mental arithmetic. *International Journal of Industrial Ergonomics*, 35(11), 991-1009.
- Santiago-Espada, Y. (2014). MATB-II: Revised multi-attribute task battery. Retrieved from <http://matb.larc.nasa.gov/>
- Santiago-Espada, Y., Myer, R. R., Latorella, K. A., & Comstock, J. R. (2011). The multi-attribute task battery II (MATB-II) software for human performance and workload research: A user's guide (NASA/TM-2011-217164). *Hampton, Virginia: NASA Langley Research Center*.
- Siegenthaler, E., Costela, F. M., McCamy, M. B., Di Stasi, L. L., Otero-Millan, J., Sonderegger, A., & Martinez-Conde, S. (2014). Task difficulty in mental arithmetic affects microsaccadic rates and magnitudes. *European Journal of Neuroscience*, 39(2), 287-294.
- Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, 5(258). doi: 10.3389/fpubh.2017.00258
- Slobounov, S. M., Fukada, K., Simon, R., Rearick, M., & Ray, W. (2000). Neuropsychological and behavioral indices of time pressure effects on visuomotor task performance. *Brain Research and Cognitive Research*, 9(3), 287-298.
- Spielberger, C. D., & Gorsuch, R. L. (1983). *State-trait anxiety inventory for adults: Manual, instrument, and scoring guide*. Mind Garden, Incorporated.
- Strauß, A., Wöstmann, M., & Oblesser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, 8(350), 1-7.
- Stuiver, A., Brookhuis, K. A., de Waard, D., & Mulder, B. (2014). Short-term cardiovascular measures for driver support: increasing sensitivity for detecting changes in mental workload. *International Journal of Psychophysiology*, 92(1), 35-41.
- Tadjdeh, Y. (2015). Future Vertical Lift Could Be Shot in the Arm for Industry. <http://www.nationaldefensemagazine.org/archive/2015/October/Pages/FutureVerticalLiftCouldBeShotintheArmforIndustry.aspx>
- Taylor, G. S., Reinerman-Jones, L. E., Szalma, J. L., Mouloua, M., & Hancock, P. A. (2013). What to automate addressing the multidimensionality of cognitive resources through system design. *Journal of Cognitive Engineering and Decision Making*, 7(4), 311-329.
- Thayer, J. F., Hansen, A. L., Saus-Rose, E., & Johnsen, B. H. (2009). Heart rate variability, prefrontal neural function, and cognitive performance: The neurovisceral integration perspective on self-regulation, adaptation, and health. *Annals of Behavioral Medicine*, 37(2), 141-153. doi:10.1007/s12160-009-9101
- Trejo, L. J., Kramer, A. F., & Arnold, J. A. (1995). Event-related potentials as indices of display-monitoring performance. *Biological Psychology*, 40, 33-71.

- Tsang, P. S., & Velazquez, V. L. (1996). Diagnosticity and multidimensional subjective workload ratings. *Ergonomics*, *39*(3), 358-381.
- Veltman, J. A., & Gaillard, A. W. K. (1998). Physiological workload reactions to increasing levels of task difficulty. *Ergonomics*, *41*(5), 656-669.
- Webb, C. M., Gaydos, S. J., Estrada, A., & Milam, L.S. (2010). *Toward an operational definition of workload: A workload assessment of aviation maneuvers*. (USAARL Report No. 2010-15) Fort Rucker, AL: U.S. Army Aeromedical Research Laboratory.
- Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), *Attention and performance VIII* (pp. 239–257). Hillsdale, NJ: Erlbaum.
- Wickens, C. D. (1984). Processing resource in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–101). New York, NY: Academic Press.
- Wickens, C. D. (2002). Situation awareness and workload in aviation. *Current directions in psychological science*, *11*(4), 128-133.
- Wickens, C. D., & Hollands, J. (1999). *Engineering psychology and human performance*. Upper Saddle River, NJ: Prentice Hall.
- Wickens, C. D., Sandry, D., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, output, and central processing. *Human Factors*, *25*, 227–248.
- Wickens, C. D. (2008). Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449-455.
- Wickens, C. D., & Tsang, P. (2015). Workload. In D. A. Boehm-Davis, F. T. Durso, & J. D. Lee (Eds.), *APA handbook of human systems integration* (pp. 277–292). Washington, DC: American Psychological Association.
- Wilson G. F., Fullenkamp P., & Davis I. (1994). Evoked potential, cardiac, blink, and respiration measures of pilot workload in air-to-ground missions. *Aviation, Space, and Environmental Medicine*, *65*, 100–105.
- Yao, Y. J., Chang, Y. M., Xie, X. P., Cao, X. S., Sun, X. Q., & Wu, Y. H. (2008). Heart rate and respiration responses to real traffic pattern flight. *Applied psychophysiology and biofeedback*, *33*(4), 203-209.

Appendix A. Map of Flight Path



Figure A1. Example of Flight Path Response to Inadvertent Meteorological Conditions Flight Scenario.



Figure A2. Example of Flight Path for Perform Non-Precision Approach Flight Scenario

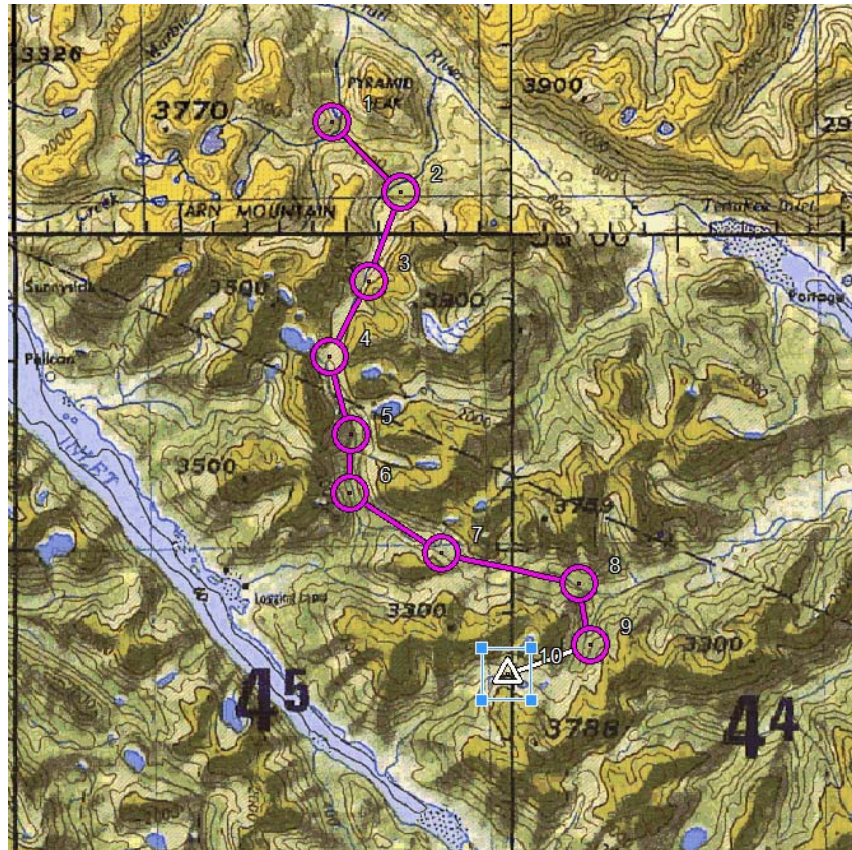


Figure A3. Example of Flight Path for Terrain Flight MEDEVAC Flight Scenario

Appendix B. t-Tests For All Domains

Table B1. t-Test Statistics for Visual Domain: Psychophysiological Data

Task	Metric	df	<i>t</i>	<i>p</i>
Terrain Flight (V1)	HR	18	0.41	0.69
	HRV	18	-0.60	0.56
	RR	18	1.52	0.15
	EOG-Duration	16	0.89	0.39
	EOG-Blink Rate	15	-0.38	0.71
	EEG-Frontal Alpha	22	-0.95	0.35
	EEG-Frontal Beta	22	0.28	0.79
	EEG-Frontal Theta	22	0.21	0.84
	EEG-Beta Ratio	22	0.27	0.79
Terrain Flight MEDEVAC	HR	15	-1.28	0.22
	HRV	15	1.76	0.10
	RR	15	0.95	0.36
	EOG-Duration	15	0.76	0.46
	EOG-Blink Rate	12	1.63	0.13
	EEG-Frontal Alpha	21	-0.18	0.86
	EEG-Frontal Beta	21	-1.43	0.17
	EEG-Frontal Theta	21	0.27	0.79
EEG-Beta Ratio	21	-0.69	0.50	

Table B2. *t*-Test Statistics for Cognitive Domain: Psychophysiological Data

Task	Metric	df	<i>t</i>	<i>p</i>
Respond to Inadvertent Instrument Meteorological Conditions (IIMC)	HR	18	-0.15	0.88
	HRV	18	-0.59	0.56
	RR	18	1.46	0.16
	EOG-Duration	15	1.03	0.32
	EOG-Blink Rate	13	-0.80	0.44
	EEG-Frontal Alpha	22	-0.64	0.53
	EEG-Frontal Beta	22	3.69	0.001
	EEG-Frontal Theta	22	-4.25	<0.001
	EEG-Beta Ratio	22	4.09	<0.001
Respond to Emergency	HR	16	-0.81	0.43
	HRV	16	0.44	0.67
	RR	16	-0.23	0.83
	EOG-Duration	18	0.04	0.97
	EOG-Blink Rate	14	0.27	0.79
	EEG-Frontal Alpha	22	1.33	0.20
	EEG-Frontal Beta	22	-2.07	0.05
	EEG-Frontal Theta	22	1.03	0.31
	EEG-Beta Ratio	22	-1.61	0.12

Table B3. *t*-Test Statistics for Auditory Domain: Psychophysiological Data

Task	Metric	df	<i>t</i>	<i>p</i>
Perform Radio Communications	HR	18	-0.79	0.44
	HRV	18	1.63	0.12
	RR	18	-0.59	0.57
	EOG-Duration	17	-0.85	0.41
	EOG-Blink Rate	16	-0.29	0.78
	EEG-Frontal Alpha	20	2.01	0.06
	EEG-Frontal Beta	20	-2.55	0.02
	EEG-Frontal Theta	20	1.86	0.08
	EEG-Beta Ratio	20	-2.58	0.02*
Perform Rescue Hoist	HR	15	-1.34	0.20
	HRV	15	0.79	0.44
	RR	15	-0.26	0.80
	EOG-Duration	14	-1.90	0.08
	EOG-Blink Rate	13	2.39	0.03
	EEG-Frontal Alpha	22	0.29	0.78
	EEG-Frontal Beta	22	-0.55	0.59
	EEG-Frontal Theta	22	0.33	0.74
EEG-Beta Ratio	22	-0.37	0.72	



Department of the Army
U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama 36362-0577
www.usaarl.army.mil



U.S. Army Medical Research and Materiel Command