USAARL-TECH-TR--2026-10

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

Physiological Monitoring Under Varied Conditions of Workload and Fatigue

**Amanda M. Kelley, Jon Vogl, Michelle Duffy, Kathryn A. Feltman,
J. Andrew Atchley, Isaiah Persson, Chris Aura, Xiaomin Yue,
Danielle McDermott, Jonathan “J” Melton, Samantha Wolf,
Jason Gerstner, & Ian Curry**

DISTRIBUTION STATEMENT A. Approved for public release: distribution is 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.

Human Subjects Use Statement

In the conduct of research involving human subjects, the investigator(s) adhered to the policies regarding the protection of human subjects as prescribed by Department of Defense Instruction 32106.02 (Protection of Human Subjects and Adherence to Ethical Standards in DoD-Supported Research) dated 8 November 2011.

IRB Determination and Number

This study was approved by the U.S. Army Medical Research and Development Command Institutional Review Board (protocol log number M-11076) on 22 April 2024 as greater-than-minimal-risk, human subjects research.

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)*
30-01-2026

2. REPORT TYPE
Technical Report

3. DATES COVERED *(From - To)*
10/01/2022 - 9/30/2025

4. TITLE AND SUBTITLE
Physiological Monitoring Under Varied Conditions of Workload and Fatigue

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
6.2

6. AUTHOR(S)
Kelley, A.[1], Vogl, J.[1], Duffy, M.[1,2], Feltman, K.[1], Atchley, A.[1,3], Aura, C.[1], McDermott, D.[1], Melton, J.[1], Gerstner, J.[1], & Curry, I.[1]

5d. PROJECT NUMBER
MO230066

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-TECH-TR--2026-10

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Development Command
Military Operational Medicine Research Program
504 Scott Street
Fort Detrick, MD 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)
USAMRDC MOMRP

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution is unlimited.

13. SUPPLEMENTARY NOTES
[1] US Army Aeromedical Research Laboratory, [2]Oak Ridge Institute for Science and Education, [3]Chickasaw Nation Industries

14. ABSTRACT
In this study, we evaluated physiological outcomes under varied levels of fatigue and cognitive workload as well as performance on a number of tasks (including basic laboratory tasks, operationally relevant laboratory measures [simulated marksmanship, low-fidelity flight simulation] and high-fidelity simulated flight). The study objectives were: 1) To evaluate changes in physiological measurements and human performance under varied conditions of workload and fatigue; and 2) to evaluate the consistency of those changes or patterns across a variety of tasks, spanning cognitive function and functional performance outcomes.

15. SUBJECT TERMS
simulated flight, cognitive workload, sleep deprivation, fatigue, operator state

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Loraine St. Onge, PhD
UNCLAS	UNCLAS	UNCLAS	SAR	70	19b. TELEPHONE NUMBER *(Include area code)* 334-255-6906

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

This page is intentionally blank.

Summary

The next generation of rotary-wing aircraft are anticipated to have advanced capabilities that will require aviators to fly for longer durations and impose a high demand in terms of cognitive workload which will likely translate to increased prevalence of fatigued and overloaded aviators in the cockpit. Considering the established link between fatigue, cognitive workload, and the degradation of performance, which can ultimately pose a risk to safety and mission success, an operator state monitoring (OSM) strategy is under development as a key mitigation measure. Ideally, an OSM system would equip aviators with wireless, wearable sensors that monitor physiological metrics continuously. This data is, theoretically, fed into a processing algorithm using changes in physiology to predict when flight performance will degrade. The system will then communicate with the aircraft to employ some level of adaptive automation as a way to maintain performance. To-date, research has documented the relationships between workload, fatigue, physiology, and performance. However, physiological metrics, a resultant predictive algorithm, and interpretation of that output as it relates to performance, has not been demonstrated under various combinations of workload level and fatigue. The goal of this study was not to test a specific set of sensors for the purpose of OSM, but rather to document how physiology responds under multiple stressors (fatigue and workload), the consistency of those responses across a variety of tasks, and how those changes in physiology can be interpreted with regard to predicted performance on operationally relevant tasks. The findings supported a number of the physiological measures, especially pupil diameter, with sensitivity to fatigue effects. The findings of this study demonstrate consistency of effects in a laboratory setting with limited experimental control (e.g., nutrition and light exposure were unregulated), on functional performance outcomes. Thus, OSM research can justifiably advance to testing in more realistic and field environments.

This page is intentionally blank.

Acknowledgments

The authors would like to express their gratitude to the research team at the U.S. Army Aeromedical Research Laboratory (USAARL) for their contributions to this project. We are especially grateful to USAARL's Flight Systems Branch and Human Integration Team for their tremendous support.

This research was supported in part by an appointment to the Postgraduate 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 Development Command.

This page is intentionally blank.

Table of Contents

Table of Contents (continued)

List of Figures

List of Tables

Page

List of Tables (continued)

This space is intentionally blank.

Introduction

Ongoing work within the military (e.g., Friedl, 2018; Feltman et al., 2022), as well as in the civilian sector (e.g., Pagnotta et al., 2022) continues to pursue the goal of detecting changes in pilots' physical and mental states through the non-invasive measurement of physiological indices such as electrocardiography (ECG), electroencephalography (EEG), electrooculography (EOG), actigraphy, and pupillometry. The overarching goal of operator state monitoring (OSM) within U.S. Army aviation is to use information indicative of an operator's state to drive the aircraft's automation, thus adapting to the operator's needs. While much of the work completed to-date has yielded promising results, many challenges remain. Specifically, the relationships between observed physiological measures, detected degraded cognitive states, and performance, have yet to be defined with respect to how they are effectively captured by current wearable sensors. To maximize the utility of OSM within an operational setting, detecting when performance is likely to be degraded, or has begun to degrade, is essential to the system making use of this information, and ultimately engaging some form of adaptive automation. To-date, the U.S. Army Aeromedical Research Laboratory (USAARL) OSM research efforts have focused on cognitive workload (Feltman et al., 2021; Duffy & Feltman, 2023; Kelley et al., 2023; Feltman et al., 2024; Vogl et al., 2025). These studies have supported the relationships between workload and physiological measures as well as revealed data trends suggesting that physiological measures can be used to predict changes in performance. However, realistic environments are not limited to one stressor, one type of task, or within experimental control. Thus, in this study, understanding the underlying relationships driving an OSM system was further evaluated with the introduction of an additional stressor, fatigue.

Operator State Monitoring

While the concept of OSM may appear straightforward, the development of such a system is highly complex. For the purposes of this study, consider a set of three broad areas essential to OSM development, measurement, prediction, and interpretation (Table 1). This is by no means an exhaustive presentation of the challenges and research questions involved in OSM development. Also, the number of states that could be considered by this type of system is extensive. We have chosen to focus on two stressors or states that have consistently been shown to impact aviator performance as well as been identified as contributing factors to mishaps, workload, and fatigue (e.g., Gaines et al., 2020; Wingelaar-Jagt et al., 2021). Thus, the remainder of the discussion is limited to workload and fatigue.

This space is intentionally blank.

Table 1. Operator State Monitoring Concept Areas

Concept	Research Questions	Description
Measurement	What observable variables are relevant? Does technology support measurement feasibility in the cockpit? What aspects of individual variability need to be measured? What method is used to process measurements?	Physiological variables, estimates of sleep quality and quantity, demographics, health status
Prediction	Is the system detecting a state (e.g., fatigue)? Is the system predicting performance?	State detection may estimate the probability of an adverse state whereas performance prediction may estimate the probability of degradation
Interpretation	How are measurements interpreted? Individually? In combination using a predictive algorithm? How are natural variations within an individual and between individuals controlled for?	Criteria thresholds for measurements or algorithmic output to indicate when automation is to be employed

Measurement.

To-date, considerable work has focused on measurement. Relationships between physiological variables, workload, fatigue, and aviation performance have been evaluated for decades yielding some consistent patterns (Table 2). Inconsistencies may be driven by differences between experimental manipulations, operational definitions of variables, and technology used for measurements, or represent aspects of the true nature of these relationships. The key factor to consider, for this study, is that physiological outcomes manifest differently under varied levels of workload and fatigue. However, in the majority of studies of workload and fatigue, these stressors are studied in isolation leading us to the first research question: *How do these variables respond when measured under a combination of stressors?* As can be seen in Table 2, these physiological measures do not always manifest the same outcomes in the presence of a singular stressor leading to inconsistent results across and within stressors. Eye fixations, for example, have been documented to both *increase* and *decrease* in the presence of high workload. Alternatively, pupil diameter has been shown to *decrease* during fatigue conditions and *increase* in high workload conditions. Again, these inconsistencies may be driven by differences in the studies, but they may also represent true differences in the nature of these relationships. Also, the combined effect of these stressors may exacerbate physiological outcomes and/or performance degradation.

Table 2. Physiological Responses to Workload and Fatigue Conditions

Physiological Measure	Response During Fatigue	Response During High Workload
Heart Rate	Increases (Wilson & Eggemeir, 2020)	Increases (Unni et al., 2017)
Heart Rate Variability (HRV)	Decreases (Stuiver & Mulder, 2014)	Decreases (Delliaux et al., 2019)
Pupil Diameter	Decreases (Morad et al., 2000)	Increases (Wanyan et al., 2014)
Eye Fixation Duration	Increases (Marquart et al., 2015)	Increases (Feng et al., 2018); Decreases (Tsai et al., 2007)
No. of Eye Fixations	Decreases (van Egmond et al., 2022)	Increases (Van Orden et al., 2001)
Blink Count or Rate	Increases (Soleimanloo et al., 2019)	Decreases (e.g., Faure et al., 2016)
Theta Wave Activity (EEG)	Increases (Heitmann et al., 2023)	Increases (Wu et al., 2017)
Alpha Wave Activity (EEG)	Decreases (Heitmann, et al., 2023)	Decreases (Wu et al., 2017)
Beta Wave Activity (EEG)	Increases (Heitmann et al., 2023)	Increases (Chikhi et al., 2022); Decreases (Hussain et al., 2021)

Prediction.

As previously stated, many studies have evaluated the relationships between stressors and physiological outcomes. This research is important to the foundation of an OSM system with respect to chosen metrics; however, it does not provide practical information such as when performance deficits are likely to occur. Ultimately, our understanding of the presence of an adverse state is contingent on the degree to which that state translates to changes in performance. Understanding these changes in performance is essential to determining whether there is an increased risk of a mishap or mission failure. This leads us to a second research question: *Are changes in cognitive states that degrade performance detectable through physiological measures?* Establishing the degree to which physiological measures are indicative of an adverse state that predicts performance levels is an essential component to OSM.

Interpretation.

In an OSM system, once the physiological measurements have been processed and yield a predicted change in performance, those predictions must be compared against a criterion or threshold to decide whether there is a meaningful risk to mission success. Individual differences within and between individuals pose significant challenges to identifying critical thresholds that can be broadly applied. For example, baseline fluid intelligence levels have been shown to predict the degree of performance degradation following total sleep deprivation on a conceptually-related fluid intelligence task (Kurinec et al., 2022). This suggests that one's normal level of functioning could potentially enhance resilience to sleep deprivation induced performance impairments. Potentially, monitoring performance elements and including them in OSM algorithms may help account for these variations. Still, the question of what predicted level of performance degradation should trigger intervention (e.g., adaptive automation) remains and will require input from end-users. This key element of OSM development is beyond the scope of the present study, but the results will inform future studies in this area.

Study Purpose

In this study, we evaluated physiological outcomes under varied levels of fatigue and cognitive workload as well as performance on a number of tasks (including basic laboratory tasks, operationally relevant laboratory measures [simulated marksmanship, aviation-related tasks] and high-fidelity simulated flight). The study objectives were: 1) To evaluate changes in physiological measurements and aviator performance under varied conditions of workload and fatigue; and 2) to evaluate the consistency of those changes or patterns across a variety of tasks, spanning cognitive function and functional performance outcomes. The impact of fatigue on performance was measured over the course of a period of sustained wakefulness (34 hours sleep deprivation). The fatigue manipulation employed has been shown to impact performance in past research (e.g., Alhola & Polo-Kantola, 2007). Performance was repeatedly measured across the study to identify the main and interaction effects of the stressors on performance and physiological measurements. The physiological measurements were modeled as predictors of performance changes in relation to the experimental conditions.

Methods

This study employed a repeated-measures, experimental design manipulating two independent variables, fatigue, and workload. Participants experienced both low and high workload conditions in each test session for all primary tasks (exceptions noted below). Fatigue was manipulated in this study using a 34-hour period of sustained wakefulness, which has been sufficient for inducing fatigue-related performance deficits in past research (e.g., Caldwell et al., 2020). Participants completed two visits to USAARL, an initial and an extended visit. The initial visit occurred at least one day but no more than one week prior to the extended visit. The USAARL sleep suites can house two participants at a time thus most participants completed the study in pairs (three participated individually). All testing was completed individually. Prior to execution, this study was reviewed and approved by the U.S. Army Medical Research and Development Command Institutional Review Board (protocol log number M-11076; USAARL study number 2024-002).

Participants

Participants, recruited from Fort Rucker, AL, were 11 rated, U.S. Army, rotary-wing aviators with a minimum of 200 flight hours, and currently cleared for flight duties. Participants were in a leave status and monetarily compensated for participation ($1500). One participant withdrew (due to fatigue-related discomfort) midway through the study, resulting in 10 complete datasets. All participants were between the ages of 27 and 46 years (*M* = 34.60 years, *SD* = 6.90). Participants were required to refrain from the consumption of: 1) stimulants (including caffeine) for a minimum of 16 hours prior, and 2) alcohol and over-the-counter medications that may induce drowsiness for 24 hours prior. Additionally, a minimum of 6 hours of sleep the night prior to participation in the sustained wakefulness portion of the study, confirmed with actigraphy-based sleep estimates, was required. Participants were also free of the following exclusion criteria, as assessed by a study physician:

- Currently taking medications that induce drowsiness, such as prescribed or over-the-counter antihistamines, affect cognitive function, or affect the physiological parameters collected in this study (e.g., heart rate). Any self-medication was assessed through self-report.
- Nicotine/tobacco use as it could negatively impact the results for these users to abstain for the duration of the study as well as have negative medical side effects.
- Excessive, regular caffeine consumption (more than 600 milligrams per day) as it could negatively impact the results for these users to abstain for the duration of the study as well as have negative medical side effects.
- Female participants with known pregnancy, who test positively for pregnancy, or refuse the urine test.
- Any history of psychological/psychiatric disorder.
- Any history of significant cardiovascular disease or hypertension.
- Any history of sleep disorders (e.g., sleep apnea)

Measures

Instruments and tasks used in this study are divided in four categories: questionnaires, physiological measures, cognitive and laboratory tasks, and military functional tasks. In addition, they are divided by frequency and timing of administration.

Single-administration questionnaires at initial visit.

Questionnaires completed during the initial visit were administered using hardcopies.

Demographics and personal history questionnaire.

The demographics and personal history questionnaire collected age, sex, ethnicity, flight experience, education, handedness, and vision information. Additionally, the questionnaire included screening questions relevant to eligibility (e.g., tobacco use, alcohol use, medical conditions, medications, and caffeine use). This questionnaire was developed in-house and was completed during the initial visit (Appendix B).

Beck Depression Inventory.

Depression symptoms were measured using the Beck Depression Inventory-II (BDI-II; Beck et al., 1996). The BDI-II is a commonly used 21-item, multiple-choice self-report measure that captures affect, cognition, and physical symptoms of depression over the most recent two-week period. The outcome is a single score, where higher scores indicate greater endorsement of depression symptoms. Threshold guidelines suggest scores of 0 to 13 indicate minimal symptoms, 14 to 19 mild, 20 to 28 moderate, and 29 to 63 severe.

Raven's Progressive Matrices.

The Raven's Progressive Matrices (Raven, 2000) consists of 48 visual items requiring participants to solve. Participants have 45 minutes to complete the assessment. The items become progressively more difficult. This measure assesses abstract reasoning and fluid intelligence. The outcome measure is a raw score converted into a standard score.

Single-administration questionnaires at onset of extended visit.

Questionnaires completed at the onset of the extended visit were done electronically.

Adult Attention Deficit/Hyperactive Disorder Self Report Scale Symptom Checklist.

The Adult Attention Deficit/Hyperactive Disorder Self Report Scale Symptom Checklist (ASRS) contains 18 items and requires 2 minutes for completion. It was developed in conjunction with the World Health Organization and the Workgroup on Adult Attention Deficit/Hyperactive Disorder (Kessler et al., 2005) and is used as a screening tool with adult patients. The items are consistent with the Diagnostic and Statistical Manual of Mental Disorders, version IV criteria (American Psychiatric Association, 2000). The outcomes included an inattentive subscale score, hyperactive subscale score, and a total score.

Sleep habits questionnaire.

The sleep habits questionnaire was developed in-house following readily available sleep history questionnaires used in medical screenings (e.g., Stanford Health Care Sleep Questionnaire). The items include questions to ascertain an individual's typical sleeping patterns (average amount of sleep, typical time in bed, typical wake time, average sleep latency, average sleep disturbances) (Appendix B).

State-Trait Anxiety Inventory.

The State-Trait Anxiety Inventory (STAI) (Spielberger & Reheiser, 2004) is a widely used 40-item, self-report anxiety inventory rated on 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. It was included for the purposes of this study to capture both an individual's short-term expression of anxiety (i.e., state anxiety) as restricted by directions in the instructions asking participants to think only of the past two weeks, and an individual's enduring experiences of anxiety symptoms. 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. The primary outcome measures are state and trait scores.

Repeated-administration questionnaires.

Two questionnaires, administered electronically, were repeated at seven timepoints throughout the extended wakefulness period to assess fatigue and mood. These timepoints, labeled sessions, approximately corresponded with the following length of wakefulness:

- Session 1: 1 hour of wakefulness
- Session 2: 7 hours of wakefulness
- Session 3: 12 hours of wakefulness
- Session 4: 18 hours of wakefulness
- Session 5: 23.5 hours of wakefulness
- Session 6: 29 hours of wakefulness
- Session 7: 31.5 hours of wakefulness

Karolinska Sleepiness Scale.

The Karolinska Sleepiness Scale (KSS) is a well-validated single item questionnaire that asks participants to rate how sleepy they feel in the moment (Kaida et al., 2006). The KSS measures daytime sleepiness with higher scores indicating greater daytime sleepiness.

Profile of Mood States – Short Form.

The Profile of Mood States – Short Form (POMS-SF) is a valid and reliable short version of the POMS, a measure of psychological distress and mood (McNair et al., 1981). The POMS-SF contains 35 items; in each an adjective is provided and the participant rates how much it describes them using a 5-point Likert scale format (Curran et al., 1995). The POMS-SF was administered to evaluate the degree to which sustained wakefulness impacted mood states. The scale outputs 7 subscale scores: tension, anger, vigor, esteem-related affect, fatigue, depression, and confusion. Additionally, a total mood disturbance score is computed by summing the scores for the “negative” affect subscales (tension, anger, fatigue, depression, and confusion) and subtracting the “positive” affect subscales (vigor and esteem-related affect).

Subjective workload assessments.

Two measures were used to assess workload associated with a subset of tasks. These were both administered electronically.

Instantaneous Self-Assessment of Workload.

To assess workload manipulations during test sessions (simulated flights and USAARL Multi-Attribute Task Battery [MATB]), participants were asked to rate workload using an adaptation of the Instantaneous Self-Assessment of Workload (ISA) technique (Brennen, 1992; Jordan, 1992). The workload ratings use those of the Crew Status Survey Workload Scale (Ames & George, 1993), provided in Table 3. Participants were asked at one-minute increments to rate their workload during simulated flights. If no response was provided, a “7” was recorded. An alternate version with a 10-point scale was used during the USAARL MATB (Table 4).

Table 3. Instantaneous Self-Assessment of Workload Used During Simulated Flights

Level	Description
1	Nothing to do; no system demands.
2	Light activity; minimal demands.
3	Moderate activity; easily managed; considerable spare time.
4	Busy; challenging but manageable; adequate time available.
5	Very busy; demanding to manage; barely enough time.
6	Extremely busy; very difficult; non-essential tasks postponed.
7	Overloaded; system unmanageable; essential tasks undone; unsafe.

Table 4. Instantaneous Self-Assessment of Workload Labels for the USAARL MATB

Level	Workload	Spare Capacity	Description
1			
2	Underutilized	Very Much	Little or nothing to do. Rather boring.
3			
4	Relaxed	Ample	More time than necessary to complete tasks. Time passes slowly.
5			
6	Comfortable	Some	The controller has enough work to keep him/her stimulated. All tasks are under control.
7			
8	High	Very Little	Certain non-essential tasks are postponed. Could not work at this level very long. Controller is working at their limit. Time passes quickly.
9			
10	Excessive	None	Some tasks are not completed. The controller is overloaded and does not feel in control.

National Aeronautics and Space Administration-Task Load Index.

The National Aeronautics and Space Administration (NASA)- Task Load Index (TLX) (Hart & Staveland, 1988) is a questionnaire that measures participant workload. The participant rates the previous task, in this case flight, 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. This instrument was completed following flight simulator, USAARL MATB, and *n*-back tasks.

Cognitive and laboratory tasks.

All tests were administered electronically.

Psychomotor Vigilance Task.

The Psychomotor Vigilance Task (PVT) (Loh et al., 2004) 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 using the Pison wrist-worn device (Khitrov et al., 2014) for three minutes. At the onset of the task, participants made a closed fist. When a light visual stimulus appeared, participants responded to the stimulus by opening their hand. Participants reset for each trial by closing their fist. Participants' reaction time to acknowledging the stimulus was recorded, along with false starts and missed responses (lapses). The PVT was completed at approximately the following times relative to a 0800-wakeup time during the extended visit.

- Session 1: 1 hour of wakefulness
- Session 2: 7 hours of wakefulness
- Session 3: 12 hours of wakefulness
- Session 4: 18 hours of wakefulness
- Session 5: 23.5 hours of wakefulness
- Session 6: 29 hours of wakefulness
- Session 7: 31.5 hours of wakefulness

Laboratory and cognitive tasks repeated with four administrations.

There were four tasks, administered electronically, that were administered at four timepoints throughout the extended wakefulness period. These were completed at approximately the following times relative to an 0800-wakeup time during the extended visit:

- Session 1: 4.5-5.5 hours of wakefulness
- Session 2: 12.5-13.5 hours of wakefulness
- Session 3: 20.5-21.5 hours of wakefulness
- Session 4: 26.5-27.5 hours of wakefulness

Additionally, one task was administered at three time points. During all tasks, EEG, EOG, and ECG were measured. Eye tracking was also measured during the USAARL MATB and *n*-back tasks.

Stroop Task.

The Stroop task is a well-established cognitive test of selective attention (Macleod, 1991). In this task, participants are presented with color words, and must name the color that the word is printed in and ignore the meaning of the word. Participants completed 10 trials of each congruent and incongruent color-word pairs. Stroop effect interference was the key outcome measure and is the mean difference in reaction time between congruent and incongruent trials.

Neurofit ONE.

The Neurofit ONE is an oculomotor assessment system that uses a desktop mounted screen with an embedded infared-based camera system to isolate and record the eyes. This system does require that the participant restrict their head movement with a chinrest to maintain a stable image of the eyes. Participants then track a visual stimulus with their eyes only. The primary outcome of interest is the "nfit" score, a composite oculometrics score that quantifies functional health of neural circuitry and pathways (https://www.neurofit.tech/neurofit-one; Liston & Stone, 2014).

Rapid Visual Information Processing Task.

The Rapid Visual Information Processing Task (RVIP) is a well-validated measure of sustained attention (Bakan, 1959). In each trial, participants are presented with a sequence of digits ranging from 2 to 9 in length and must detect "target" sequences within those presented. Difficulty is manipulated using the length of the "target" sequence as well as the speed of the sequence presentation (2 levels: slow [1,200 milliseconds (ms)], fast [600 ms]). Participants completed six blocks of trials. During the task, EEG, EOG, and ECG were measured.

USAARL Multi-Attribute Task Battery.

The USAARL MATB serves as a basic multitasking simulation that engages operators in a combination of four subtasks typical to aviation, system monitoring, target tracking, communications, and resource management. This classic task battery, originally developed by NASA, has been used in many studies (e.g., Bowers et al., 2014; Sato et al., 2020). The USAARL MATB offers the same suite of four subtasks that have been utilized over the decades but with additional customized features (see Vogl et al., 2024, for details).

To interact with the subtasks, the participants used a joystick and mouse to input their responses. Participants were instructed to use their right hand for the mouse control and left for the joystick control. Task demand was individually tailored based on participant performance during a training session (see Appendix C for a full description). Participants completed five trials within each of the four administration sessions. Performance for each of the USAARL MATB subtasks was recorded for each simulation loop and a composite USAARL MATB score was derived for each trial. Instantaneous workload ratings (every 30 seconds) were also collected throughout completion of the MATB. For the purposes of this report, only the composite scores are reported. At the end of each trial, the participant completed the NASA-TLX.

Domain-Specific n-Back Battery.

The multimodal *n*-back battery administers the classic *n*-back paradigm across different cognitive domains varying by stimuli and presentation types (Wickens et al., 2002) (Figure 1). The *n*-back paradigm is presented as a sequential memory task that asks participants to store and update a series of data in their working memory. As participants are presented additional stimuli, the participant is tasked with comparing the currently presented stimulus with the stimulus that was presented *n* iterations prior (Kirchner, 1958). Participants indicate that the comparison yields a match by pressing the 'Match' button, or that the stimuli do not match by pressing the 'Not Match' button. The task is adaptive in nature such that the *n* adjusts based on performance. For each subtask within each session, the *n* would start at 1 in the first trial. If accuracy exceeded 80%, the *n* would increase in the next trial, if accuracy was below 50%, the *n* would decrease in the next trial but not below an *n* of 1, and if the accuracy was between 50% and 80%, the *n* would stay the same. Three, one-minute trials for each of the four subtasks were presented in each session (12 minutes of testing total). The *n* value in the *n*-back task typically varies between 0 and 5, beyond which is outside of normal human performance limits (i.e., without extensive training).

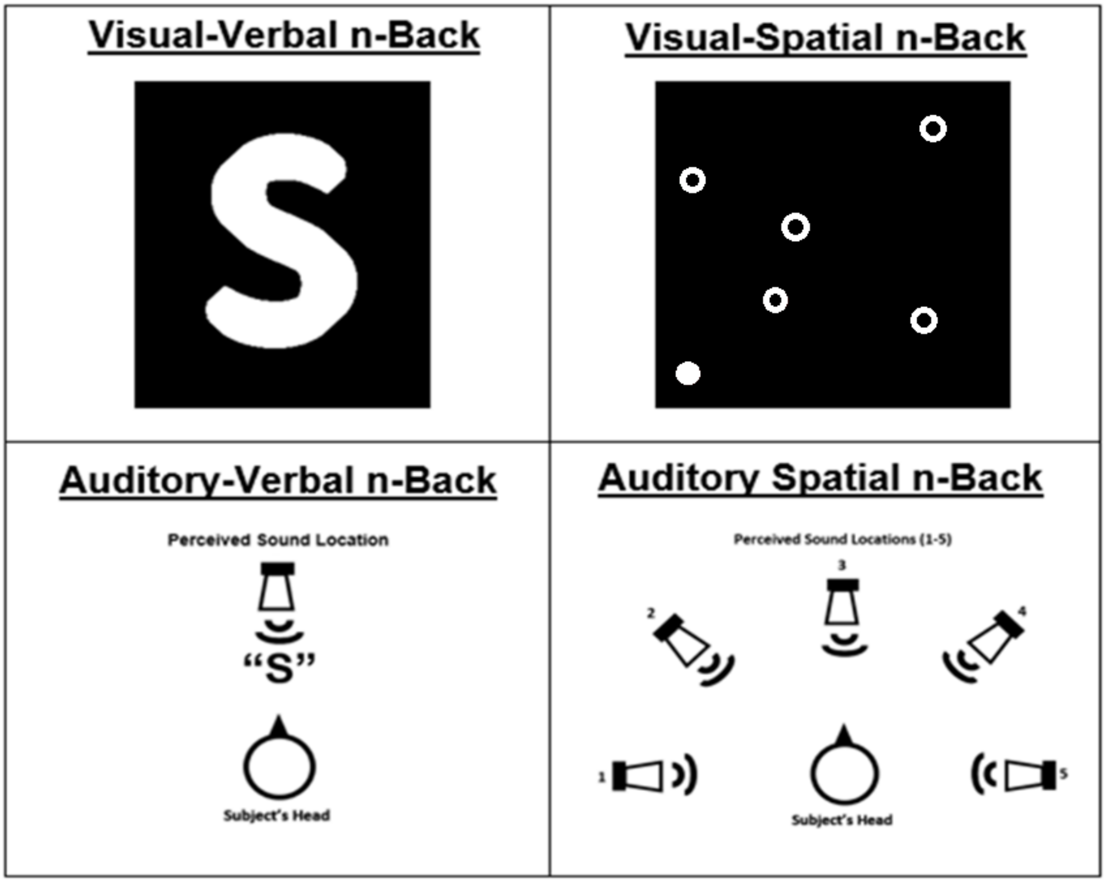

Figure 1. A depiction of the stimuli used in the multimodal *n*-back task battery featuring four variations of the *n*-back paradigm.

The *n*-back sessions were completed at approximately the following times relative to a 0800-wakeup time during the extended visit.

- Session 1: 7.5-8 hours of wakefulness
- Session 2: 21.5-22 hours of wakefulness
- Session 3: 29.5-30 hours of wakefulness

Functional tasks.

Standard marksmanship task.

The weapons simulator used for this task is the Engagement Skills Trainer (EST) 2000, a small arms training device used by the U.S. Army. As can be seen in Figure 2, a participant fires from a lane (USAARL's EST 2000 has a five-lane configuration) at "targets" which appear on a projection screen at a distance of 26 feet (ft) 3 inches from the firing line. In the standard marksmanship qualifying task, participants shoot at 40 targets presented sequentially using a rifle (M4) that has been modified to use with the EST 2000 but maintain their form, fit, feel, and function. The targets vary in distance, from 50 to 300 meters. The scenario entails the participant firing from three positions, prone supported, prone unsupported, and kneeling. The key dependent variable is throughput (accurate shots per second). Qualification scores are defined by number of hits: 36-40 for “expert,” 30-35 for “sharpshooter,” 23-29 for “marksman,” and less than 23 is “unqualified.” During the initial visit, participants zeroed their weapon (i.e., aligned the laser sensor to the equivalent of the mechanical weapon zero) and conducted a practice session. The simulated marksmanship task was completed at approximately the following times relative to an 0800-wakeup time during the extended visit:

- Session 1: 8-9 hours of wakefulness
- Session 2: 14.5-15.5 hours of wakefulness
- Session 3: 30-31 hours of wakefulness

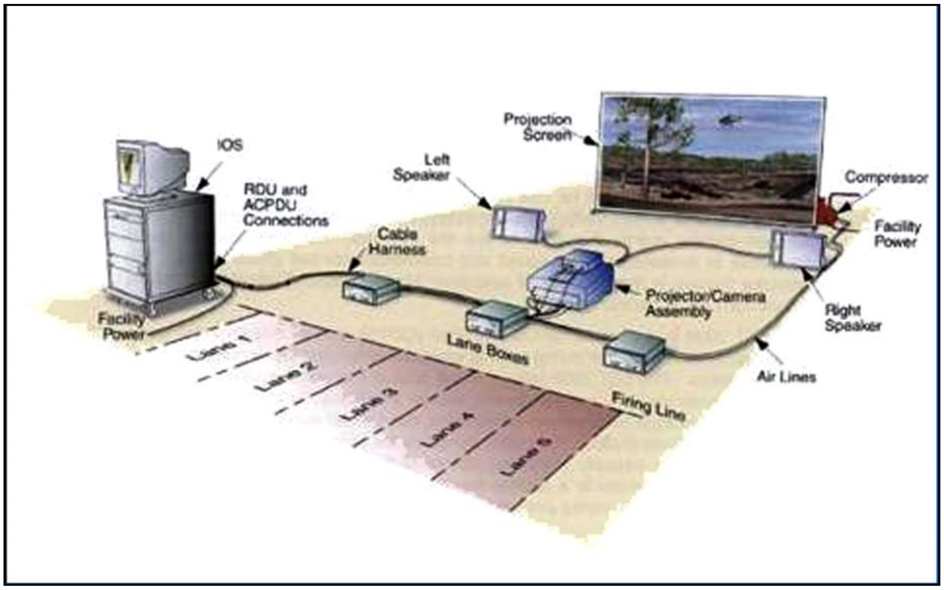

Figure 2. EST 2000 set-up.

Flight simulator and flight tasks.

The USAARL NUH-60 flight simulator simulates a UH-60M aircraft and consists of a simulator compartment containing a cockpit, visual display system, instructor/operator station, control room observation station, and a six-degree-of-freedom motion system. The cockpit includes all the UH-60M avionics, upper and lower console switches, power control quadrant and circuit breakers. The visual display system consists of twelve image generator channels, a ten-foot radius collimated optical display providing a 200 x 45 degrees field of view and two chin displays. The visual system simulates the natural helicopter environment surroundings for day, dusk, or night, with blowing sand or snow options for a degraded visual environment. At the beginning of each session, the eye tracking system was calibrated for the participant.

Participants completed a total of eight flight scenarios over four sessions during the extended visit. Two flight scenarios were completed during the flight familiarization block on the initial visit. All scenarios were designed to include both “on-the-controls” flight performance measures (e.g., altitude) and additional tasks in the cockpit (e.g., fuel calculations) which were graded as pass/fail by the research pilot to increase face validity of the sessions. These flight scenarios were designed such that those in session 1 were comparable to those in session 3, and session 2 to 4 (details to follow in results section). Routes and data collection forms are included in Appendix D. The key outcome variables of interest were deviations from specified airspeed, deviations from specified altitude, and rate of closure (speed at which aircraft approaches the ground). The simulated flights were completed at approximately the following times relative to an 0800-wakeup time during the extended visit:

- Session 1: 2-3 hours of wakefulness
- Session 2: 8-9 hours of wakefulness
- Session 3: 24-25 hours of wakefulness
- Session 4: 32-33 hours of wakefulness

Physiological measures.

Physiological measures (ECG, EEG, and eye tracking) were recorded during the completion of the following tasks: Stroop, RVIP, USAARL MATB, *n*-back, and flight simulation. Physiological data were recorded and synchronized with performance data using Lab Streaming Layer (LSL).

At the beginning of each session, a one-minute baseline ($n = 0$) was collected using the *n*-back display to control for luminance changes. For this baseline, participants selected “Match” for each response as soon as each new stimuli appeared.

EEG.

EEG was recorded using the *Advanced Brain Monitoring B-Alert X-24* (Advanced Brain Monitoring, Carlsbad, CA) wireless wet electrode system with 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). See Appendix E for a description of data processing. Outcome measures included power spectral density values (delta, 1-4 hertz (Hz); theta, 4-8 Hz; alpha, 8-13 Hz; and beta, 13-30 Hz), workload-related metrics, calculations described in Appendix E, frontal theta/alpha ratio, frontal beta/alpha ratio, parietal alpha power, global band activity.

ECG.

The Polar H10 is a chest strap ECG device that uses patented dry electrodes placed within the chest strap itself to record a bipolar differential voltage between the electrodes into its output electrocardiogram. The ECG is recorded with a sample rate of 200 Hz. Custom in-house developed software was used to perform processing stages and compute the heart rate and R-R intervals (distance between two consecutive R waves) for subsequent HRV analyses. See Appendix E for ECG data processing description. Outcome measures included heart rate (beats per minute [BPM]), three time-domain HRV measures (standard deviation of R-R intervals, root mean square of successive differences in R-R intervals [RMSSD], percentage of successive R-R interval differences exceeding 50 milliseconds [ms]), and one frequency-domain HRV measure (low-frequency to high-frequency ratio).

Eye tracking and pupillometry.

Eye tracking and pupillometry were measured using two systems. During the cognitive and laboratory tasks, the Pupil Labs Pupil Core Binocular Headset was used, which consists of three small cameras mounted on a lightweight eye frame, without lenses, that are worn on the face like eyeglasses. A forward-facing camera records the visual scene in front of the person at a rate of up to 200 Hz at a resolution of 1920 x 1080 pixels. Two smaller cameras positioned on the frames and facing each of the participant's eyes record each eye with an infra-red sensor at a rate of up to 120 Hz and a resolution of 400 x 400 pixels. These images are processed by calibrated computer vision algorithms which estimate both pupil diameter, eyelid position, and the direction of gaze mapped to the forward-facing camera image in near-real-time.

In the flight simulator, the SmartEye ProDX system was used, which consists of an array of four cameras mounted around the participant, each looking at the participant's head at a different angle to provide effective tracking over a much larger range of head positions compared to remotely mounted systems that use only a single camera system. See Appendix E for data processing. Outcome measures for both included pupil diameter and blink count; see Appendix E for descriptions on how these were calculated.

Actigraphy measures.

Two devices were used to measure actigraphy, with one also including a version of the PVT.

Actigraphy.

The Phillips actiwatch is a wrist activity monitor that collects human activity data and yields data that, using a validated algorithm, estimates sleep latency, sleep quality, total sleep, number of sleep bouts, mean length of sleep, number of immobile minutes, number of immobile phases, and mean length of immobility. The actiwatch is an acceptable substitute for use under circumstances where EEG is not possible to derive sleep estimates (e.g., Kilgore et al., 2009).

Pison Salus.

The Pison Salus device is a lightweight, wrist-worn device that contains electrodes that record electromyography (EMG) data from five bipolar channels at a sample frequency of 1 kilohertz (kHz). The Pison sensor is a wearable wristband that uses EMG to detect electrical signals in peripheral nerves that initiate hand and finger movements. It contains two electrodes in contact with the inside of the wrist that record EMG signals, along with onboard processing for signal conditioning and transmission. Data is streamed wirelessly to a mobile device for analysis. Additionally, this device administers the PVT using hand motion (measured with EMG) as a response mechanism.

Procedure

Initial visit.

The duration of the initial visit was approximately 8 hours. During the initial visit, volunteers were briefed on the study and given a tour of the sleep suites where they would be staying during the extended visit. An investigator on the study team reviewed the consent form with the volunteer(s) and if they chose to participate, participants then provided written informed consent. Females were administered a urine pregnancy test. Next, participants met privately with a study physician to complete the demographics (e.g., age, sex; military and aviation experience such as number of deployments and flight hours) and personal history questionnaire to determine eligibility. Once cleared for participation by the study physician, participants were enrolled in the study and issued an actiwatch to wear for the duration of the study (including the period between the initial visit and extended visit) and received training on operating the watch. Participants completed practice and familiarization sessions with the cognitive and laboratory tasks (including the Instantaneous Self-Assessment of Workload), flight simulator tasks, and on the marksmanship trainer. Short breaks occurred between tasks and for lunch. Prior to release from the initial visit, participants reviewed the requirements (e.g., restrictions on caffeine and alcohol, what to bring) with a member of the study team.

Extended visit.

The extended visit duration was 46 to 49 hours (depending on length of recovery sleep) with a period of 34-hours of sustained wakefulness. On the first day of the extended visit, participants confirmed adherence to participation requirements with a member of the study team, including review of the sleep estimates provided by the actiwatch and self-report. Participants again met briefly and privately with the study physician to ensure continued eligibility. Once cleared, participants began the 34-hour total sustained wakefulness phase of the study. During this period, the order and timing of activities was maintained as closely as possible to ensure that the experimental activities were administered on a consistent and repeating schedule. Technical errors prevented exact timing and strict adherence to the study schedule, which is unsurprising given the number of recording devices and pieces of equipment used concurrently. Participants were always accompanied by a research team member for safety purposes as well as to ensure they maintained wakefulness. Participants were provided meals three times each day and given one-hour for physical training each day. They were allowed full access to the designated laboratory space and gym equipment. During breaks, participants had access to caffeine-free beverages and snacks. Following completion of the 34-hour sustained wakefulness period, participants were allowed up to 13.5 hours of recovery sleep. Upon waking, the study physician met briefly with each participant, provided counseling regarding sleepiness and safety, and released them from the study.

Throughout the duration of the study, participants completed schedule blocks according to Figure 3. Descriptions of the measures and durations of each block are provided in Table 5.

	Initial Visit	Extended Visit		
6:00			Shower/breakfast/ recreation	
7:00			Q & I	
8:00		Arrival	Flight simulator	Shower/ breakfast
9:00	Arrival/informed consent	Initial Q & I		Screen/ release
10:00	Medical screening Cognitive and lab tasks practice	Flight simulator	Break Cognitive and lab tasks	
11:00				
12:00	Lunch	Lunch	Lunch	
13:00	Flight familiarization	Cognitive and lab tasks	Q & I *n*-back	
14:00	Marksmanship calibration	Break	Marksmanship	
15:00	Release	Q & I *n*-back	Break Q & I	
16:00		Flight simulator and marksmanship	Flight simulator	
17:00				
18:00		PT	Dinner/shower Recovery sleep	
19:00		Dinner/shower		
20:00		Q & I Cognitive and lab tasks		
21:00				
22:00		Marksmanship		
23:00				
0:00		Recreation		
1:00				
2:00		Q & I Cognitive and lab tasks (including short break)		
3:00				
4:00				
5:00		*n*-back		

Figure 3. Study schedule. Q = questionnaires, I = instrumentation.

Table 5. Descriptions of All Study Schedule Blocks

	Block	Physiological Measures	Tasks	Duration	Number of Sessions
Initial Visit	Informed consent & medical screening	NA	Demographics and personal history questionnaire Raven's progressive matrices	60 to 90 minutes	1
	Cognitive tasks practice	NA	Stroop test RVIP Neurofit	30 minutes	1
	Lab tasks practice	NA	USAARL MATB tutorial and practice trials	60 minutes	1
	Flight familiarization	NA	Simulated flight tasks ISA	60 to 90 minutes	1
	Marksmanship calibration and practice	NA	EST 2000 weapon zeroing EST 2000 practice round of record fire	30 minutes	1
Extended Visit	Initial instrumentation	NA	Application of all physiological devices EEG impedance checks Calibration of eye tracking system 5-minute baselines for each ECG and EEG	60 to 90 minutes	1
	Instrumentation	NA	EEG impedance checks Reapplication of electrodes	15 minutes	6
	Initial questionnaires	NA	Profile of Mood States Sleep habits questionnaire Beck Depression Inventory State Trait Anxiety Inventory Adult ADHD Self-Report Scale Symptom Checklist	20 minutes	1
	Questionnaires	NA	Karolinska Sleepiness Scale POMS-SF PVT (Pison watch)	10 minutes	6
	Flight simulator	EEG ECG EOG Pison watch Eye tracking	2 simulated flight scenarios ISA (1-minute intervals during flight scenarios) NASA-TLX (following each flight scenario)	60 minutes	4
	Lab tasks	EEG ECG EOG Pison watch	USAARL MATB task series ISA NASA-TLX	60 minutes	4

	Eye tracking			
Cognitive tasks	EEG ECG EOG Pison watch Eye tracking	Stroop test RVIP Neurofit	60 minutes	4
n-back	EEG ECG EOG Eye tracking Pison watch	*n*-back task series ISA NASA-TLX	30 minutes	3
Recreation, meals, & breaks	NA	Recreational activities (e.g., watch television, ping pong, board games, cards)	Varied	10
Marksmanship	EEG EOG ECG	Standard marksmanship qualification task	40 minutes	3
Recovery sleep	NA	Sleep in private bedroom	Up to 13.5 hours	1

Note. NA = not applicable

Statistical Analysis and Quality Control

All data were inspected for impossible values and technical errors prior to analyses.

The study employed a repeated measures design manipulating workload and fatigue. Prior to any analyses, data distributions and descriptive statistics were explored. Linear mixed-effects models (LMMs) were used to assess main effects of fatigue and workload and an interaction effect of fatigue and workload on performance outcomes, physiological measures, and perceived workload levels. Family-wise error rate was controlled for by using a Bonferroni correction for post-hoc pairwise comparisons, and all relevant *p*-values are corrected using this approach. Specifications of analyses varied by task and are noted in the results.

Results and Discussion

Descriptive statistics on the questionnaires completed at the initial visit and sleep estimates provided by the actiwatch for the time between the initial and extended visits are provided in Table 6. Results of these measures do not suggest any problematic values/responses with respect to inclusion/exclusion criteria and confounding factors.

Table 6. Descriptive Statistics for All Questionnaires

Measure/Questionnaire	*n*	Mean (*SD*)	Median
Flight Experience			
UH-60 Hours	9	1471.11 (*1433.83*)	1000
CH-72 Hours	1	500 (NA)	500
AH-64 Hours	1	590 (NA)	590
TH-67 Hours	6	76.33 (*13.06*)	80
UH-72 Hours	7	648.57 (*835.27*)	120
OH-58 Hours	2	84.00 (*90.51*)	84
Raven's Progressive Matrices			
Standardized Score	11	114.00 (*12.60*)	116
STAI			
State Score	11	24.91 (*5.70*)	23
Trait Score	11	26.27 (*4.88*)	25
ASRS			
Total Score	11	20.64 (*8.54*)	18
Hyperactivity Score	11	10.27 (*5.04*)	11
Inattention Score	11	10.36 (*4.41*)	9
BDI			
Total Score	11	1.18 (*1.40*)	1
Sleep Habits			
Avg sleep latency - minutes	11	20.40 (*14.80*)	10
Avg sleep duration - hours	11	7.90 (*0.84*)	8
Avg sleep disruption - # of awakenings	11	3.11 (*2.93*)	2
Actiwatch			
Sleep disturbances pre-extended visit	10	33.90 (*10.96*)	35.5
Sleep disturbances post-recovery sleep	8	35.25 (*11.82*)	33
Sleep duration pre-extended visit	10	488.10 (*119.72*)	499
Sleep duration post-recovery sleep	8	626.13 (*99.88*)	641
Sleep latencies pre-extended visit	10	98.33 (*216.27*)	40.33
Sleep latencies post-recovery sleep	8	8.71 (*13.81*)	1.915

Note. SD = standard deviation

Fatigue Manipulation Check

In order to assess the efficacy of the sustained wakefulness period in inducing fatigue, LMMs were conducted on outcomes from four assessments completed throughout the extended visit; however, the number of administration sessions varied between assessments: KSS (7 sessions), PVT (7 sessions), POMS-SF (6 sessions), and Neurofit (4 sessions). All missing data were due to technical recording errors. Results supported the fatigue manipulation across all tasks (Table 7 and Figure 4). Specifically, KSS scores, PVT number of lapses, PVT mean reaction time, and POMS-SF scores increased across sessions, whereas Neurofit scores decreased.

Table 7. Fatigue Manipulation Check Summary Results Presenting *p* Values (Effect Size [Partial Eta Squared])

Task/Questionnaire & Outcome Measures	**Fatigue (Session)**
KSS	
Score	< 0.001 (0.68)
PVT	
Mean reaction time	< 0.001 (0.48)
Number of lapses	< 0.001 (0.46)
POMS-SF	
Fatigue subscore	< 0.001 (0.70)
Total mood disturbance score	< 0.001 (0.73)
Neurofit	
Nfit score	0.03 (0.51)

Note. Green shading indicates statistical significance.

This space is intentionally blank.

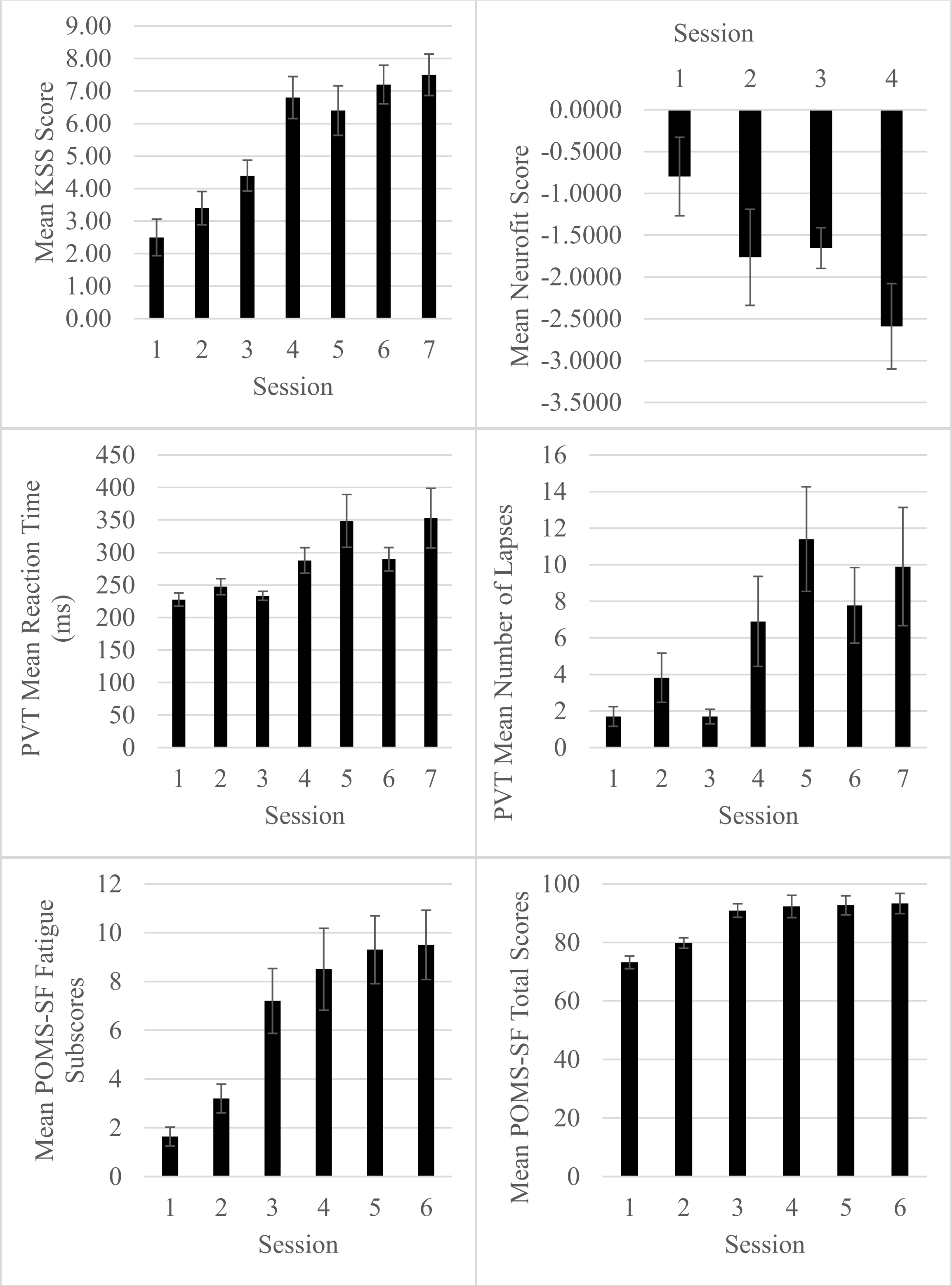

Figure 4. Mean scores, reaction times, and lapses across sessions. Error bars represent standard error of the mean.

Fatigue Effects on Tasks without Workload Manipulations, and Physiological Metrics

Stroop Task.

LLMs were used to analyze these data in the context of a repeated measures design (Table 8). Family-wise error rate was controlled for by using a Bonferroni correction for post-hoc pairwise comparisons, and all relevant *p* values are corrected using this approach (13 tests).

Table 8. Stroop Task Results Summary of Main Effect and Interaction *p* Values (Effect Size [Partial Eta Squared])

Metric	Fatigue
Performance Metrics	
Accuracy	0.93 (0.18)
Reaction Time	0.14 (0.26)
Physiological Metrics	
Heart Rate (BPM)	1.00 (0.07)
HRV (RMSSD)	1.00 (0.11)
Pupil Diameter	0.007 (0.38)
Blink Count	1.00 (0.10)
Frontal Theta: Alpha	1.00 (0.12)
Frontal Beta: Alpha	0.75 (0.19)
Parietal Alpha	1.00 (0.17)
Global Alpha	0.75 (0.19)
Global Beta	0.03 (0.33)
Global Delta	1.00 (0.12)
Global Theta	0.15 (0.26)

Note. Green shading indicates statistical significance.

No effects of fatigue were seen on performance. Mean accuracy was stable across sessions and exceeding 90%, suggesting that there may be a ceiling effect.

There were significant effects on physiological metrics, specifically pupil diameter and global beta activity derived from EEG (Figure 5). Pupil diameter decreased as is expected with increasing levels of fatigue from sessions 1 and 2 to sessions 3 and 4. Global beta activity, however, tended to decrease relative to the first session which may be an indication of a practice effect rather than indicative of a fatigue effect.

This space is intentionally blank.

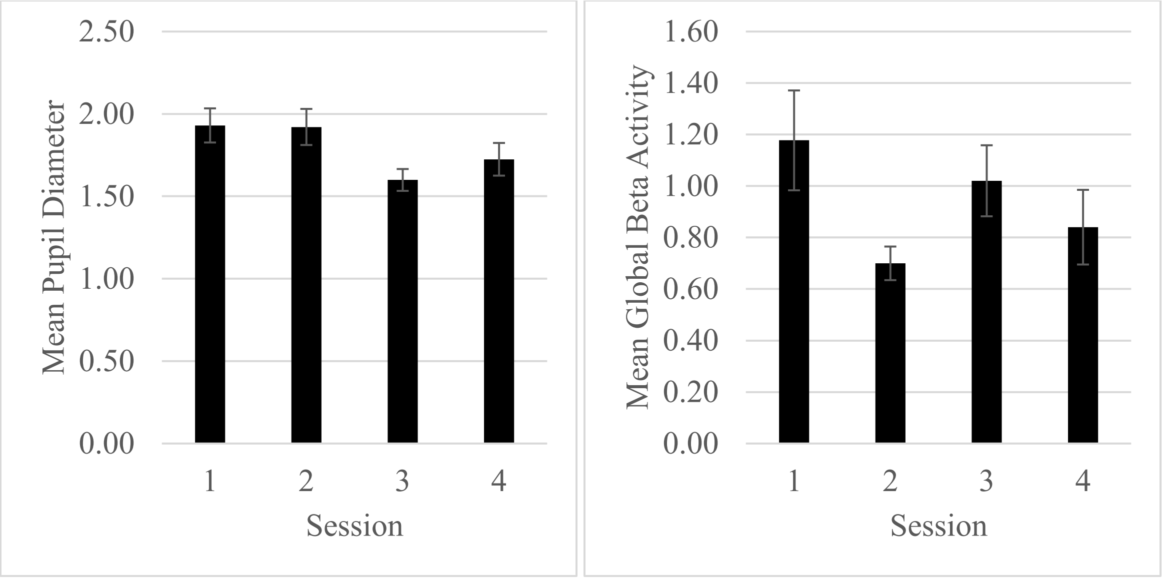

Figure 5. Main effects of fatigue on physiological outcomes during Stroop task execution. Error bars represent standard error of the mean.

Marksmanship.

Nine participants had complete datasets and were included in the analysis (the excluded participant did not yield any reliable data). Of those, 0 were "expert" during their first extended visit test session, 2 were "sharpshooter," 3 were "marksman," and 4 were "unqualified." An LMM was used to evaluate the impact of fatigue (three sessions) on marksmanship performance (throughput) which did not yield any significant results, $p = 0.29$. Considering that marksmanship skill level was skewed toward the lesser skilled categories, a floor effect may have masked any potential fatigue effects.

Workload and Fatigue Effects on Performance Outcomes, Subjective Measures, and Physiological Metrics

RVIP task.

LMMs were used to analyze these data in the context of a repeated measures design (Table 9). Family-wise error rate was controlled for by using a Bonferroni correction for post-hoc pairwise comparisons, and all relevant *p* values were corrected using this approach.

This space is intentionally blank.

Table 9. RVIP Task Results Summary of Main Effect and Interaction *p* Values (Effect Size [Partial Eta Squared])

Metric	Workload	Fatigue	Workload * Fatigue
Performance Metrics			
Perceptual Sensitivity (d’)	< 0.001 (0.278)	< 0.001 (0.312)	1.00 (0.109)
Accuracy	0.001 (0.242)	< 0.001 (0.346)	1.00 (0.027)
Reaction Time	< 0.001 (0.527)	0.002 (0.282)	1.00 (0.067)
Physiological Metrics			
Heart Rate (BPM)	1.00 (0.028)	0.437 (0.172)	1.00 (0.055)
HRV (RMSSD)	1.00 (0.020)	0.305 (0.183)	1.00 (0.011)
Pupil Diameter	1.00 (0.001)	0.002 (0.306)	1.00 (0.025)
Blink Count	0.001 (0.244)	0.001 (0.318)	1.00 (0.011)
Frontal Theta: Alpha	1.00 (0.004)	1.00 (0.065)	1.00 (0.031)
Frontal Beta: Alpha	1.00 (0.003)	1.00 (0.090)	1.00 (0.011)
Parietal Alpha	0.005 (0.210)	< 0.001 (0.339)	1.00 (0.033)
Global Alpha	0.003 (0.227)	< 0.001 (0.426)	1.00 (0.058)
Global Beta	1.00 (0.004)	1.00 (0.074)	1.00 (0.012)
Global Delta	0.118 (0.133)	1.00 (0.113)	1.00 (0.004)
Global Theta	1.00 (0.022)	0.072 (0.212)	1.00 (0.003)

Note. Green shading indicates statistical significance.

Performance measures.

Workload and fatigue main effects were seen with all three performance outcomes (Figures 6 and 7). Specifically, accuracy, mean reaction time, and d’ were greater in the low workload condition versus the high condition. While the decreased accuracy and d’ values in the high workload versus low condition is in line with expectations, the slower mean reaction times in the high condition may suggest that participants sacrificed accuracy for speed in response. Alternatively, the fatigue effects were as expected with greater reaction times and reduced accuracy and d’ values across sessions.

This space is intentionally blank.

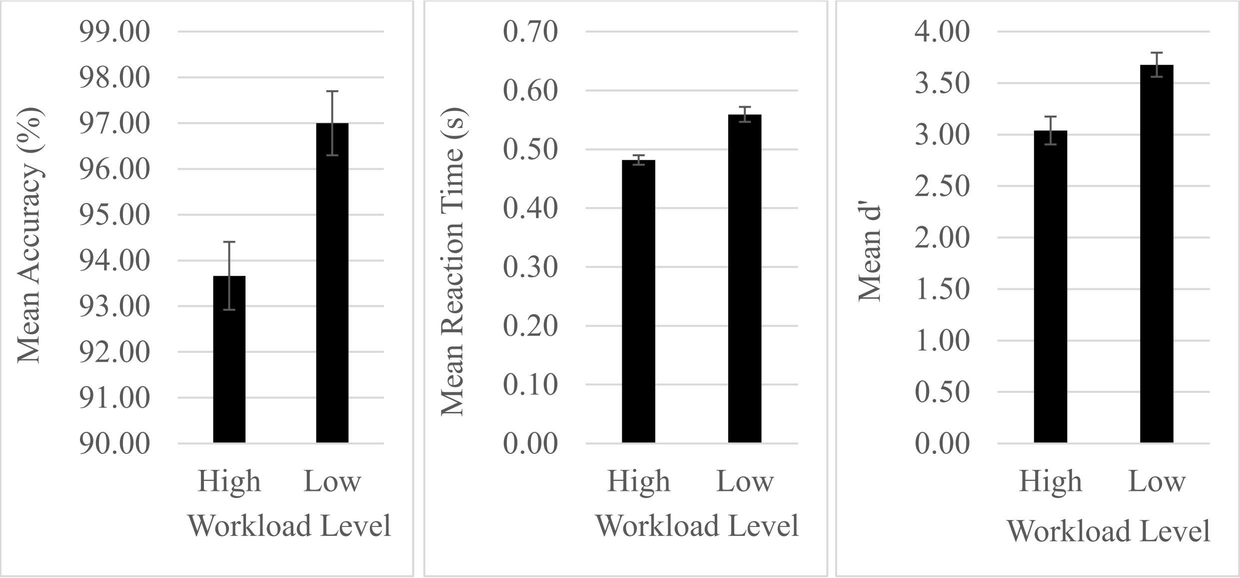

Figure 6. Main effects of workload on RVIP performance outcomes. Error bars represent standard error of the mean.

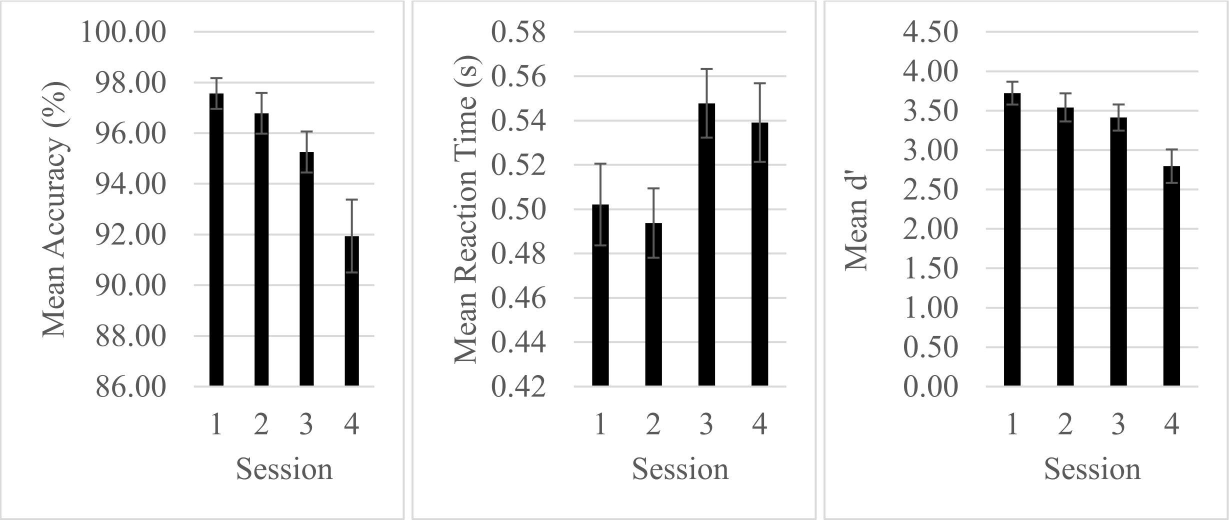

Figure 7. Main effects of fatigue (session) on RVIP performance outcomes. Error bars represent standard error of the mean.

Physiological measures.

There were significant main effects of workload (Figure 8) and fatigue (Figure 9) on parietal alpha activity derived from EEG, global alpha activity, and blink count. Pupil diameter was significantly affected by fatigue only. The decreased alpha activity in the high versus low workload condition are as expected, whereas the increased alpha activity in later sessions is not as expected. The decreased blink count in high versus low workload is as expected given that the RVIP workload manipulations are in the visual domain. Interestingly, blink count significantly increased in comparison to the initial session aligning with expectations for a fatigue effect.

Considering that reaction time was quicker in the high versus low workload condition (inconsistent with expectations), it is possible that these unexpected findings in reaction time, blink counts, and alpha brain activity (relative to fatigue) are a reflection of a confounding factor (e.g., motivation, participant strategy, practice).

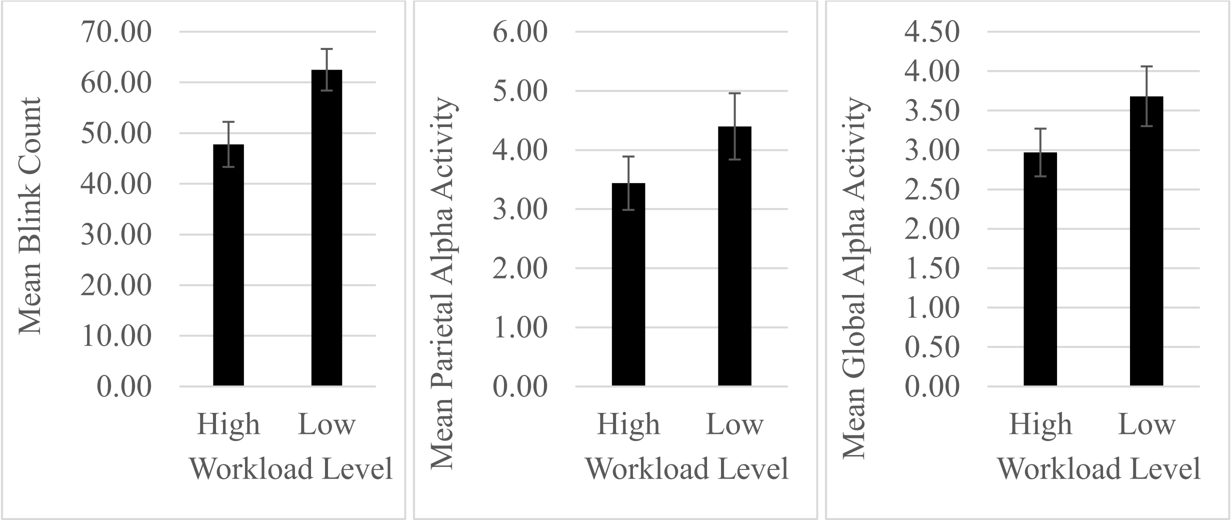

Figure 8. Main effect of workload on physiological measures during RVIP execution. Error bars represent standard error of the mean.

This space is intentionally blank.

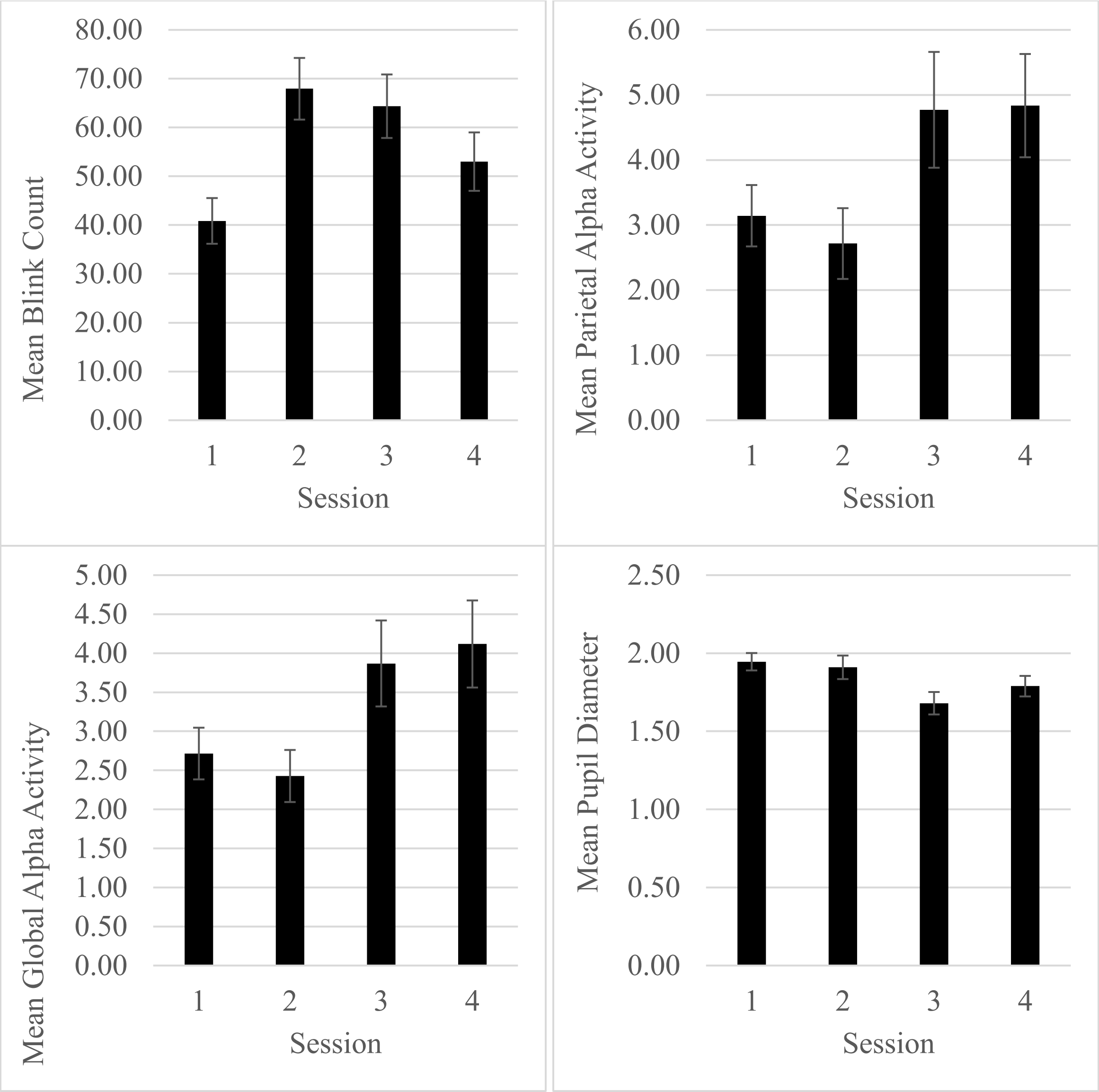

Figure 9. Main effects of fatigue on physiological measures during RVIP execution. Error bars represent standard error of the mean. Pupil diameter is heavily skewed to the lower end of the acceptable range, a byproduct of the brightness of the monitors.

USAARL MATB.

LLMs were used to analyze these data in the context of a repeated measures design. Participants' workload levels in the experimental sessions are defined using feedback provided during the training session. Unexpectedly, this feedback yielded two indistinguishable levels for some participants, thus these two levels were collapsed in order to avoid inflating within-subjects comparisons for these participants only. For the remaining participants, these levels were treated as distinct. This hybrid approach preserved true workload distinctions while ensuring model validity. Family-wise error rate was controlled for by adopting a Bonferroni correction for post-hoc pairwise comparisons, and all relevant *p* values are corrected using this approach. Table 10 summarizes the results.

Table 10. USAARL MATB Task Results Summary of Main Effects and Interaction *p* Values (Effect Size [Partial Eta Squared])

Metric	**Workload**	**Fatigue**	**Workload * Fatigue**
Performance Metrics			
Multitasking Efficiency	< 0.001 (0.811)	< 0.001 (0.201)	0.028 (0.176)
Composite Performance	<0 .001 (0.708)	< 0.001 (0.182)	< 0.001 (0.283)
System Monitoring	< 0.001 (0.336)	0.502 (0.066)	0.148 (0.156)
Communications	< 0.001 (0.152)	0.003 (0.120)	< 0.001 (0.479)
Tracking	< 0.001 (0.872)	< 0.001 (0.221)	1.00 (0.047)
Resource Management	< 0.001 (0.585)	0.382 (0.069)	1.00 (0.064)
Subjective Metrics			
ISA Scores	< 0.001 (0.759)	0.006 (0.129)	1.00 (0.131)
ISA Reaction Time	0.005 (0.146)	0.001 (0.151)	0.080 (0.186)
NASA TLX Composite	< 0.001 (.556)	1.00 (0.040)	0.024 (0.179)
Mental Demand	< .001 (.585)	1.00 (0.026)	0.005 (0.197)
Physical Demand	< .001 (.332)	0.008 (0.111)	1.00 (0.121)
Temporal Demand	< .001 (.619)	1.00 (0.030)	0.011 (0.189)
Performance	< .001 (.378)	0.116 (0.083)	1.00 (0.098)
Effort	< .001 (.477)	1.00 (0.020)	1.00 (0.091)
Frustration	< .001 (.335)	1.00 (0.008)	0.471 (0.141)
Physiological Metrics			
Heart Rate (BPM)	1.00 (0.028)	< 0.001 (0.297)	1.00 (0.034)
HRV (RMSSD)	0.024 (0.063)	< 0.001 (0.171)	1.00 (0.032)
Pupil Diameter	1.00 (0.042)	< 0.001 (0.220)	1.00 (0.033)
Blink Count	1.00 (0.043)	1.00 (0.051)	1.00 (0.028)
Frontal Theta: Alpha	1.00 (0.016)	0.229 (0.079)	1.00 (0.026)
Frontal Beta: Alpha	1.00 (0.023)	0.006 (0.119)	1.00 (0.025)
Parietal Alpha	0.021 (0.118)	< 0.001 (0.335)	1.00 (0.084)
Global Alpha	0.167 (0.094)	< 0.001 (0.212)	1.00 (0.071)
Global Beta	1.00 (0.058)	1.00 (0.041)	1.00 (0.047)
Global Delta	1.00 (0.028)	0.066 (0.093)	1.00 (0.044)
Global Theta	1.00 (0.017)	0.028 (0.103)	1.00 (0.031)

Note. Green shading indicates statistical significance.

Performance measures.

The main effect of workload was significant across all performance metrics, with multitasking efficiency, composite performance, system monitoring, communications, tracking, and resource management all demonstrating strong effects (all *p* values < .001, partial η^2 = .152-.872). These large effect sizes are in line with the observed consistent degradation in performance with increasing workload. In contrast, fatigue effects were seen on the multitasking efficiency and communications subtasks, suggesting impairments in specific task components, rather than global impairments, as fatigue accumulated. A significant workload * fatigue interaction emerged for multitasking efficiency, composite performance, and communications (*p* < .05, partial η^2 = .176-.479).

Physiological measures.

Workload effects on physiological measures were significant for HRV (RMSSD) and parietal alpha ($p < .05$), although the effect sizes were small to moderate (partial $\eta^2 = .063\text{-}.118$). Follow-up comparisons for HRV showed that no post-hoc pairwise workload level comparisons met the Bonferroni corrected criteria (all p values $\geq .446$), indicating that while workload affected HRV overall, the effect was diffuse across levels. Similarly, post-hoc comparisons for parietal alpha activity indicated no significant pairwise differences after correction ($p > 0.05$). These results indicate that only some physiological markers were sensitive to workload changes, though less robustly than behavioral or subjective indicators.

Fatigue showed particularly strong effects on physiological metrics (e.g., heart rate, HRV (RMSSD), pupil diameter, parietal alpha, and global alpha, all p values $< .001$, partial $\eta^2 = .171\text{-}.335$), highlighting these as sensitive indicators of prolonged sleep deprivation. Specifically, post-hoc comparisons indicated that session 3 (20.5-21.5 hours of wakefulness) was associated with significantly lower HR than all other sessions (all p values $< .05$), while sessions 1 and 2 exhibited the highest mean HR values. HR values decreased from session 2 to session 3 by approximately 7 BPM across all workload levels. This fatigue-related suppression in heart rate was robust and consistent regardless of workload condition. Post-hoc comparisons confirmed that pupil diameter significantly decreased in sessions 3 and 4 compared to sessions 1 and 2. Descriptive means showed a drop in pupil size of approximately 0.3–0.4 millimeters (mm) from the first two to the last two sessions, consistent across workload levels. For parietal alpha activity, multiple significant pairwise differences emerged; session 3 exhibited significantly greater parietal alpha activity than sessions 1 ($p < .001$), 2 ($p < .001$), and 4 ($p = .002$), and session 4 was significantly greater than session 2 ($p < .001$) and session 1 ($p = .019$). Overall, parietal alpha increased significantly with fatigue level, peaking during session 3. Other EEG measures, including frontal beta:alpha, global alpha, and global theta, also reflected fatigue-related changes. Interaction effects were not present among the physiological metrics after correction, suggesting that most physiological responses to workload and fatigue occurred independently.

Physiological responses were the least uniformly sensitive across the domains compared to performance and subjective metrics. Only two metrics, HRV and parietal alpha, showed reliable workload effects after correction, and the effect sizes were small. This indicates that, at least in the present task context, many physiological indicators may lack the resolution to detect workload variations unless paired with more fine-grained or time series analyses. In contrast, physiological sensitivity to fatigue was more pronounced. Heart rate, HRV, pupil diameter, and multiple EEG bands (especially alpha-related activity) responded robustly to sleep deprivation fatigue, consistent with neurophysiological models linking fatigue with cortical slowing and autonomic dysregulation (Lal & Craig, 2002; Jap et al., 2009). Notably, interaction effects were not present among physiological metrics, which may suggest that fatigue and workload modulate physiology via separate pathways (autonomic versus cortical) rather than interacting together. However, further analyses are needed to confirm this interpretation.

Subjective measures.

Workload significantly influenced all subjective workload measures, including ISA scores, ISA reaction time, and each NASA-TLX metric (all p values < .001, partial η^2 = .146-.759), demonstrating large perceived changes in workload with increasing task load. Fatigue effects were more nuanced. ISA scores, ISA reaction times, and the NASA-TLX physical demand subscale were significantly affected by fatigue (p < .05), though effect sizes were small (partial η^2 = .111-.151). Only three interaction effects emerged among the NASA-TLX scores (total, mental demand, and temporal demand) with significant workload and fatigue interactions (p < .05), but with modest effect sizes (partial η^2 = .179-.197).

Subjective workload ratings were broadly sensitive to both workload and fatigue, but the drivers differed. All NASA-TLX and ISA-based measures increased reliably with workload, mirroring participants' perceptions of elevated cognitive demand as presented by the distinct workload levels presented in the study. Fatigue effects were more subtle, reaching significance only for ISA scores, ISA reaction time, and physical demand. Interestingly, only a few subjective metrics showed interaction effects, and these were modest in size, suggesting that while fatigue and workload may independently shape perceptions, their compounded effects are not always additive. These results emphasize the need to interpret self-report metrics as reflecting distinct experiential components of operator state rather than as simple proxies for performance decline.

***n*-back task.**

LMMs were utilized to analyze these data based on workload (*n*-Level), fatigue (session), and interaction effects. A Bonferroni correction was applied to control family-wise error rate. Results are summarized in Table 11.

This space is intentionally blank.

Table 11. *n*-Back Results Summary of Main Effect and Interaction *p* Values (Effect Size [Partial Eta Squared])

Metric	**Workload**	**Fatigue**	**Workload * Fatigue**
Performance Metrics			
d’	< 0.001 (0.15)	0.005 (0.05)	1.00 (0.00)
Criterion	1.00 (0.01)	1.00 (0.00)	1.00 (0.00)
RT for Correct Responses	< 0.001 (0.10)	1.00 (0.01)	< 0.001 (0.10)
RT for Erroneous Responses	< 0.001 (0.14)	1.00 (0.04)	0.015 (0.06)
Subjective Metrics			
NASA TLX Total Score	< 0.001 (0.17)	1.00 (0.02)	1.00 (0.02)
Mental Demand	< 0.001 (0.25)	0.013 (0.05)	1.00 (0.02)
Physical Demand	1.00 (0.00)	< 0.001 (0.12)	1.00 (0.00)
Temporal Demand	< 0.001 (0.12)	0.372 (0.03)	1.00 (0.00)
Performance	0.138 (0.04)	1.00 (0.02)	1.00 (0.00)
Effort	< 0.001 (0.10)	1.00 (0.01)	1.00 (0.02)
Frustration	1.00 (0.00)	1.00 (0.00)	1.00 (0.02)
Heart Rate (BPM)	1.00 (0.00)	< 0.001 (0.07)	1.00 (0.01)
HRV (RMSSD)	1.00 (0.00)	0.007 (0.06)	1.00 (0.00)
Pupil Diameter	0.431 (0.03)	< 0.001 (0.12)	1.00 (0.00)
Blink Count	1.00 (0.00)	0.004 (0.05)	1.00 (0.00)
Frontal Theta: Alpha	0.503 (0.03)	< 0.001 (0.13)	1.00 (0.01)
Frontal Beta: Alpha	1.00 (0.01)	0.003 (0.06)	1.00 (0.03)
Parietal Alpha	0.067 (0.04)	< 0.001 (0.09)	1.00 (0.02)
Global Alpha	1.00 (0.01)	0.011 (0.05)	1.00 (0.00)
Global Beta	1.00 (0.00)	1.00 (0.02)	1.00 (0.02)
Global Delta	0.353 (0.03)	0.038 (0.04)	1.00 (0.01)
Global Theta	1.00 (0.01)	0.163 (0.03)	1.00 (0.00)

Note. Green shading indicates statistical significance. RT = reaction time.

Performance measures.

Workload had a significant main effect on the performance metrics d-prime, reaction time for correct responses, and reaction time for erroneous responses. However, there was no impact of workload on criterion. Fatigue had a significant effect on one performance metric, d-prime. Interestingly, there were significant interaction effects on two performance measures, reaction time for correct responses and that for incorrect responses. Specifically, both reaction times for correct and incorrect responses were faster for *n* = 1 trials in session 1 compared to sessions 2 and 3 (Figure 10).

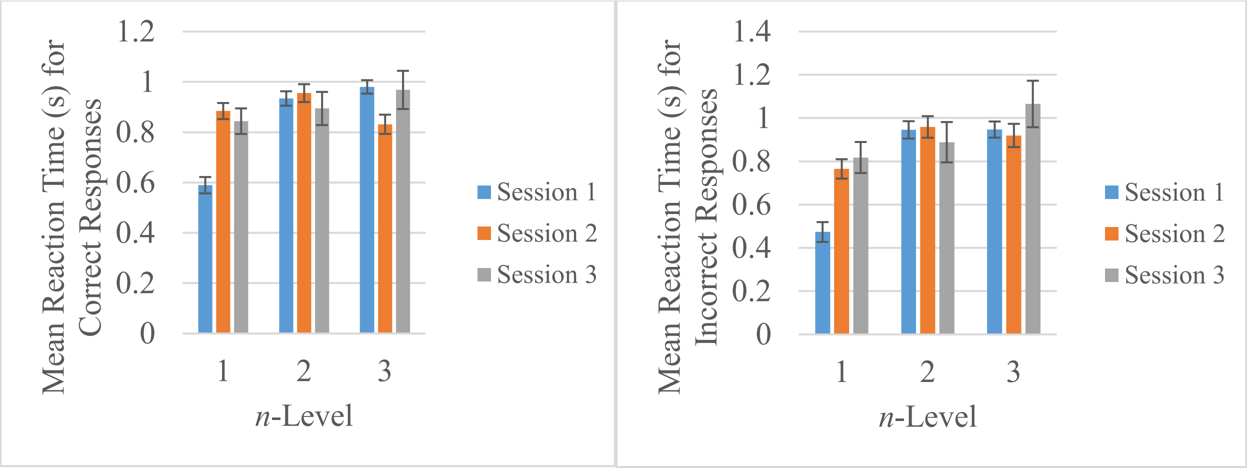

Figure 10. Reaction times for correct (left) and incorrect (right) responses by *n*-level and session.

The results indicate that the performance metrics for the adaptive *n*-back task were generally worse at higher *n* levels, which is consistent with previous literature (Brouwer et al., 2012; Mehler et al., 2009). Specifically, d-prime decreased with higher levels of *n* and the reaction times for both correct and erroneous responses increased with higher levels of *n*. Fatigue also negatively impacted d-prime, which is consistent with previous findings that noted that accuracy on the *n*-back decreased due to sleep deprivation (Lo et al. 2012, Martinez-Cancino et al., 2015). The interaction shown is consistent with previous work indicating that the sharpest decrease in reaction time occurs for the lowest *n* trials of the *n*-back when sleep deprivation occurs (Martinez-Cancino et al., 2015). These findings indicate that the manipulation of workload in this study was effective, and the impact of fatigue on performance is consistent with the literature.

Physiological measures.

No workload main effects, nor interaction terms were significant on any of the physiological measures recorded. There was a main effect of fatigue on all the physiological measures, except for global beta, and global theta. It is possible that fatigue effects on physiology overshadowed any workload effects or that sensitivity is such that a more fine-grained approach (time-series analysis) is needed.

Subjective measures.

A workload main effect on NASA-TLX composite scores showed that scores increased with difficulty level (*n*-Level). There was no main effect of fatigue (session) nor an interaction on NASA-TLX total scores. When analyzing subscores, workload significantly impacted mental demand, temporal demand, and effort whereas fatigue impacted physical demand and temporal demand. No interaction effects were seen on any of the NASA-TLX subscores. The observed differences in mental demand suggest that operators found the *n*-back more difficult when they were more fatigued. However, the impact on physical demand is more likely a reflection of the physical toll of fatigue rather than the physical demand of the *n*-back task itself.

Simulated flight.

Workload manipulation check.

To evaluate whether the workload manipulations in the simulated flights effectively induced changes to perceived workload, paired-samples *t*-tests were used to determine whether mean NASA-TLX scores (subscale scores and total score) differed between high and low workload flights measured following the first simulator flight session. From these, ratings on the physical demand, effort, and temporal demand subscales were significantly different between workload conditions (Table 12).

Table 12. NASA-TLX Paired-Samples *t*-Test Results for Session One, Significant Outcomes Only

NASA-TLX Subscale	*t*-value	*p*-value	High Workload Mean (SE)	Low Workload Mean (SE)
Physical Demand	4.11	0.003	29.50 (3.91)	18.00 (3.00)
Effort	2.50	0.037	40.00 (6.75)	23.00 (3.59)
Temporal Demand	2.68	0.025	38.50 (4.35)	26.50 (4.95)
Total	3.02	0.014	30.83 (4.12)	22.08 (3.18)

Note. SE indicates standard error of the mean.

Overall results.

LMMs were used to analyze these data in the context of a repeated measures design. Family-wise error rate was controlled for by using a Bonferroni correction for post-hoc pairwise comparisons, and all relevant *p* values are corrected using this approach. Results are summarized in Table 13. All measures were aggregated for the duration of the flights.

This space is intentionally blank.

Table 13. Simulator Flight Results Summary of Main Effects and Interaction *p* Values (Effect Size [Partial Eta Squared])

Metric	Workload	Fatigue	Workload * Fatigue
Performance			
Airspeed Deviations *Sessions 2 vs. 4*	0.61 (0.01)	<0.01 (0.28)	<0.001 (0.74)
Rate of closure deviations *Sessions 1 vs. 3*	0.06 (0.13)	0.045 (0.14)	<0.001 (0.69)
Physiological Metrics			
Heart Rate (BPM)	1.00 (0.002)	< 0.001 (0.196)	1.00 (0.020)
HRV (RMSSD)	1.00 (0.007)	0.009 (0.081)	0.295 (0.050)
Pupil Diameter	1.00 (0.000)	< 0.001 (0.113)	1.00 (0.021)
Blink Rate	1.00 (0.008)	1.00 (0.017)	0.066 (0.064)
Frontal Theta: Alpha	1.00 (0.001)	< 0.001 (0.106)	0.004 (0.087)
Frontal Beta: Alpha	1.00 (0.003)	1.00 (0.010)	1.00 (0.011)
Parietal Alpha	1.00 (0.002)	.279 (.050)	1.00 (0.017)
Global Alpha	1.00 (0.000)	1.00 (0.028)	1.00 (0.020)
Global Beta	1.00 (0.001)	1.00 (0.036)	1.00 (0.006)
Global Delta	1.00 (0.000)	0.675 (0.042)	1.00 (0.017)
Global Theta	1.00 (0.002)	1.00 (0.030)	1.00 (0.019)
Subjective Metrics			
Mental Demand	0.305 (0.02)	<0.001 (0.39)	0.585 (0.03)
Physical Demand	0.724 (0.000)	<0.001 (0.31)	0.106 (0.10)
Effort	0.966 (0.000)	<0.001 (0.39)	0.073 (0.12)
Temporal Demand	0.603 (0.000)	0.004 (0.21)	0.032 (0.14)
Performance	0.383 (0.01)	0.607 (0.03)	0.20 (0.01)
Frustration	1.00 (0.00)	.003 (0.22)	0.22 (0.08)
Total Score	0.63 (0.00)	<0.001 (0.36)	0.22 (0.08)

Note. Green shading indicates statistical significance.

Performance measures.

Flight scenarios were designed such that performance outcomes in session 1 were comparable to those in session 3, and session 2 to 4; resulting in two sets of flight data. Flight scenarios, outcome measures, and comparisons are outlined in Table 14. LLMs were used to evaluate effects of workload, fatigue, and the interaction of the two for each set. The primary performance outcome for sessions 1 and 3 was the deviation from specified rate of closure while airspeed and altitude deviations were used as the primary performance outcomes in sessions 2 and 4.

Table 14. Summary of Flight Scenarios and Data Collected

Session	Route	Event #	Workload	Task	Task Parameters	Outcome(s)
			Comparison Set #1			
1	2	5	High	LZ Approach (SMOKE)	Rate of closure, descend < 300 fpm	Rate of closure (RMSD)
1	C	2	Low	Load at PZ	Rate of closure, descend < 300 fpm	Rate of closure (RMSD)
3	3	2	High	Ridgeline Landing (SMOKE)	Rate of closure, descend < 300 fpm	Rate of closure (RMSD)
3	B	5	Low	Landing DVE (SMOKE)	Rate of closure, descend < 300 fpm	Rate of closure (RMSD)
			Comparison Set #2			
2	4	2	High	Frequency change	Adjust frequency, maintain 4000' MSL and 100 KIAS	Airspeed deviation (RMSD) Altitude deviation (RMSD)
2	E	4	Low	Loitering	Flight profile after takeoff is 1500' MSL and 100 KIAS	Airspeed deviation (RMSD)
4	5	3	High	Redirect for LZ	Flight profile after takeoff is 1500' MSL and 100 KIAS	Altitude deviation (RMSD)
4	D	4	Low	Attitude change	Flight profile after takeoff is 3000' MSL and 100 KIAS	Airspeed deviation (RMSD)

Note. LZ = landing zone, PZ = pick-up zone, DVE = degraded visual environment, fpm = feet per minute, MSL = mean sea level, KIA = knots indicated airspeed, RMSD = root mean squared deviation.

For set number 1 (sessions 1 and 3), there was a main effect of fatigue (session), $p = 0.045$, partial $\eta^2 = 0.14$ and interaction effect between session and workload, $p < 0.001$, partial $\eta^2 = 0.69$, on rate of closure. Contrasts showed faster rate of closure (feet per minute) during high workload at session 1, compared to low workload, session 3, (t (26) = -3.96, $p = 0.0005$). The opposite was found for the low workload conditions, with a faster rate of closure in feet per minute during session 3 as compared to session 1 (t (26) = 6.76, $p < 0.001$) (Figure 11).

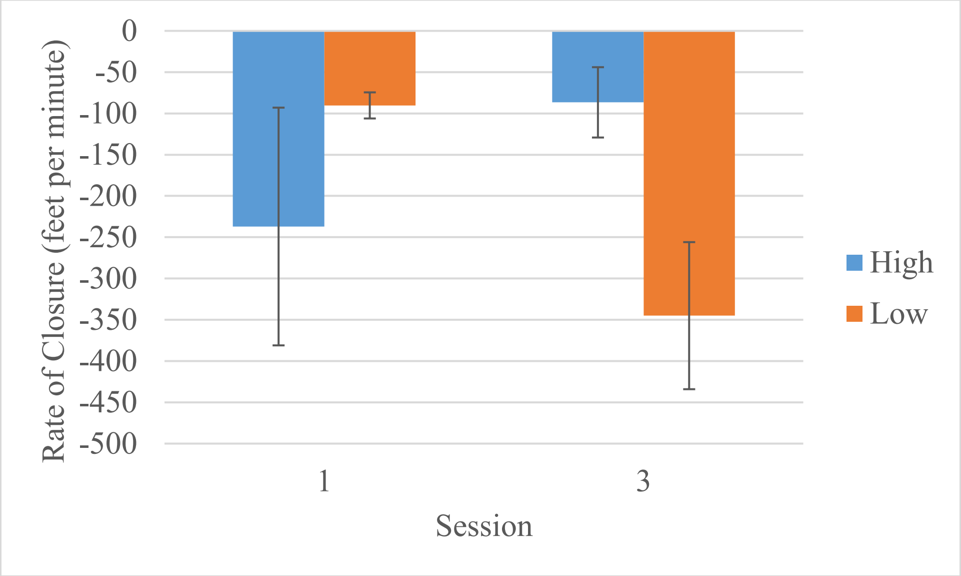

Figure 11. Interaction effect of fatigue (session) and workload on rate of closure. Error bars represent standard error of the mean.

For set number 2 (sessions 2 and 4), there was a main effect of fatigue (session), p = 0.003, partial $\eta^2 = 0.28$, and interaction effect between session and workload, $p < 0.001$, partial $\eta^2 = 0.74$, on airspeed deviations. Contrasts showed greater airspeed deviations in session 2 high workload vs. session 4 high workload, $t(28.3) = 4.37$, $p = 0.0002$, whereas the opposite pattern was seen in the low workload condition, with greater deviations in session 4 vs. session 2, $t(29.2)$ = -8.66, $p < 0.0001$ (Figure 12).

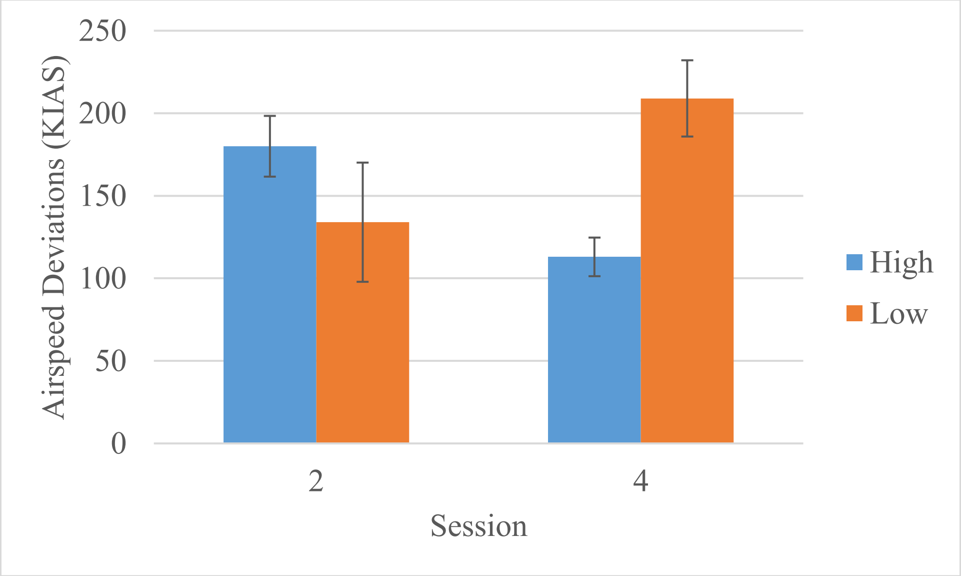

Figure 12. Interaction effect of fatigue (session) and workload on airspeed deviations. Error bars represent standard error of the mean.

During the flights, participants completed additional tasks not on the controls which were rated as pass/fail by the research pilots. To evaluate the effects of fatigue and workload on proper task execution, the frequency of "unsatisfactory" ratings were calculated across each participants' flights. Then, an LMM was used to evaluate the effects of fatigue (session) and workload (high vs. low), along with their interaction, on the frequency of unsatisfactory ratings. The model included a random intercept for each participant to account for repeated measures. There was a significant main effect of fatigue on the number of unsatisfactory ratings, $F(3,63) = 5.19$, $p = 0.003$. Compared to flight session 1, the number of unsatisfactory ratings were significantly higher in session 2 ($\beta = 0.80$, $SE = 0.17$, $t(72) = 3.43$, $p = 0.001$) and in session 4 ($\beta = 0.50$, $SE = 0.17$, $t(72) = 2.14$, $p = 0.04$). The main effect of workload was not significant ($p = 0.14$), nor was there a significant interaction between fatigue and workload ($p = 0.51$). Figure 13 displays the estimated marginal means by session number and workload level. While there was no significant interaction with workload, Figure 13 suggests a trend toward more unsatisfactory ratings in the high workload conditions for sessions 2 and 4.

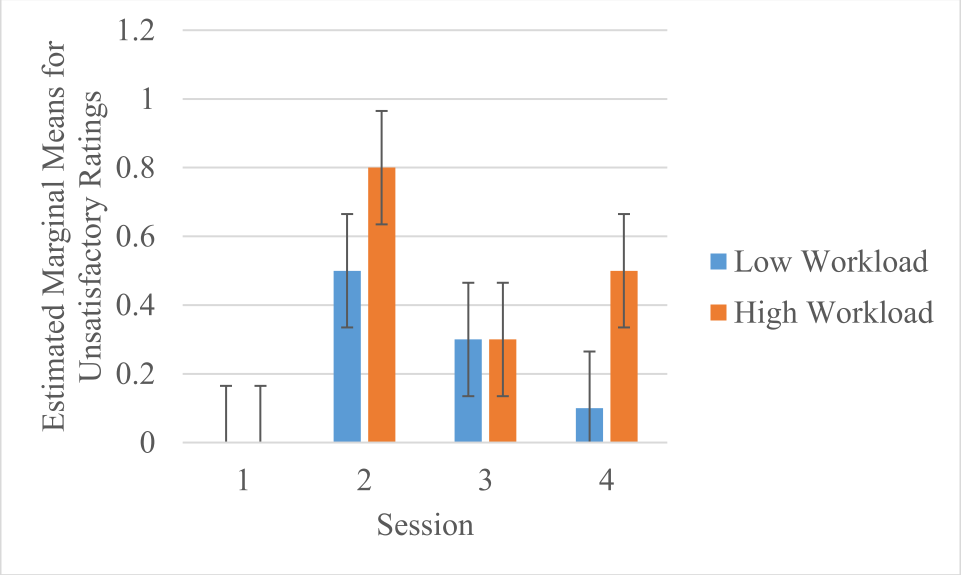

Figure 13. Estimated marginal means by fatigue (session) and workload level for unsatisfactory flight performance. Error bars represent standard error of the mean.

Flight performance overall was impacted by both workload and fatigue. In sessions 1 and 3, the prescribed rate of closure was 300 fpm, thus increasingly negative rates of closure are "faster" and considered "stronger" performance. "Faster" is not uniformly "stronger" performance but is considered as such in the routes used. There was an interesting interaction effect such that, in session 3, participants had faster rates of closure during the low workload session compared to high, whereas participants performed better in the high versus low workload condition during session 1. The worst performance was seen in the high workload condition during session 3, suggesting that fatigue exacerbated the effect of high workload. Alternatively, the worst performance in session 4 (compared to session 2), was seen in the low workload condition which may be an indication of a lack of engagement or motivation (participants were observed nodding off during this session). The key takeaway from these findings is that, when fatigued, performance can be negatively impacted regardless of level of workload and that "lower" level of workload may not be "better" than high workload levels.

Physiological measures.

Main effects of fatigue were seen with heart rate, HRV, pupil diameter, and the ratio of frontal theta to alpha activity. Additionally, an interaction effect between fatigue (session) and workload was evident on the ratio of frontal theta to alpha activity. No effects were seen with respect to workload, however, which suggests that the workload manipulations were not strong enough to elicit an effect in the analyzed event windows. Time-series analyses are likely needed to truly evaluate the patterns with respect to workload interjections.

Subjective measures.

All NASA-TLX subscales except the performance subscale resulted in significant main effects of fatigue, while the temporal demand score was the only one with a significant interaction. Comparisons are provided in Table 15.

Table 15. Summary Statistics and Pairwise Comparison Results

	Flight Session 1		Flight Session 2		Flight Session 3		Flight Session 4	
	Low Workload	High Workload	Low Workload	High Workload	Low Workload	High Workload	Low Workload	High Workload
	Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)	
Mental Demand	33.90^{a,b,c} (19.30)	37.80^{a,b,c} (13.90)	(51.10) ^{a} (18.00)	(56.70) ^{a} (19.20)	49.40^{b} (11.60)	56.10^{b} (14.50)	61.10^{c} (17.60)	57.20^{c} (18.40)
Physical Demand	18.90^{a,b,c} (9.61)	31.70^{a,b,c} (10.90)	44.40^{a} (19.60)	32.80^{a} (21.50)	42.80^{b} (12.30)	41.70^{b} (10.90)	52.20^{c} (19.10)	47.20^{c} (23.30)
Effort	28.30^{a,b,c} (15.40)	41.70 ^{a,b,c} (10.00)	60.60 ^{a} (17.20)	51.10 ^{a} (22.60)	53.30 ^{b} (21.20)	55.60 ^{b} (13.10)	62.80 ^{c} (15.40)	(57.20) ^{c} (17.70)
Temporal Demand	24.40 $^{a, b,c}$ (11.00)	43.30 (19.70)	48.90 ^{a} (25.50)	54.40 (24.20)	46.70 ^{b} (13.70)	40.60 (14.70)	56.70 ^{c} (21.40)	46.10 (13.40)
Performance	24.40 (18.80)	28.90 (16.90)	27.20 (21.10)	31.70 (15.60)	34.40 (18.10)	32.80 (18.60)	27.20 (11.50)	32.80 (16.40)
Frustration	11.10 ^{a} (7.41)	16.70 ^{a} (13.50)	18.90 (15.80)	27.20 (29.30)	27.80 (17.90)	20.00 (11.70)	35.60 ^{a} (22.40)	29.40 ^{a} (14.70)
Total Score	23.50 ^{a,b,c} (9.53)	33.30 ^{a,b,c} (11.00)	41.90 ^{a} (13.10)	42.30 ^{a} (16.90)	42.40 ^{b} (11.60)	41.10 ^{b} (7.16)	49.30 ^{c} (14.40)	45.00 ^{c} (12.40)

Note. Values with matching superscripts within a row significantly differed by $p < 0.05$.

The distinctions in workload manipulations (the methodology used to induce workload changes) likely contributed to the differences seen in subscores. Specifically, high workload conditions involved engine fuel issues and an engine fire that likely placed greater temporal demands on the participant compared to the low workload conditions where participants were to calculate estimated time of arrival to the airfield and experienced a hydraulics failure. The manipulation check was only done on two of the flight routes and a more detailed validation study to evaluate the routes used is necessary to further evaluate the efficiency of these interjections. These distinctions also highlight the importance of evaluation of NASA-TLX subscores and not just the total score. Similarly, when considering the total NASA-TLX scores in the first session, the values are fairly low compared to previous studies (e.g., Feltman et al., 2021).

Overall Discussion

The current study aimed to examine performance and physiological outcomes under varied levels of fatigue (34-hours of sustained wakefulness) and cognitive workload on a series of tasks spanning cognitive function and functional performance. All physiological measures were synchronized and tasks spanned visual, auditory, and cognitive domains. The first objective was to evaluate the main and interaction effects of fatigue and workload by task, which ultimately served to meet the second objective, to establish consistencies across the tasks in order to estimate robustness of the effects. When the task results are considered together, the strength of the effects can be discussed agnostic of task and performance outcome type. Considering that an OSM system will likely not be tailored to specific functional outcomes or even the same operational conditions, determining which parameters are most sensitive to detecting the effects of stressors will inform their utility in such a system.

Consistency of Fatigue and Workload Effects on Physiological Measures

Three main categories of physiological measurements (ECG, EEG, eye tracking) were taken across five tasks spanning cognitive, marksmanship, and aviation-specific, functional performance: Stroop (measuring selective attention), RVIP (measuring sustained attention), multimodal *n*-back (sequential memory), USAARL MATB (simulated aviation-like multi-tasking), and simulated flight (high-fidelity simulated flight routes). An overview of significant effects seen across tasks demonstrates four key findings: 1) fatigue effects were seen with a variety of physiological measures across all tasks, 2) pupil diameter showed sensitivity to fatigue in all tasks whereas at least one measure in each category showed sensitivity to fatigue in the majority of tasks, 3) workload effects were seen with at least one measure in each physiological measures category but only on two tasks (RVIP, MATB), and 4) interaction effects were largely non-existent with all physiological measures across all tasks.

This space is intentionally blank.

Table 16. Overview of Main and Interaction Effects on Physiological Measures Across All Tasks

Physiological Measure	Workload Effects (tasks)	Fatigue Effects (tasks)	Workload*Fatigue Effects (tasks)
		ECG	
Heart Rate	_	MATB *n*-back Simulated flight	_
Heart Rate Variability	MATB	MATB *n*-back Simulated flight	_
		EEG	
Global Alpha Activity	RVIP	RVIP MATB *n*-back	_
Global Beta Activity	_	Stroop	_
Global Delta Activity	_	*n*-back	_
Global Theta Activity	_	MATB	_
Parietal Alpha	RVIP MATB	RVIP MATB *n*-back	–
Frontal Theta: Alpha	–	*n*-back Simulated flight	Simulated flight
Frontal Beta: Alpha	–	MATB *n*-back	–
		Eye Tracking	
Blink Count	RVIP	RVIP *n*-back	–
Pupil Diameter	_	Stroop RVIP MATB *n*-back Simulated flight	_

Unsurprisingly, fatigue's effects on performance were seen on most tasks with multiple performance outcomes. Considerable research has shown the deleterious effects of fatigue in numerous domains, particularly in aviation (e.g., Bendak & Rashid, 2020). However, the extent to which fatigue impacted physiological measures and the consistency of such impacts suggests that development of OSM systems should consider a variety of measures. Rather, numerous physiological measures are sensitive to fatigue effects. Of particular note are the recurrent fatigue effects seen on pupil diameter. Recall, pupil diameter response to high workload (increased diameter) contrasts that to fatigue (decreased diameter) which could impact the accuracy of an OSM system. A multi-sensor suite will ultimately be needed to circumvent ambiguity based on one physiological measure. However, the findings here did not yield workload effects nor interaction effects with pupil diameter, which could suggest any of the following: 1) fatigue effects could simply mask workload effects, 2) the workload manipulations in this study were insufficient in the context of 34-hours of sustained wakefulness to yield significant effects, 3) workload effects are masked by environmental luminance, 4) temporal resolution in these analyses are too imprecise, or 5) specificity of pupil diameter as a proxy of workload is inadequate. Specificity has yet to be fully examined for any of these measures and will play a vital role in OSM effectiveness.

Workload effects on *physiological* measures were less robust than fatigue effects in this study but were consistently seen across two of the tasks (RVIP, USAARL MATB). Workload effects on *performance* were seen on three tasks (RVIP, USAARL MATB, *n*-back). It is worth noting that the workload manipulation checks on the simulated flights with respect to the perceived workload scores was positive but in comparison to similar studies, perceived workload was relatively low (e.g., Feltman et al., 2021). As candidate sensors and measures are down-selected for potential inclusion in an OSM system, it will be vital to consider not only sensitivity to changes in workload (perceived or objective) but also performance. This will be an important consideration for future studies as measures are chosen for inclusion and evaluated in terms of utility when developing OSM algorithms.

Interaction effects were largely non-existent with all physiological measures across all tasks; however, they were detected on performance of three tasks (RVIP, MATB, *n*-back). While not specific to the overarching objectives of this study, these interactions on performance highlight the complex nature of realistic environments where multiple stressors are present. For example, workload level can be impactful above and below comfortable thresholds and these thresholds are likely to shift when coupled with fatigue. This was highlighted in particular with the USAARL MATB results, suggesting that in some cases, the detrimental effects of fatigue may be exacerbated when facing a high workload (Matthews & Desmond, 2002). Further evaluation of motivation and engagement indices will be important steps in understanding the dynamic nature of optimal performance conditions.

Taken together, these findings reinforce the multidimensional nature of operator state assessment. Performance metrics emerge as the most reliable markers of workload and fatigue with partial overlap, subjective measures best capture workload, and physiological indicators appear to be most sensitive to fatigue, but not to workload, under these task conditions. The lack of consistent interaction between workload and fatigue may suggest that these factors primarily impact operator states through independent cognitive and physiological mechanisms, rather than additive or multiplicative effects. This independence aligns with neuroergonomic frameworks highlighting distinct pathways, such as acute attentional demands versus cumulative physiological depletion (Hancock & Warm, 1989; Matthews & Desmond, 2002). Practically, this independence may necessitate separate assessment and management strategies for workload and fatigue in operational settings, as each requires distinct interventions. These distinctions highlight the importance of integrating complementary data streams for adaptive automation and workload-sensitive systems, as reliance on any single indicator may fail to capture the nuanced and dynamic nature of cognitive state changes.

Limitations

The current study was extremely complex from a logistical execution standpoint and the research team chose to sacrifice some experimental control in order to provide a more realistic scenario. Thus, nutrition, wake-up time at the start of the extended visit, and light exposure (e.g., use of cell phones, access to windows) were not controlled. Having controlled these factors may have provided more explicit insight into the effects of interest. However, for an OSM system to be useful, the effects to which it is sensitive will have to be large and robust (as were the fatigue effects demonstrated here). Not only are a number of human factors and individual differences going to confound results and introduce error, but environmental factors will also compromise the integrity of the sensors and recording devices. The simulator in this study is full-motion and the sensors used here generally provided good quality data but the sensors were also monitored closely and reapplied/adjusted as needed, which will not be possible in the field. Also, time of day was manipulated in this study to a limited degree specifically with the simulated flights. A broader range of test times should be included in future efforts.

Another important consideration in terms of limitations is that the flight routes and manipulations were not directly evaluated and validated prior to inclusion in this study. Across all flight studies, performance outcomes are challenging and workload manipulations are heavily influenced by experience. While the routes included here were designed to include a variety of cockpit tasks and outcome measures, they may not have been sufficiently challenging to elicit workload effects. There is always a question of participant engagement in a simulator as well. The participants know that they are safe and likely are not as motivated to perform as optimally as in the aircraft. Lastly, workload and fatigue effects and interactions may have been masked in this report given the aggregate nature of the analyses. Efforts are underway to conduct time-series analyses and to evaluate predictive performance deficits using physiological measures.

Conclusions

Performance and physiological outcomes were measured under varied levels of fatigue and cognitive workload on a series of tasks spanning cognitive function and functional performance as part of the USAARL's ongoing OSM research efforts. The findings show robust and consistent effects of fatigue on physiological measures across the tasks, supporting the utility of the measures in OSM research and system design. Moreover, these results demonstrate that effective physiological monitoring can be achieved using a variety of physiological signals, as several, including pupil diameter, are demonstrated here to be particularly sensitive to fatigue. This suggests that a robust OSM system using multiple synchronized, physiological signals can be effectively implemented with commercial grade wearable sensors. Finally, the findings of this study demonstrate consistency of effects in a laboratory setting with limited experimental control (e.g., nutrition and light exposure were unregulated) and on functional performance outcomes. Thus, OSM research can justifiably advance to testing in more realistic and field environments.

References

Alhola, P., & Polo-Kantola, P. (2007). Sleep deprivation: Impact on cognitive performance. *Neuropsychiatric Disease and Treatment, 3*(5), 553–567.

American Psychiatric Association. (2000). *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.).

Ames, L. L., & George, E. J. (1993). *Revision and verification of a seven-point workload estimate scale* (No. AFFTCTIM9301). U.S. Air Force Flight Test Center.

Bakan, P. (1959). Extraversion-introversion and improvement in an auditory vigilance task. *British Journal of Psychology, 50*(4), 325–332.

Beck, A. T., Steer, R. A., Ball, R., & Ranieri, W. F. (1996). Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. *Journal of Personality Assessment, 67*(3), 588–597.

Bendak, S., & Rashid, H. S. (2020). Fatigue in aviation: A systematic review of the literature. *International Journal of Industrial Ergonomics, 76*, 102928.

Bowers, M. A., Christensen, J. C., & Eggemeier, F. T. (2014). The effects of workload transitions in a multitasking environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 220-224). SAGE Publications.

Brennan, S. D. (1992). *An experimental report on rating scale descriptor sets for the instantaneous self assessment (ISA) recorder* (Technical Memorandum [CAD5] 92017). DRA Maritime Command and Control Division.

Brouwer, A. M., Hogervorst, M. A., Van Erp, J. B., Heffelaar, T., Zimmerman, P. H., & Oostenveld, R. (2012). Estimating workload using EEG spectral power and ERPs in the n-back task. *Journal of Neural Engineering, 9*(4). http://dx.doi.org/10.1088/1741-2560/9/4/045008

Caldwell, J. L., Schroeder, V. M., Kunkle, C. L., & Stephenson, H. G. (2020). Differential effects of modafinil on performance of low-performing and high-performing individuals during total sleep deprivation. *Pharmacology Biochemistry and Behavior, 196*, 172968.

Chikhi, S., Matton, N., & Blanchet, S. (2022). EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology, 59*(6), e14009.

Curran, S. L., Andrykowski, M. A., & Studts, J. L. (1995). Short Form of the Profile of Mood States (POMS-SF): Psychometric information. *Psychological Assessment, 7*(1), 80–83. https://doi.org/10.1037/1040-3590.7.1.80

Delliaux, S., Delaforge, A., Deharo, J. C., & Chaumet, G. (2019). Mental workload alters heart rate variability, lowering non-linear dynamics. *Frontiers in Physiology, 10*, 565.

Duffy, M. J., & Feltman, K. A. (2023). *A systematic literature review of operator state detection using physiological measures* (USAARL-TECH-TR--2023-11). U.S. Aeromedical Research Laboratory.

Faure, V., Lobjois, R., & Benguigui, N. (2016). The effects of driving environment complexity and dual tasking on drivers' mental workload and eye blink behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, *40*, 78–90.

Feltman, K. A., Bernhardt, K. A., & Kelley, A. M. (2021). Measuring the domain specificity of workload using EEG: Auditory and visual domains in rotary-wing simulated flight. *Human Factors, 63*(7), 1271–1283. https://doi.org/10.1177/0018720820928626

Feltman, K. A., Randles, L. W., Goldie, C., & Bernhardt, K. A. (2022). Improving ecological validity in workload manipulation: Moving beyond a one-size-fits-all approach. *Applied Ergonomics, 102*, 103736. https://doi.org/10.1016/j.apergo.2022103736

Feltman, K. A., Vogl, J. F., McAtee, A., & Kelley, A. M. (2024). Measuring aviator workload using EEG: An individualized approach to workload manipulation. *Frontiers in Neuroergonomics*, *5*. https://doi.org/10.3389/fnrgo.2024.1397586

Feng, C., Wanyan, X., Yang, K., Zhuang, D., & Wu, X. (2018). A comprehensive prediction and evaluation method of pilot workload. *Technology and Health Care, 26*(1_suppl), 65–78.

Friedl, K. E. (2018). Military applications of soldier physiological monitoring. *Journal of Science and Medicine in Sport, 21*(11), 1147–1153.

Gaines, A. R., Morris, M. B., & Gunzelmann, G. (2020). Fatigue-related aviation mishaps. *Aerospace Medicine and Human Performance, 91*(5), 440–447.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*(5), 519–537. https://doi.org/10.1177/001872088903100503

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Human Mental Workload*. (pp. 139-183). North-Holland.

Heitmann, H., Zebhauser, P. T., Hohn, V. D., Henningsen, P., & Ploner, M. (2023). Resting-state EEG and MEG biomarkers of pathological fatigue–A transdiagnostic systematic review. *NeuroImage: Clinical, 39*, 103500.

Hussain, I., Young, S., & Park, S. J. (2021). Driving-induced neurological biomarkers in an advanced driver-assistance system. *Sensors, 21*(21), 6985.

Jap, B. T., Lal, S., Fischer, P., & Bekiaris, E. (2009). Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Systems with Applications, 36*(2), 2352–2359. https://doi.org/10.1016/j.eswa.2007.12.043

Jordan, C.S. (1992). *Experimental study of the effect of an instantaneous self assessment workload recorder on task performance* (Technical Memorandum [CAD5] 92017). DRA Maritime Command and Control Division.

Kaida, K., Takahashi, M., Åkerstedt, T., Nakata, A., Otsuka, Y., Haratani, T., & Fukasawa, K. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology, 117*(7), 1574–1581.

Kelley, A. M., McAtee, A., Duffy, M., & Feltman, K. A. (2023). *Evaluation of inter-subject variability in physiological metrics and workload perception: Implications for operator state monitoring* (USAARL-TECH-TR--2023-16). U.S. Aeromedical Research Laboratory.

Kessler, R. C., Adler, L., Ames, M., Demler, O., Faraone, S., Hiripi, E. V. A., Howes, M. J., Jin, R., Secnik, K., Spencer, T., Ustun, T. B., & Walters, E. E. (2005). The World Health Organization Adult ADHD Self-Report Scale (ASRS): A short screening scale for use in the general population. *Psychological Medicine, 35*(2), 245–256.

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. *Behavior Research Methods, 46*(1), 140–147.

Kilgore, W. D., Lipizzi, E. L., Grugle, N. L., Killgore, D. B., & Balkin, T. J. (2009). Handedness correlates with actigraphically measured sleep in a controlled environment. *Perceptual and Motor Skills, 109*(2), 395–400.

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology, 55*(4), 352–358. https://doi.org/10.1037/h0043688

Kurinec, C., Stenson, A., Hinson, J., Whitney, P., Honn, K., & Van Dongen, H. (2022). Examining premorbid cognitive abilities to predict task-specific, inter-individual differences in resilience to total sleep deprivation. *Sleep Medicine, 100*, S31–S32.

Lal, S. K. L., & Craig, A. (2002). Driver fatigue: Electroencephalography and psychological assessment. *Psychophysiology, 39*(3), 313–321. https://doi.org/10.1017/S0048577201393095

Liston, D. B., & Stone, L. S. (2014). Oculometric assessment of dynamic visual processing. *Journal of Vision, 14*(14), 12–12.

Lo, J. C., Groeger, J. A., Santhi, N., Arbon, E. L., Lazar, A. S., Hasan, S., Schantz, M. V., Archer, S. N., & Dijk, D. J. (2012). Effects of partial and acute total sleep deprivation on performance across cognitive domains, individuals and circadian phase. *PLoS ONE, 7*(9), e45987. https://doi.org/10.1371/journal.pone.0045987

Loh, S., Lamond, N., Dorrian, J., Roach, G., & Dawson, D. (2004). The validity of psychomotor vigilance tasks of less than 10-minute duration. *Behavior Research Methods, Instruments, & Computers, 36*(2), 339–346.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

Marquart, G., Cabrall, C., & De Winter, J. (2015). Review of eye-related measures of drivers' mental workload. *Procedia Manufacturing, 3*, 2854–2861.

Martínez-Cancino, D. P., Azpiroz-Leehan, J., & Jiménez-Angeles, L. (2015). The effects of sleep deprivation in working memory using the n-back task. In *Proceedings of the International Federation of Medical and Biological Engineering, 49*(1). https://doi.org/10.1007/978-3-319-13117-7_108

Matthews, G., & Desmond, P. A. (2002). Task-induced fatigue states and simulated driving performance. *The Quarterly Journal of Experimental Psychology Section A, 55*(2), 659–686. https://doi.org/10.1080/02724980143000527

McNair, D. M., Lorr, M., & Droppleman, L. F. (1981). *Profile of Mood States, revised (POMS)*. EdITS/Educational and Industrial Testing Service.

Mehler, B., Reimer, B., Coughlin, J. F., & Dusek, J. A. (2009). Impact of incremental increases in cognitive workload on physiological arousal and performance in young adult drivers. *Transportation Research Record, 2138*(1), 6–12. https://doi.org/10.3141/2138-02

Morad, Y., Lemberg, H., Yofe, N., & Dagan, Y. (2000). Pupillography as an objective indicator of fatigue. *Current Eye Research, 21*(1), 535–542.

Pagnotta, M., Jacobs, D. M., de Frutos, P. L., Rodríguez, R., Ibáñez-Gijón, J., & Travieso, D. (2022). Task difficulty and physiological measures of mental workload in air traffic control: A scoping review. *Ergonomics, 65*(8), 1095–1118.

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41*(1), 1–48. https://doi.org/10.1006/cogp.1999.0735

Sato, T., Yamani, Y., Liechty, M., & Chancey, E. T. (2020). Automation trust increases under high-workload multitasking scenarios involving risk. *Cognition, Technology & Work, 22*(2), 399–407

Soleimanloo, S.S., Wilkinson, V. E., Cori, J. M., Westlake, J., Stevens, B., Downey, L. A., Shiferaw, B. A., Rajaratnam, S. M. W., & Howard, M. E. (2019). Eye-blink parameters detect on-road track-driving impairment following severe sleep deprivation. *Journal of Clinical Sleep Medicine, 15*(9), 1271–1284.

Spielberger, C. D., & Reheiser, E. C. (2004). Measuring anxiety, anger, depression, and curiosity as emotional states and personality traits with the STAI, STAXI and STPI. In M. J. Hilsenroth & D. L. Segal (Eds.), *Comprehensive Handbook of Psychological Assessment, Vol. 2. Personality Assessment* (pp. 70–86). John Wiley & Sons, Inc.

Stuiver, A., & Mulder, B. (2014). Cardiovascular state changes in simulated work environments. *Frontiers in Neuroscience, 8*, 399.

Tsai, Y. F., Viirre, E., Strychacz, C., Chase, B., & Jung, T. P. (2007). Task performance and eye activity: Predicting behavior relating to cognitive workload. *Aviation, Space, and Environmental Medicine, 78*(5), B176–B185.

Unni, A., Ihme, K., Jipp, M., & Rieger, J. W. (2017). Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study. *Frontiers in Human Neuroscience, 11*, 167.

van Egmond, L. T., Meth, E. M., Bukhari, S., Engström, J., Ilemosoglou, M., Keller, J. A., Zhou, S., Schlioth, H. B., & Benedict, C. (2022). How sleep-deprived people see and evaluate others' faces: An experimental study. *Nature and Science of Sleep, 14*, 867–876.

Van Orden, K. F., Limbert, W., Makeig, S., & Jung, T. P. (2001). Eye activity correlates of workload during a visuospatial memory task. *Human Factors, 43*(1), 111–121.

Vogl, J., McCurry, C. D., Bommer, S., & Atchley, J. A. (2024). The United States Army Aeromedical Research Laboratory Multi-Attribute Task Battery. *Frontiers in Neuroergonomics, 5*, 1435588.

Vogl, J., O'Brien, K., & St. Onge, P. (2025). One size does not fit all: A support vector machine exploration of multiclass cognitive state classifications using physiological measures. *Frontiers in Neuroergonomics, 6*. https://doi.org/10.3389/fnrgo.2025.1566431

Wanyan, X., Zhuang, D., & Zhang, H. (2014). Improving pilot mental workload evaluation with combined measures. *Bio-medical Materials and Engineering, 24*(6), 2283–2290.

Wickens, C. D. (2002). Multiple resources and performance prediction. *Theoretical Issues in Ergonomics Science, 3*(2), 159–177. https://doi.org/10.1080/14639220210123806

Wilson, G. F., & Eggemeier, F. T. (2020). Psychophysiological assessment of workload in multi-task environments. In D. Damos (Ed.), *Multiple Task Performance* (pp. 329-360). CRC Press.

Wingelaar-Jagt, Y. Q., Wingelaar, T. T., Riedel, W. J., & Ramaekers, J. G. (2021). Fatigue in aviation: Safety risks, preventive strategies and pharmacological interventions. *Frontiers in Physiology, 12*, 712628.

Wu, Y., Miwa, T., & Uchida, M. (2017). Using physiological signals to measure operator's mental workload in shipping – an engine room simulator study. *Journal of Marine Engineering & Technology, 16*(2), 61–69. doi:10.1080/20464177.2016.1275496

Appendix A. Acronyms and Abbreviations

ASRS	Attention Deficit/Hyperactivity Disorder Self Report Scale
BDI	Beck’s Depression Inventory
BPM	Beats per Minute
ECG	Electrocardiography
EEG	Electroencephalography
EMG	Electromyography
EST	Engagement Skills Trainer
EOG	Electrooculography
HRV	Heart Rate Variability
Hz	Hertz
ISA	Instantaneous Self-Assessment of Workload
KSS	Karolinska Sleepiness Scale
LMM	Linear Mixed-Effects Model
LSL	Lab Streaming Layer
MATB	Multi-Attribute Task Battery
NASA	National Aeronautics and Space Administration
NASA-TLX	NASA-Task Load Index
OSM	Operator State Monitoring
POMS	Profile of Mood States
PSD	Power Spectral Density
PVT	Psychomotor Vigilance Task
RMSSD	Root Mean Square of Successive Differences
RVIP	Rapid Visual Information Processing Task
SD	Standard Deviation
STAI	State-Trait Anxiety Inventory
USAARL	U.S. Army Aeromedical Research Laboratory

Appendix B. In-house Developed Questionnaires

Participant # ____________

Date ___________________

Demographics and Personal History Questionnaire

Do you have a current DD-2292 (cleared for flight duties)? .... YES / NO

If you answered "**No**" to this question, **please stop here**, and return the questionnaire to personnel.

NICOTINE USE

Do you use tobacco products containing nicotine? .... YES / NO
– including e-cigarettes, patches/gum, cigars, smokeless

If you answered "**Yes**" to this question, **please stop here**, and return the questionnaire to personnel.

PSYCHIATRIC CONDITIONS

Have you ever been diagnosed with a psychiatric illness? YES / NO
Have you ever been hospitalized for a psychiatric or psychological reason? YES / NO

If you answered "**Yes**" to this question, **please stop here**, and return the questionnaire to personnel.

Age ____________ **Gender:** _____Male ______Female

Ethnicity (if you prefer not to disclose, leave blank):

_____ African American (non-Hispanic) _____ Asian/Pacific Islander

_____ Caucasian (non-Hispanic) _____ Latino or Hispanic

_____ Native American _____ Other: ________________________

Handedness: _____Right _____Left _____Ambidextrous

Highest Level of Education Completed:

High School Some College Associate Bachelor's Graduate

Do you have 20/20 uncorrected vision? Yes / No If No do you wear:

_______Glasses _______Contacts _______Both _______Neither

Do you have any other visual impairment(s), such as color blindness? YES / NO

If you answered “Yes” to the question above, please state the impairment(s) below:

____

____

ALCOHOL USE

On average, how many beers do you drink per week? ___________
On average, how many glasses of wine do you drink per week? ___________
On average, how many mixed drinks or shots do you drink per week? ___________

MEDICATION USE

Are you currently taking any prescription drugs? .... YES / NO
If Yes, what? ____________________________
For what? ____________________________

Are you currently taking any over-the-counter drugs? .... YES / NO
If Yes, what? ____________________________
For what?____________________________

Are you currently taking any vitamins or dietary supplements? .... YES / NO
If Yes, what? ____________________________
For what?____________________________

Do you use caffeine products or over the counter supplements: YES / NO

If yes, please indicate the type and frequency product is used. On an average day of use, Please include number of times per day.

Product	**Frequency (e.g., occasionally, monthly, weekly, daily)**	**Enter average amount consumed on a day of use**
Caffeine products (e.g., NOS [16 oz], Red Bull [8.46 oz], Coffee [8 oz], Soda/Soft Drink [12 oz], or gum)		
Protein powder		
Multi-vitamin		
Other (Specify): __________		

Number of deployments: ________________

Total Length of Deployments Combined (months): ________________

Time in military aviation career (months): ________________

Please list the type(s) of aircraft you have flown **during your current military career** and the approximate hours in each type:

Type	**Hours (actual)**	**Hours (simulator)**	**Years Flown**
1.			
2.			
3.			
4.			
5.			

What was the approximate amount of total **on-the controls** time (in hours) you received in the:

Last year_________________

Last 90 days________________

Last 30 days________________

To be filled out by Study Physician and Technician

STUDY PHYSICIAN____

TECHNICIAN ADMINISTERING ____

MEDICAL HISTORY

Have you ever had, or do you now have: (Y=YES N=NO D=DON'T KNOW)

IF YES, obtain (1) Type; (2) Mo/Yr of occurrence; (3) Is it current?

Y N D .... Attention Deficit Disorder?

Y N D .... Neurocognitive Disorders (e.g., cognitive disabilities)?

Y N D .... Psychiatric Disorders (e.g., depression, anxiety, bipolar)?

Y N D .... Do you have normal, or corrected to normal, vision (or any eye trouble)?

Y N D .... Do you wear glasses or contacts?

Y N D .... Do you have normal hearing?

Y N D .... Rapid or Pounding heartbeat? (circle which)

Y N D .... High or Low blood pressure? (circle which)

Y N D .... Sleep disorders (e.g., sleep apnea)?

Do you have a current DD-2292 (cleared for flight duties)? .... **YES** **NO**

Have you consumed alcohol within the last 24 hours?................. **YES** **NO**

Pregnancy test administered and negative? .... **YES** **NO** **NA**

Reported more than 600 mg of caffeine per day? .... **YES** **NO**

Nicotine user?................. **YES** **NO**

Current medications affecting drowsiness, cognitive function, or physiological parameters of the study?................. **YES** **NO**

Eligible? (Study Physician initials)________

Participant #____________

Date__________________

Sleep Habits Questionnaire

Please respond to the below questions regarding your sleeping patterns.

Typical bedtime: _____________ weekday ________________ weekend

Typical awakening time: _____________ weekday ________________ weekend

Typical hours in bed: ________________ hours

Typical hours of sleep: _______________ hours

Typical amount of time it takes to fall asleep: _____________ hours

Typical number of awakenings per night: _________________

Time it takes to fall back asleep after wakening: ____________

My sleep pattern is regular: ______ yes ________ no

I awaken early in the morning still tired but unable to return to sleep:

______ yes ________ no

Appendix C. USAARL MATB – Description of Workload Levels.

The individually-tailored task demand was controlled through modification of a pre-generated parameter file that defined the frequency and magnitude of task events. The demand levels were designed to span the spectrum of task demand, from extreme underload to extreme overload. Each participant experienced five demand level trials that were scaled to their normal demand level derived through the training process. During each session, the participant would receive a counterbalanced trial of their normal demand level (established during the initial visit practice session), their normal demand level ± 3 demand levels (slightly under the defined just noticeable difference identified in Vogl et al., 2025), and extreme high and low demand levels designed to be over and under loading, respectively. This yielded a total of 20 trials across the 4 session times (i.e., 5 trials per session)

Appendix D. Flight Routes and Data Collection Forms

SLEEP STUDY - Route 1 (IC 1, PPC 1)

Subject:

RP:

SET UP:

ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: Route 1; INI: Empty 15700, Crew 600

EVENT MARK WHEN TRANSFERRING CONTROLS

EVENT	P*	ACTION	Response	U/S	REMARKS
1	RP	Calculate ETA, Fuel Req'd	10:00, 140 lbs		Event Mark Stop/Start
		Notes:			
2	RP	ENG 1 FUEL PRESS LOW	IAW FRC		
		Notes:			
Pass controls if ENG OUT					
3	RP	ENG 1 FIRE	IAW FRC		WILL NOT OCCUR IF ENG 1 OUT
		Notes:			
Pass controls if ENG OUT					
4	Subject	Roll-On Landing	IAW ATM		DO NOT PROMPT
		Notes:			

EVALUATION CRITERIA

1	ETE: +/- 1 minute; FUEL: +/- 100 lbs, Time to Calculate
2	Eng #1 in XFD before ENG OUT (30 seconds)
3	Confirm, Request A/S 80 KIAS or less, EPCL - OFF, FIRE T - PULL, MAIN/RES (Underline)
4	Min SE A/S < Ground Speed < 60 KIAS, +/- 5 deg RWY alignment, descent <540 fpm

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You hoave just completed a single-ship passenger move to LZ Kingfisher. We are at the LZ preparing to return to our home airfield.

Flight profile is 1500’ MSL and 100 KIAS unless otherwise directed/required by conditions.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

SLEEP STUDY - Route A (IC 2, PPC 2)

Subject:

RP:

SET UP:

ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: Route A; INI: Empty 15700, Crew 600

EVENT MARK WHEN TRANSFERRING CONTROLS

EVENT	P*	ACTION	Response	U/S	REMARKS
1	Subject	Traffic Avoidance	30 right, 600'		
		Notes:			
2	Subject	Resume Own Nav	Resume		
		Notes:			
Take controls at AACP1					
3	RP	Calc ETA to Airfield	6:00		Cue at AACP1, Event Mark Start/Stop
		Notes:			
4	RP	#1 HYD FAIL NO BUP	IAW FRC		Force continued flight to the airfield
		Notes:			
Pass controls after EP Resolved					
5	Subject	Roll-On Landing	IAW ATM, FRC		DO NOT PROMPT
		Notes:			

EVALUATION CRITERIA

1	Descend to 600', turn 30 deg right
2	Return to/hold 1500' and 100 KIAS
3	ETE: +/- 1 minute, Time to Calculate
4	BUP ON (Underline), Urgency (Possible)
5	40 KIAS < Speed < 60GS, +/- 5 deg RWY alignment, descent <540 fpm, no aerodynamic braking

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You are conducting a single ship training flight to an un-towered airfield to conduct traffic patterns.

Flight profile after take-off is 1500’ MSL and 100 KIAS unless otherwise directed/required by conditions.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

SLEEP STUDY - Route 2 (IC 2, PPC 2)

Subject:

RP:

SET UP:

ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: Route 2; INI: Empty 15700, Crew 600

EVENT MARK WHEN TRANSFERRING CONTROLS

EVENT	P*	ACTION	Response	U/S	REMARKS
1	Subject	Gunfire; 3 o'clock	Terrain Mask		
		Notes:			
Take controls when EP starts					
2	RP	Compressor Stall	IAW FRC		Slow back for EP resolution
		Notes:			
Pass controls when EP resolved					
3	Subject	Radar Threat	Terrain Mask		**STAY OVER THE LAKE**
		Notes:			
4	Subject	Normal Approach	IAW ATM		
		Notes:			

EVALUATION CRITERIA

1	200' AHO or below, obstacle avoidance, general flight plan routing
2	Reduce power, EPCL reduce (IDLE), EPCL FLY
3	200' AHO or below, obstacle avoidance, general flight plan routing
4	Obstacle avoidance, appropriate alignment, appropriate rate of closure, descent <300fpm

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You are conducting a single ship passenger movement from PZ Bluebird to LZ Eagle. This scenario will take place in an enemy-threat environment. There have been unconfirmed reports of small recon teams with radar weapon systems along our route.

Flight profile after take-off is 1500’ MSL and 100 KIAS unless otherwise directed/required by conditions. **DO NOT** discuss specifics of 2800/2900 series tasks during these scenarios.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

SLEEP STUDY - Route B (IC 2, PPC 2)

Subject:

RP:

SET UP:

ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: Route B; INI: Empty 15700, Crew 600

EVENT MARK WHEN TRANSFERRING CONTROLS

EVENT	P*	ACTION	Response	U/S	REMARKS
1	RP	Weigh (HVR), Hover Checks	18300, IAW ATM		0 KTS WIND, Event Mark Stop/Start
		Notes:			
Pass controls after climb out					
2	Subject	Missile Launch	Terrain Mask		
		Notes:			
3	Subject	Redirect LZ	Adjust Flt Profile		WPT 033
		Notes:			
4	Subject	Normal Approach	IAW ATM		
		Notes:			

EVALUATION CRITERIA

1	GWT +/-500 lbs; verify GWT, verify no exceedence of GWT, verify power for all manuevers
2	200' AHO or below, obstacle avoidance, general flight plan routing
3	Turn to waypoint as directed by RP, continue terrain flight
4	Obstacle avoidance, appropriate alignment, appropriate rate of closure, descent <300fpm

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You are conducting a three-ship passenger move from PZ Crow to LZ Skylark. This scenario will take place in an enemy-threat environment. There have been confirmed reports of enemies conducting small-arms attacks on aircraft traversing the area.

Flight profile after take-off is 1500’ MSL and 100 KIAS unless otherwise directed/required by conditions. DO NOT discuss specifics of 2800/2900 series tasks during these scenarios.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

SLEEP STUDY - Route 3 (IC 4, PPC 3)

Subject:

RP:

SET UP:

ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: No Route; INI: Empty 15700, Crew 600, Cargo: 1500

EVENT MARK WHEN TRANSFERRING CONTROLS

EVENT	P*	ACTION	Response	U/S	REMARKS
0	RP	Left Traffic Pattern	N/A		JVMF Message Appears 2+00
		Notes:			
1	RP	JVMF, Fuel Req'd, ETA	Fuel, 9:03		Event Mark Stop/Start, Hospital WP34
		Notes:			
Pass controls before takeoff					
2	Subject	Ridgeline Landing	IAW ATM		
		Notes:			
3	Subject	STAB AUTO MODE FAIL	IAW FRC		
		Notes:			
4	Subject	Routine Landing	IAW ATM		DO NOT PROMPT STAB
		Notes:			

EVALUATION CRITERIA

0	
1	ETE: +/- 1 minute; FUEL: +/- 100 lbs, Time to Calculate
2	Obstacle avoidance, appropriate alignment, appropriate rate of closure, descent <300fpm
3	CMSSUS - As required (Underline), Auto Reset, 0 above 40 and full down below 40
4	Obstacle avoidance, app alignment, app rate of closure, descent <360fpm, **STAB full down**

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You are on the 1st-Up MEDEVAC duty crew, conducting a training flight while on shift. You will be executing traffic patterns at your home stagefield.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

SLEEP STUDY - Route C (IC 1, PPC 1)					
Subject:					
RP:					
SET UP:					
ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: N/A; INI: Empty 15700, Crew 600					
EVENT MARK WHEN TRANSFERRING CONTROLS					
EVENT	**P***	**ACTION**	**Response**	**U/S**	**REMARKS**
1	RP	Calc Fuel, ETA	9:54, 136		JVMF 00+30, Event Mark Stop/Start
		Notes:			
Pass controls before takeoff					
2	Subject	Routine Landing	IAW ATM		add 1400 lb cargo after WOW at LZ
		Notes:			
3	RP	Dynamic Update/HVR Check	IAW ATM		
		Notes:			
Pass controls after climb out					
4	Subject	Routine Landing	IAW ATM		
		Notes:			

EVALUATION CRITERIA	
1	ETE: +/- 1 minute; FUEL: +/- 100 lbs, Time to Calculate
2	Obstacle avoidance, appropriate alignment, appropriate rate of closure, descent <300fpm
3	MTA, MGWT, GWT, HVR Q; verify GWT, no exceedence of GWT, power for all manuevers
4	Obstacle avoidance, appropriate alignment, appropriate rate of closure, descent <300fpm

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You are flying a supply route and have just dropped a group of passengers. You are preparing to RTB from your current location.

Flight profile after take-off is 1500’ MSL and 100 KIAS unless otherwise directed/required by conditions.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

SLEEP STUDY - Route 4 (IC 3, PPC 2)

Subject:

RP:

SET UP:

ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: Route 4; INI: Empty 15700, Crew 600

EVENT MARK WHEN TRANSFERRING CONTROLS

EVENT	P*	ACTION	Response	U/S	REMARKS
1	Subject	ITO	IAW ATM		
		Notes:			
2	Subject	Frequency Change	Acknowledge call		
		Notes:			
3	Subject	Altitude Change	Adjust ALT		
		Notes:			
Pass controls after climb					
4	RP	DUAL GEN FAIL	IAW FRC		Anticipate unusual att. recovery
		Notes:			
5	RP	Frequency Change	Adjust FMS		
		Notes:			

END ROUTE AT WP*

EVALUATION CRITERIA

1	400 AGL prior to turn, announced ROC (min 230-260 fpm), announced climb airspeed
2	
3	Climb to 5000'
4	Request A/S 80 KIAS, GEN RESET, GEN OFF, APU ON **OR** EMER APU ON
5	Adjust freq as req'd

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You are conducting an IFR training flight. You have received the following clearance: cleared to Sonoma County as filed (direct), climb and maintain 4000, departure frequency 121.35, squawk 0421. You are about to conduct a diverse departure from your current location.

Flight profile after take-off is 4000’ MSL and 100 KIAS unless otherwise directed/required by conditions.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

SLEEP STUDY - Route D (IC 3, PPC 2)					
Subject:					
RP:					
SET UP:					
ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: Route D; INI: Empty 15700, Crew 600					
EVENT MARK WHEN TRANSFERRING CONTROLS					
EVENT	**P***	**ACTION**	**Response**	**U/S**	**REMARKS**
1	Subject	ITO	IAW ATM		
		Notes:			
Pass controls after climb out					
2	RP	Level Off Checks			
		Notes:			
3	Subject	Unusual Att Recovery	IAW ATM		Attain during level off checks
		Notes:			
Emergency transfer after establishing unusual attitude					
4	Subject	Attitude Change	Adjust as req'd		
		Notes:			
END ROUTE AT WP**					

EVALUATION CRITERIA	
1	400 AGL prior to turn, announced ROC (min 230-260 fpm), announced climb airspeed
2	AHTTA
3	Climb to 5000'
4	Request A/S 80 KIAS, GEN RESET, GEN OFF, APU ON **OR** EMER APU ON

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You are conducting an IFR training flight. You have received the following clearance: cleared to Mountain View direct, climb and maintain 3000, departure frequency 135.025, squawk 4612. You are about to conduct a diverse departure from your current location.

Flight profile after take-off is 3000’ MSL and 100 KIAS unless otherwise directed/required by conditions.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

SLEEP STUDY - Route 5 (IC 2, PPC 2)

Subject:

RP:

SET UP:

ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: Route 5; INI: Empty 15700, Crew 600
CARGO HOOK: NORM, COCKPIT, ARMED

EVENT MARK WHEN TRANSFERRING CONTROLS

EVENT	P*	ACTION	Response	U/S	REMARKS
1	RP	Allowable weight calc	4288		want OGE HVR capable, Mark Stop/Start
		Notes:			
Pass controls before take off					
2	Subject	Attach Load	IAW ATM		15-20 feet over WP 36, IO trigger
		Notes:			
3	Subject	Redirect for ROZ	Adjust as req'd		MAGNUS WP 32
		Notes:			
Take controls after redirect call					
4	RP	Calculate Endurance	Time Remaining		Mark Stop/Start
		Notes:			
Pass controls after calculations					
5	Subject	Drop Load	IAW ATM		15-20 feet AGL
		Notes:			

EVALUATION CRITERIA

1	GWT +/- 500 lbs, time to calculate
2	Stable hover 15-20 feet above load for 15 sec prior to "LOAD HOOK"
3	
4	Fuel Burn, BINGO requirement, calculate fuel and time margins, time to calculate
5	Stable hover 15-20 feet above LZ for 15 sec prior to releasing load (P* triggered)

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You are conducting a sling load operation moving a 3,400lb blivet from your airfield to LZ Raven. All cargo hook checks have been completed and the switches are configured as required. You will pick up the load at 15' RADALT. The scenario will begin prior to picking up your cargo.

Flight profile after take-off is 1500’ MSL and 100 KIAS unless otherwise directed/required by conditions. Reminder that with a sling load you are limited to 30 degree bank angles and 140 KIAS VNE.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

SLEEP STUDY - Route E (IC 4, PPC 3)

Subject:

RP:

SET UP:

ICS: All pins down, 1 & 3-7 max volume, 2 1/2 volume; FPN: N/A; INI: Empty 15700, Crew 600, Cargo 1500

EVENT MARK WHEN TRANSFERRING CONTROLS

EVENT	P*	ACTION	Response	U/S	REMARKS
1	RP	Calculate Power, HVR Check	##, IAW ATM		
		Notes:			
Pass controls before take off					
2	Subject	Terrain Flight	IAW ATM		
		Notes:			
3	Subject	Hoist (50 AGL)	IAW ATM		
		Notes:			
4	Subject	Loiter			
		Notes:			
5	Subject	Hoist (50 AGL)	IAW ATM		
		Notes:			

EVALUATION CRITERIA

1	MTA, MGWT, GWT, HVR Q (**Stop/Start Mark**); verify GWT, no exceedence of GWT, power for all manuevers
2	200' AHO or below, obstacle avoidance, general flight plan routing
3	Hover over hoist WP at 50 AGL for 15 seconds, HDG +/- 10 deg, Alt +/- 5', Drift NTE 10'
4	
5	Hover over hoist WP at 50 AGL for 15 seconds, HDG +/- 10 deg, Alt +/- 5', Drift NTE 10'

BRIEF

You are a PI flying under the callsign FORGE25. You will execute all pilot on the controls/off the controls duties and any actions directed by the RP.

You are a MEDEVAC crew responding to a hoist mission in the nearby mountains. All hoist checks have been completed. You will drop off your medic, loiter, and then return to recover both them and the patient. Conditions dictate that the hover height will be 50’.

Flight profile after take-off is 1500’ MSL and 100 KIAS unless otherwise directed/required by conditions.

Just a reminder – pay attention for the audio cues to provide your workload ratings. Questions?

Appendix E. Physiological Data Processing Procedures and Descriptions

EEG Data Processing

EEG recordings were processed in fixed-duration windows to compute workload-related spectral metrics and generate topographic scalp maps. Processing was implemented in MATLAB using the EEGLAB toolbox functions readlocs and topoplot for channel montage handling and scalp plotting. The pipeline proceeded as follows:

1. Channel Labeling and Montage Mapping
    a. EEG channel data were provided in the order:
        i. {'Fp1','F7','F8','T4','T6','T5','T3','Fp2','O1','P3','Pz','F3','Fz','F4','C4','P4','POz','C3','Cz','O2','EOG'}.
    b. Standard 10-20 system electrode coordinates were loaded from the EEGLAB standard-10-5-cap385.elp file using readlocs. Channels from the input dataset were matched to their standard positions with a custom label-matching function, ensuring accurate spatial mapping for topographic plotting.
2. Window Segmentation
    a. The continuous EEG signal was divided into non-overlapping time windows of variable duration, based on events recorded in the flight simulator. For each window, data segments from all channels were extracted for feature computation.
3. Spectral Analysis
    a. Power spectral density (PSD) estimates were computed using Welch's method (pwelch) with a 2-second Hamming window and no overlap. Four canonical frequency bands were defined:
        i. Delta: 1-4 Hz
        ii. Theta: 4-8 Hz
        iii. Alpha: 8-13 Hz
        iv. Beta: 13-30 Hz
    b. Mean band power was computed for each channel by averaging PSD values across the corresponding frequency bins.
4. Workload-Related Metrics
    a. The following features were derived from channel-specific band powers:
        i. Frontal Theta/Alpha Ratio: Mean theta-to-alpha ratio across frontal electrodes (Fp1, Fp2, F3, F4, Fz).
        ii. Frontal Beta/Alpha Ratio: Mean beta-to-alpha ratio across the same frontal electrodes.
        iii. Parietal Alpha Power: Mean alpha power across parietal electrodes (P3, P4, Pz).
        iv. Global Band Powers: Mean beta, alpha, theta, and delta power across all available channels.
5. Topographic Mapping (Figure E1)
    a. Two spatial maps were prepared for each analysis window:
        i. Alpha Map: Log-transformed alpha power across channels.

        ii. Theta/Alpha Ratio Map: Channel-wise ratio of theta to alpha power.
    b. When enabled, both maps were plotted using topoplot (EEGLAB) with interpolated scalp shading and consistent color scaling and saved as PNG files labeled by time window.
6. Output Structuring
    a. For each time window, computed metrics were stored in a structured array. Each element of the array corresponded to one analysis window and contained all numeric metrics along with the time boundaries of that window. This method allows for consistent quantification of EEG spectral features related to cognitive workload, while also providing spatial visualization through EEGLAB's topographic plotting capabilities.

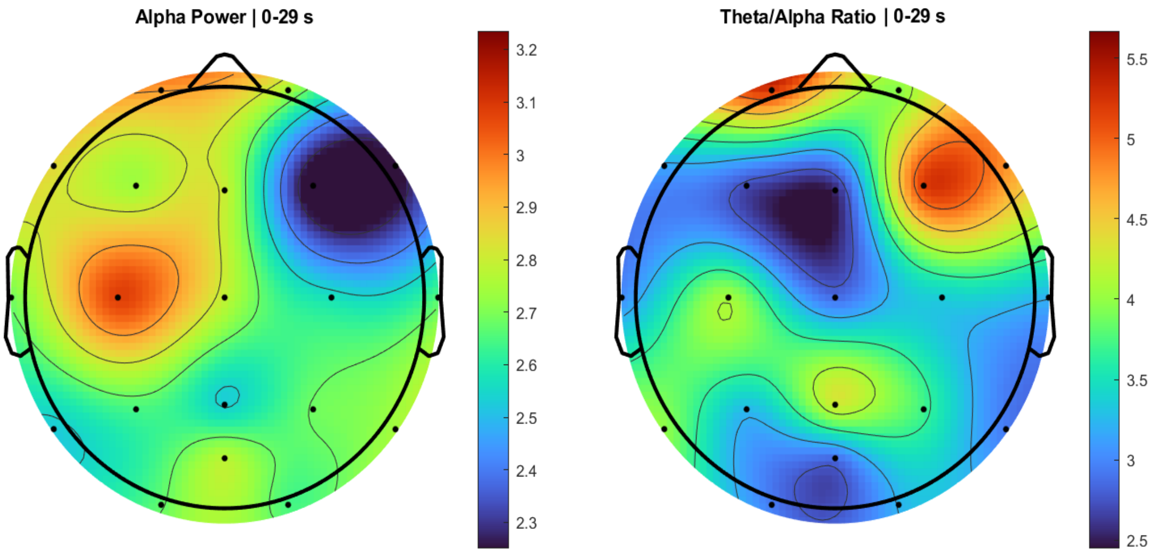

Figure E1. Example of EEG topographic plots.

ECG Data Processing

ECG recordings were processed in fixed-duration time windows (e.g., corresponding to flight simulator events, during task performance) to extract heart rate (HR), HRV, and morphological metrics. The analysis was performed using a custom MATLAB function with the following steps:

1. Window Segmentation
    a. The continuous ECG signal was divided into analysis windows corresponding to the events occurring in the flight simulator.
2. Bandpass Filtering
    a. Within each window, the raw ECG was zero-phase Butterworth bandpass filtered (1-40 Hz, 4th order) to attenuate baseline drift and high-frequency noise while preserving the QRS complex morphology.
3. Artifact Detection and Masking

    a. Signal artifacts were identified using a 2-second moving standard deviation. Samples exceeding a threshold of five times the median moving standard deviation were marked as artifacts and set to NaN to exclude them from peak detection and metric calculations.
4. R-Peak Detection
    a. R-peaks were identified in the cleaned ECG signal using a peak-finding algorithm (findpeaks) with a minimum inter-peak distance of 0.3 s (~200 BPM upper limit) and a minimum height threshold of 200 μV (a consistent value surpassed by peaks recorded with the Polar H10 system). This ensured robust detection of QRS complexes while avoiding false positives from noise.
5. RR Interval Cleaning
    a. Inter-beat (RR) intervals were calculated from consecutive R-peaks and constrained to a physiologic range of 0.3-2.0 s (30-200 BPM). RR intervals outside this range were discarded to reduce the impact of spurious detections.
6. Heart Rate and Time-Domain HRV Metrics
    a. From the cleaned RR intervals, the following metrics were computed:
        i. HR (BPM): 60 divided by the mean RR interval.
        ii. SDNN: Standard deviation of RR intervals, reflecting overall HRV.
        iii. RMSSD: Root mean square of successive differences in RR intervals, reflecting short-term HRV.
        iv. pNN50: Percentage of successive RR interval differences exceeding 50 ms.
7. Frequency-Domain HRV Metric
    a. Welch's power spectral density estimate was applied to the RR series to compute the ratio of low-frequency (LF: 0.04-0.15 Hz) to high-frequency (HF: 0.15-0.40 Hz) power. This LF/HF ratio was used as an index of parasympathetic and sympathetic system balance.
8. QRS Amplitude Measurement
    a. For each detected R-peak, the QRS amplitude was calculated as the difference between the R-peak amplitude and the preceding trough. The mean QRS amplitude across the window was reported.
9. Handling Insufficient Data
    a. If a window contained fewer than two valid RR intervals after cleaning, all metrics for that window were set to NaN to indicate insufficient data for reliable calculation.
10. Visualization and Output (Figure E2)
    a. When enabled, two-panel plots were generated for each window: (a) raw vs. filtered ECG and (b) cleaned ECG with R-peak markers. Figures were automatically saved as PNG files named according to the participant ID, window index, and time range.
    b. This processing pipeline provides a consistent, windowed approach to ECG analysis, incorporating both time- and frequency-domain HRV

metrics as well as morphological features, while mitigating the influence of noise and artifacts through filtering and artifact masking.

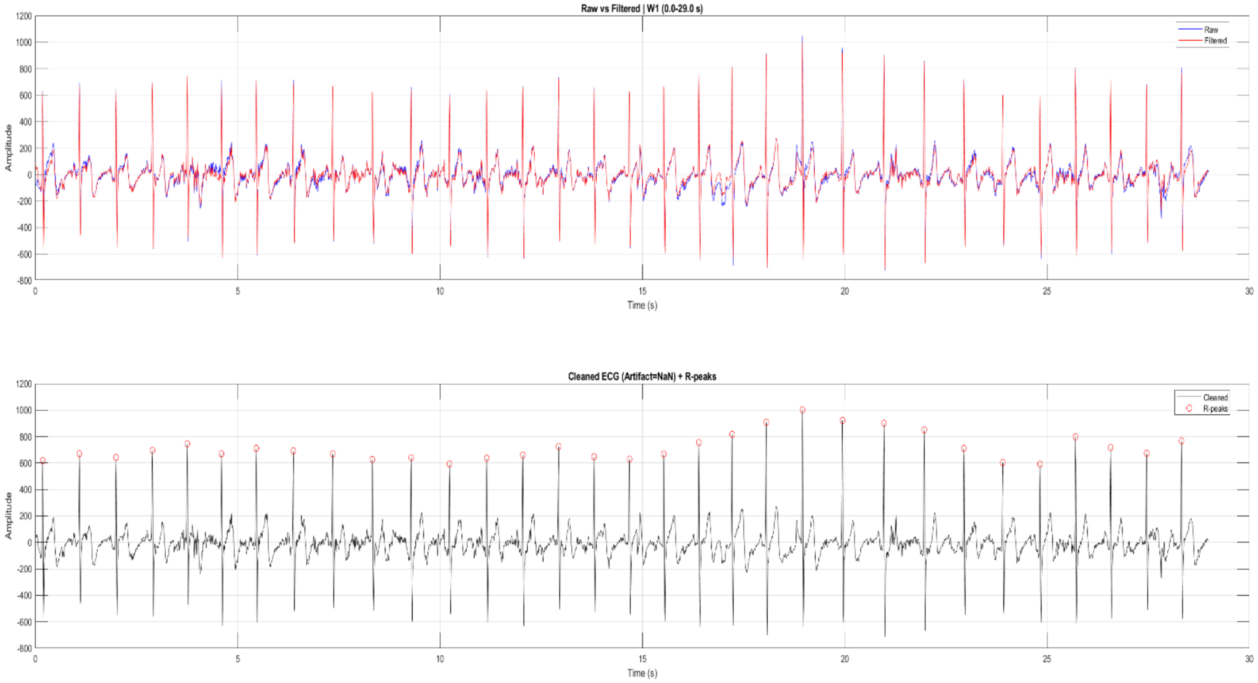

Figure E2. Example of two-panel plots displaying raw vs. filtered ECG and cleaned ECG with R-peak markers.

Eye Tracking and Pupillometry Data Processing

Pupil diameter data from the left and right eyes were first preprocessed prior to automated analysis. All values greater than 8 mm or less than 1 mm were removed to exclude physiologically implausible measurements. Additionally, any samples with a data quality score below 0.1 (a deliberately liberal threshold) were removed to eliminate low-confidence points from subsequent analysis.

After initial cleaning, pupil data were processed using a custom MATLAB function with the following steps:

1. Time Alignment and Reference Adjustment
    a. Timestamps for left eye, right eye, and electrooculography (EOG) signals were converted to column vectors and shifted so that all signals shared a common zero-time reference corresponding to the earliest recorded sample.
2. Down-sampling
    a. To ensure computational efficiency and reduce high-frequency noise, pupil signals sampled above the nominal sample rate for an eye tracking system (e.g., 60 Hz for Smart Eye and 120 Hz for Pupil Labs) were down-sampled to approximately 50 Hz.

3. Median Filtering of Short Spikes
    a. Short-duration spikes (~200 ms) were suppressed using a median filter, reducing the influence of transient noise while preserving true pupil responses.
4. EOG-Based Blink Detection
    a. Blinks were identified using bandpass-filtered (0.5-8 Hz) EOG signals and z-score thresholding. Blink onset and offset times were determined by merging temporally close blink events and anchoring to peak deflections. The number, duration, and amplitude of blinks were computed from the EOG signal.
5. Time Span Synchronization
    a. Left and right pupil time series were truncated to a common time span to ensure consistent analysis across both eyes.
6. Mapping Blinks to Pupil Data
    a. Blink periods detected in the EOG signal were mapped to corresponding pupil timestamps for each eye.
7. Combining Eyes into a Single Signal
    a. Left and right pupil data were interpolated onto a common time grid at the pupil sampling rate and averaged, omitting missing values.
8. Blink-Related Data Removal and Gap Filling
    a. Data ±100 ms before and ±200 ms after each blink were removed (set to NaN) to eliminate blink-induced artifacts. Short gaps were filled using linear interpolation, provided the gap duration was below a specified maximum.
9. Correlation-Based Blink Refinement
    a. A second pass identified additional blink artifacts by correlating EOG deflections and pupil constriction patterns (EOG z-score > 1 and pupil z-score < -1). Affected data segments were removed and short gaps interpolated.
10. Velocity Filtering and Baseline Floor
    a. Samples with absolute pupil velocity greater than 1 mm/s were removed, as were filtered values below a 0.5 mm baseline floor. Short gaps were interpolated.
11. Frequency-Domain Smoothing
    a. Pupil data were low-pass filtered at 4 Hz (zero-phase Butterworth filter) to suppress high-frequency noise.
12. Metric Computation
    a. Windowed pupil metrics were computed, including mean diameter, peak diameter, standard deviation, blink count, blink rate, spectral unrest power (0.05-1 Hz), and the ratio of low-frequency to high-frequency power (0.05-1 Hz vs. 1-4 Hz). EOG-derived blink metrics were also reported.
13. Visualization and Output (Figure E3)
    a. If enabled, time series plots of the left, right, and averaged pupil signals were generated, overlaid with blink periods. Processed EOG traces with blink onsets were also plotted. Figures were saved to the designated output directory.

b. This method integrates both spatial (binocular combination) and temporal (artifact removal and smoothing) processing steps, with EOG-driven blink detection ensuring robust artifact rejection. The resulting cleaned and interpolated pupil time series enable reliable computation of physiologically meaningful metrics.

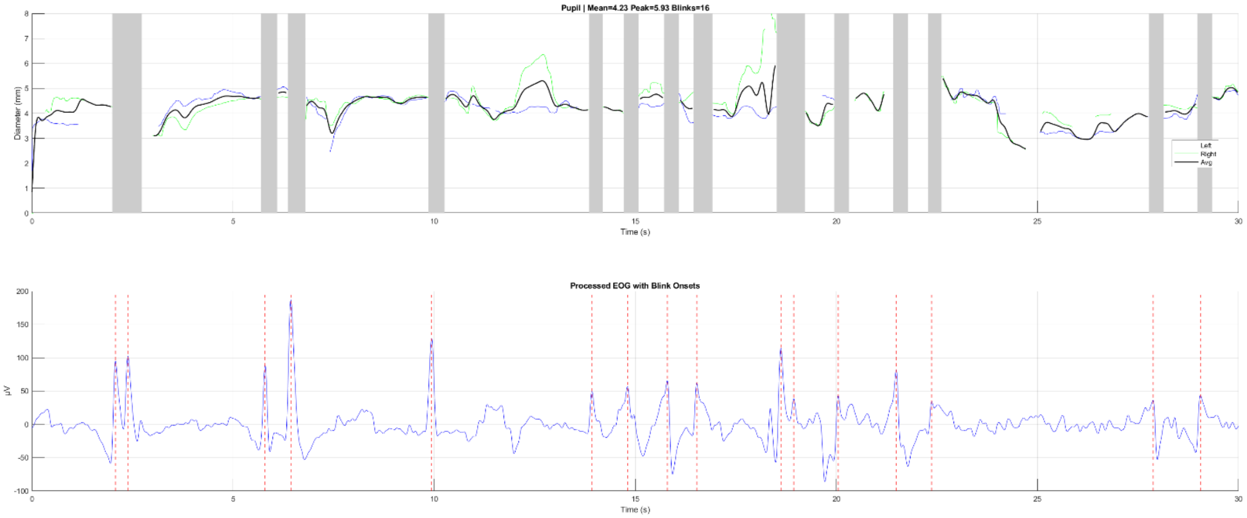

Figure E3. Example of EOG plots.

U.S. ARMY AEROMEDICAL RESEARCH LABORATORY

FORT RUCKER, ALABAMA

Optimizing
HUMAN PROTECTION
AND PERFORMANCE
since 1962

All of USAARL’s science and technical informational documents are available for download from the Defense Technical Information Center.
https://discover.dtic.mil/results/?q=USAARL

U.S. ARMY

T2COM

MRDC