



AFRL-RH-WP-TR-2023-0072

**COGNITIVE & PHYSIOLOGIC PERFORMANCE (CPP):
EXPERIMENT 2**

Hana Ulman, Tim McEwen, John Flach, Kaleb Embaugh

Mile Two, LLC

Robert Folker

PatchPlus Consulting

JULY 2023

Final Report

Distribution Statement A: Approved for public release, distribution is unlimited.

See additional restrictions described on inside pages

**AIR FORCE RESEARCH LABORATORY
711TH HUMAN PERFORMANCE WING,
AIRMAN SYSTEMS DIRECTORATE,
WRIGHT-PATTERSON AIR FORCE BASE, OH 45433
AIR FORCE MATERIEL COMMAND
UNITED STATES AIR FORCE**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Air Force Research Laboratory Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

AFRL-RH-WP-TR-2023-0072 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

DEL
ROSSA.ANTONINO.MI
CHELE.1538356391

Digitally signed by DEL
ROSSA.ANTONINO.MICHELE.1
538356391
Date: 2023.12.20 11:27:50
-05'00'

ANTONINO DEL ROSSA, 1 Lt, USAF
Program Manager
Applied Cognitive Neuroscience Research Area
Performance Optimization Branch
Airman Biosciences Division

WILLIAMS.LOGAN.
ANDREW.12735976
34

Digitally signed by
WILLIAMS.LOGAN.ANDREW.127
3597634
Date: 2024.01.02 13:46:46 -05'00'

LOGAN A. WILLIAMS, DR-III, PhD
Human Performance Product Area Lead
Operational Product Section
Product Development Branch
Airman Biosciences Division

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ORGANIZATION.				
REPORT DATE July 2023	REPORT TYPE Final	3. DATES COVERED		
		START DATE	END DATE	
TITLE AND SUBTITLE Cognitive & Physiologic Performance (CPP): Experiment 2				
5a. CONTRACT NUMBER FA8650-21-C-6271		5b. GRANT NUMBER		5c. PROGRAM ELEMENT NUMBER
5d. PROJECT NUMBER		5e. TASK NUMBER		5f. WORK UNIT NUMBER H12C
6. AUTHOR(S) ¹ Hana Ulman / ¹ Tim McEwen / ¹ John Flach / ² Robert Folker / ¹ Kaleb Embaugh				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) ¹ Mile Two, LLC 601 E. 3rd Street Dayton, OH 45432			8. PERFORMING ORGANIZATION REPORT NUMBER	
² PatchPlus Consulting 2 North Main Street, Suite 203B Medford, NJ 08055				
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory 711th Human Performance Wing Airman Systems Directorate Wright-Patterson Air Force Base, OH 45433			10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-RH-WP-TR-2023-0072
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution Statement A. Approved for public release. Distribution is unlimited.				
13. SUPPLEMENTARY NOTES AFRL-2023-5650, cleared 19 December 2023				
14. ABSTRACT This Interim Technical Report summarizes the second experiment conducted as part of the Cognitive & Physiologic Performance (CPP) research program. The primary objective of the CPP research program is to explore the feasibility of using real time physiological measures to gauge the state of operators working in Distributed Common Ground Systems (DCGS). This is motivated by reports of high levels of stress as well as physical and emotional fatigue, and concerns that these may impact operational readiness and performance				
15. SUBJECT TERMS				
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF	18. NUMBER OF PAGES
REPORT U	ABSTRACT U	THIS PAGE U	ABSTRACT SAR	80
19a. NAME OF RESPONSIBLE PERSON Antonino Del Rossa			19b. PHONE NUMBER (Include area code) N/A	

TABLE OF CONTENTS

Section	Page
LIST OF FIGURES.....	iv
LIST OF TABLES	v
1.0 EXECUTIVE SUMMARY	1
2.0 INTRODUCTION.....	2
3.0 METHODS.....	4
3.1 Participants.....	4
3.2 Materials and Equipment	4
3.2.1. Hardware	4
3.2.2. JADE software	4
3.2.3. Self-report instruments.....	6
3.2.4. Laboratory setup.....	6
3.3 Experimental Design	7
3.3.1. Fatigue.....	7
3.3.2. Task difficulty	8
3.3.3. Teaming level.....	8
3.3.4. Time	9
3.4 Dependent Variables	9
3.4.1. Long- and short-term Fatigue.....	9
3.4.2. Intelligence Task Performance.....	10
3.4.3. Time spent viewing Intelligence documents.....	11
3.4.4. Well-being.....	11
3.4.5. Subjective workload.....	12
3.4.6. Difficulty Survey.....	12
3.4.7. Heart rate and heart rate variability.....	13
3.4.8. Blood pressure.....	14
3.4.9. Eye metrics.....	14
3.5 Procedure.....	16
4.0 RESULTS & DISCUSSION.....	17
4.1 Well-being.....	17
4.1.1. Maslach Burnout Inventory.....	18
4.1.2. State-Trait Anxiety Inventory	19

4.1.3. Emotion Regulation Questionnaire	20
4.1.4. Baseline blood pressure.....	21
4.1.5. Well-being summary	22
4.2 Fatigue Manipulation	23
4.2.1. Epworth Sleepiness Scale (ESS).....	23
4.2.2. Karolinska Sleepiness Scale (KSS).....	23
4.2.3. Psychomotor Vigilance Task (PVT)	24
4.2.4. Fatigue summary	25
4.3 Difficulty Manipulation.....	25
4.3.1. Modified NASA TLX results	25
4.3.2. Global team workload	26
4.3.3. Difficulty questionnaire results	26
4.3.4. Difficulty summary	27
4.4 Intelligence Task Performance.....	28
4.4.1. Correct responses	28
4.4.2. SME ratings.....	28
4.4.3. Intelligence task performance summary	29
4.5 Heart Rate Variability	30
4.5.1. High & low frequency normalized unit metrics.....	30
4.5.2. Heart rate variability summary.....	31
4.6 Eye Tracking Metrics.....	31
4.6.1. Averaged across the entire scenario	31
4.6.2. Eye tracking metrics by Intelligence document type	32
4.6.3. Eye metrics summary	34
4.7 Physiological Metrics by Performance Level	35
4.7.1. HRV × Performance level.....	35
4.7.2. Eye Metrics × Performance level.....	36
4.7.3. Physiological measures × Performance level summary.....	37
4.8 Time Spent Viewing Documents in JADE	38
4.9 Team Survey & Case Study	39
5.0 GENERAL DISCUSSION & CONCLUSION.....	41
5.1 Fatigue.....	41
5.2 Difficulty	41

5.3	Teaming.....	42
6.0	RECOMMENDATIONS	44
7.0	REFERENCES.....	45
	APPENDIX A. PARTICIPANT CHARACTERISTICS	49
	APPENDIX B. JADE SOFTWARE	56
	APPENDIX C. TEAM SURVEY	65
	APPENDIX D. COMPARISON BETWEEN THE TOBII PRO GLASSES 2 & PUPIL CORE EYE TRACKING GLASSES	68
	LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS.....	69

LIST OF FIGURES

Figure	Page
Figure 1. Laboratory setup for individual and two-person group conditions.....	7
Figure 2. Diagram illustrating the procedure and sequence of activities.	16
Figure 3. The percentages of respondents who reported concerning levels on the three components of the MBI (i.e., 20 or above on the Exhaustion and Cynicism scales; and 12 or below on the Professional Efficacy scale).....	19
Figure 4. Mean S-Anxiety (left) and mean T-Anxiety scores (right) as a function of Fatigue and Team condition..	20
Figure 5. ERQ Cognitive Reappraisal and Expressive Suppression subscale means..	21
Figure 6. Mean baseline systolic blood pressures measured prior to the experimental sessions as a function of Team, Fatigue, and Difficulty..	22
Figure 7. Mean baseline diastolic blood pressures measured prior to the experimental sessions as a function of Team, Fatigue, and Difficulty..	22
Figure 8. Mean ESS scores as a function of Fatigue and Team conditions.	23
Figure 9. Mean pre- and post-test KSS sleepiness scores as a function of Team, Difficulty, and Fatigue.....	24
Figure 10. Mean pre- and post-test PVT response times (in ms) as a function of Team, Fatigue, and Difficulty.....	25
Figure 11. Mean composite TLX ratings as a function of Difficulty, Fatigue, and Team variables.	26
Figure 12. Participants' ratings of subjective scenario difficulty, plotted as a function of Difficulty/Scenario (Califon Hard vs MVO Easy) and Individual vs. Team conditions.....	27
Figure 13. Mean Report ratings as a function of Fatigue and Team.	29
Figure 14. Baseline Normalized HF and LF (in nu; top and bottom, respectively) averaged throughout the entire scenario across Fatigue, Difficulty, and Team.	30
Figure 15. Mean fixation duration (in ms) as a function of Fatigue, Team, and Difficulty.....	32
Figure 16. Differences in HF and LF (in nu) between Fatigue conditions as a function of Performance. $HF\Delta_F$ (left) and $LF\Delta_F$ (right).	36
Figure 17. Differences in HF and LF (in nu) between Task Difficulty conditions as a function of Performance. $HF\Delta_D$ (left) and $LF\Delta_D$ (right)..	36
Figure 18. Differences in blink rate across rank as a function of Fatigue (top) and Difficulty (bottom).....	37
Figure 19. Mean time (in seconds) spent viewing Noise documents as a function of Fatigue and Difficulty.....	39
Figure 20. The distribution of responses to the Team Questionnaire items.....	39
Figure 21. (Top) HR across the entire scenario for a cohesive team, (Bottom) HR across the entire scenario for a non-cohesive team.....	40

LIST OF TABLES

Table	Page
Table 1. Instrument names, sources, status, and method of administration.	6
Table 2. The independent variables and brief descriptions of their manipulation.	7
Table 3. ICD 203 Standards, the type of measure, and the number of raters conducting the assessment in this experiment.	10
Table 4. Summary of the three self-reported questionnaires and associated constructs.	12
Table 5. Identification and categorization of heart rate variability metrics related to cognitive load and fatigue.	13
Table 6. Identification and categorization of eye tracking metrics related to cognitive load and fatigue.	15
Table 7. Test-retest correlation coefficients (Pearson's r) for the three well-being surveys.	18
Table 8. Intraclass Correlations (ICC) of SME ratings of Performance on the Process measures.	28
Table 9. Eye tracking metrics broken down by Intelligence document type that were significant or approached significance.	34
Table 10. Labels and descriptions related to HF and LF differences for Fatigue and Task Difficulty	35

Acknowledgements

We would like to thank Dr. Rena Nicholas, and Dr. Regina Buccello-Stout for their oversight and technical contributions. We would also like to thank and acknowledge Dr. Logan Williams and Dr. Scott Galster for conceptual contributions and support. Additionally, Kati Walker assisted in collecting data. John Fay, Lindsay Huling and Brandon Canfield provided important technical support for the design and programming of the Joint Activity Design Evaluation (JADE) platform and for the associated data acquisition functions. We are also very lucky to have had the cooperation of many dedicated servicemembers from the various Intelligence, Surveillance, and Reconnaissance (ISR) and ART (Air Force Resiliency Team) units who helped make this research a success. To preserve their identities, the words of David Foster Wallace seem appropriate: “The best way [we] can think of to show [our] appreciations to these men and women is to decline to thank them by name.”

1.0 EXECUTIVE SUMMARY

Previous research has observed that Intelligence operators often operate under stressful conditions, such as dealing with complex decisions during long duty shifts while experiencing disrupted sleep cycles that can induce chronic fatigue. The Cognitive and Physiologic Performance (CPP) program was undertaken to determine if physiological sensors might be a practical method for detecting when individual analysts were stressed to a degree that could compromise performance on their Intelligence tasks. To accomplish this an easily transportable Intelligence simulation environment was developed and taken to an operational site to collect data from Intelligence operators performing the simulation task as either individuals or as two-person teams. Scenarios representing different levels of difficulty were developed and used on either rest days or following a duty shift to capture different levels of fatigue. While performing the simulations the Intelligence operators wore unobtrusive sensors to collect heart and eye activity. The results showed that multiple physiological measures were sensitive to both task difficulty and fatigue – including, cardiovascular measures and eye metrics. There were several interesting findings in the team analyses, including one case where an argument between team members was clearly detected in the heart activity measures. It seems clear that heart and eye metrics are sensitive to variations in cognitive demands and fatigue and may have promise as measures of team functioning. However, further research is needed to determine and validate the practicality for use under operational conditions within the Distributed Common Ground System (DCGS).

2.0 INTRODUCTION

Realizing the importance of human performance to achieving the operational functions of DCGS sites, the goal of the CPP research program is to develop methods appropriate for assessing and monitoring the health and performance capacity of individuals and teams of analysts within a DCGS. This is a prerequisite to the design of any intervention intended to mitigate potential impacts on the functional capacity of future DCGS systems.

Over the last decade there has been increasing awareness and concerns about the psychological health of the Air Force's 'remote warriors.' The remote warriors are typically stationed safely outside the theater of operation, and are actively engaged in combat (e.g., drone pilots) or Intelligence activities (e.g., full motion video analysts). Though these remote warriors are not under direct threat from an enemy, they are often involved in making and observing the consequences of potentially life and death decisions. Chappelle, Prince & Goodman (2019) summarized the results of psychological health assessments of operators in DCGS sites from 2011 and 2013:

"Intel operators were at an elevated risk for emotional, social, and behavioral health problems. Previous research identified approximately 26–27% of the DCGS Intel operator population experienced high levels of emotional exhaustion, and 14–16% reported medically significant psychological distress impairing both social and occupational functioning. Additionally, this population reported an array of negative health behaviors and outcomes, largely attributed to work related stress. These included but were not limited to elevated alcohol use (16–20%), problematic caffeine use (34–35%), inadequate amounts of physical exercise (38–40%), increasing prescription (14–17%) and over the counter medication (10–17%) usage, problematic headaches (28–37%) and musculoskeletal pain (22–30%), work-related anxiety (9–14%) and depression (4–11%), and insufficient sleep (55–63%)." (p. 451-452)

According to Bryant-Lees, Prince, Goodman, Chappelle & Thompson (2021), Air Force Special Operations Command (AFSOC) pilots and Air Combat Command (ACC) Intelligence operators were also noted as "being at elevated risk for negative psychological outcomes compared to their fellow operators." (p. 793). They continued to explain that "ACC Intelligence operators report experiencing higher levels of stress related to low manning and the nature of work and significantly higher level of psychological stress than AFSOC Intelligence operators." (p. 793 - 794)

It is well substantiated in the literature that cognitive workload and fatigue are associated with changes in autonomic nervous system (ANS) activity. Additionally, there have been recent advances in the design of physiologic monitoring technologies to make them relatively non-invasive with respect to impacting performance while remaining sensitive to ANS activity. Therefore, the focus of this study was to explore the impact of physiological measures in a synthetic task environment (STE) that closely mimicked the operational tasks and duties required of Intelligence operators as a function of Fatigue state, Task Difficulty (cognitive workload), and Teaming. The specific question addressed in the experimental research is:

Can physiological monitoring be used to anticipate potential performance deficits in both individual and team problem solving that might result from fatigue, cognitive overload, and/or stress?

The study reported in this document is the second of a two-part research effort. Experiment 1 and its associated report (AFRL-RH-WP-TR-2023-0019) evaluated the validity of the STE at administering scenarios representative of the work domain and verified the feasibility of wearable physiological devices that could be adopted for use in an operational setting. Results from the first study demonstrated that the manipulation of Task Difficulty was successful via measures of subjective workload, difficulty ratings, and performance outcomes. Feedback surrounding the realism of the task scenario and the engagement of the experiment was overwhelmingly positive. Although there were no significant differences in heart rate variability or eye tracking measures during the task itself during the first study, there were interesting findings related to performance, where top performers had different patterns of response whilst completing more difficult scenarios than their middle and bottom performing counterparts.

Findings from the first study, in conjunction with the conclusions drawn from Chappelle et al. (2019) and Bryant-Lees et al. (2021), motivated the design and structure of the present study. Namely, we aimed to understand whether the current DCGS population continues to experience high levels of burnout and psychological stress, the degree and impact of a work shift on an operators' fatigue state and subsequent performance, effects of cognitive workload on operators' functional state and physiology, and the impacts of collaboration. This was largely achieved via the manipulation of three variables: (1) Fatigue State (High Fatigue vs. Low Fatigue); (2) Task Difficulty (Hard scenario vs. Easy scenario); and (3) Teaming (Individual vs. Pair).

The inclusion of Teaming was motivated by the changes within the DCGS from a platform-centric Process, Exploitation, and Dissemination (PED) structure to an Analysis Exploitation Teams (AET) organization. The shift from Intelligence work in permissive environments to contested or near-peer engagements motivated the DCGS to evolve into the current AET model which is platform agnostic and asks for teams of analysts from different backgrounds to solve increasingly complex problems. In the AET structure each analyst is trained "to think holistically and historically, leveraging, cumulative knowledge, experience, and context to comprehensively and creatively address problems" (Borukhovich & Morton, 2020). The expectation is that the AET format will encourage a "swarm mindset" with increased synergy within the DCGS towards collaborative problem solving. However, there is also the potential that the need for real-time collaboration may introduce other stresses on operators. It is within this context that we examine both individual and team problem solving in this study.

3.0 METHODS

The overall hypothesis guiding the experiment is that reductions in the quality of analyst or team performance as a function of increases in cognitive load, fatigue, and stress will be associated with physiological changes that can be monitored prior to or during work shifts. Fatigue, Task Difficulty, and Team Size were manipulated in this experiment. Additional performance, behavioral, and subjective report measures were included to verify the manipulations of cognitive load and fatigue, and to provide converging evidence relative to associations between analyst states and capabilities and physiological measures.

To ensure that the results from this study would generalize to operational contexts, great care was taken to create an experimental task that is representative of intelligence work in a DGCS. To this end, the JADE software platform was developed and used in this study as an STE. The purpose of an STE is to enable researchers to balance the tradeoff between experimental control (i.e., internal validity) and domain fidelity (i.e., external validity). In this case, JADE allowed researchers to present simulated tasks/scenarios that were representative of work typically done by DCGS intelligence analysts while simultaneously allowing rich observations of analyst behaviors (e.g., responses to prompts, time spent viewing scenario documents).

3.1 Participants

This study was approved by the Air Force Research Laboratory (AFRL) Institutional Review Board (Protocol number FWR20230048H). Data were collected from 40 Active-Duty Intelligence operators stationed within multiple ISR military units located at a United States Air Force base. Ninety percent of participants reported having previous experience working within an AET format (as opposed to PED only). See Appendix A for more information about participant characteristics.

During each session, participants were offered snacks and non-caffeinated beverages. Caffeinated beverages were, however, offered at the conclusion of their data collection session. Participants were compensated with Visa/Mastercard gift cards of \$100 for each session. No participants withdrew from the study, so all received the total possible compensation of \$200.

3.2 Materials and Equipment

3.2.1. Hardware

The experiments employed the following hardware equipment: Microsoft Windows-based laptop computers, Polar H10 (Polar Electro, Kempele, Finland) electrocardiogram (ECG) wearable devices, and Pupil Core eye tracking glasses (Pupil Labs GmbH, Berlin, Germany). Blood pressure was taken using the Omron 7 Series[®] Upper Arm Blood Pressure Monitor (Omron Healthcare Co., Ltd. Kyoto, Japan). Those in the Paired Team condition were also fitted with Movo LV1-UC omnidirectional Lavalier microphones (MOVO Photo, Los Angeles, CA) to capture audio from conversations between the team members. Please note that the recorded audio data is to be analyzed by AFRL as part of a separate, but related research effort and is not covered in this report.

3.2.2. JADE software

The JADE software was created to serve as the STE by administering simulated Intelligence scenarios and tools that replicate key functions of Intelligence analysis work while gathering data from a variety of inputs. For more information about JADE, please see Appendix B.

Once participants signed into JADE, they were presented with several modules. A member of the research team guided participants through these modules during the research session. Each module was responsible for administering some aspect of the research study, including surveys, a tutorial on how to use the JADE software to complete the task, and the experimental scenario.

The survey module included the Demographics and the Intake questionnaires. The Demographics questionnaire inquired into participant backgrounds and experience. These responses were not anticipated to change across sessions. The Intake questionnaire, on the other hand, focused primarily on transient characteristics that could vary between sessions, such as whether they were wearing their contacts if they needed corrected vision, whether they had taken a prescribed stimulant (e.g., Adderall), their recent caffeine consumption, and their recent sleep behavior.

The tutorial module used example Intelligence documents but was otherwise identical to the user interface (UI) for the experimental scenario. The tutorial was directed by a member of the research team and included a walkthrough of the various tools and interactions within the UI to complete the task. The tutorial was not repeated during a participant's final session, but participants were able to request help with the software at any point.

The experimental scenario module presented participants with an Intelligence database, document viewer, notepad, and report generation tool. The Intelligence database served as a repository for various types of Intelligence documents that were needed for participants to effectively solve the scenario. Several background documents were made available for viewing from the onset of the scenario from within the Intelligence database while the remaining Intelligence documents were unable to be opened until those first initial background documents were reviewed. The background documents contained a Request for Information (RFI) and tasked the participants with identifying a site of interest corresponding to each fictional scenario (see Section 3.3.2 for details). The remaining Intelligence documents served to provide additional details that were either diagnostic or confounding in relation to correctly answering the RFI. All documents were in the Intelligence database. Participants were able to take notes about the various documents using the notepad functionality.

Each time a participant opened an Intelligence document for the first time, they were required to fill out two forms: (1) noteworthy findings and (2) impact to working hypothesis. In the case of the Paired Teaming condition, if an individual's partner viewed the document, they had the option to toggle a slider, to "Yes, my partner is reading this document" to avoid duplicating work that was already being captured on the partner's computer. Upon subsequent viewings of the same document, participants were not required to fill out these textboxes unless they believed their mental model had changed. The intent of having these inputs was to track the cognitive processes as analysts worked through the scenario.

Once all the documents had been reviewed, participants were able to move onto writing the final Intelligence report (hereafter referred to as 'Report'). On this screen participants had access to a non-editable version of their notes and were able to view any of the documents in the Intelligence database. Participants could navigate to this page at any time after all the Intelligence documents in the Intelligence database had been opened. However, participants were encouraged to begin drafting their IR when they reached 45 minutes (i.e., 17 minutes before the 52 minutes allotted for the task had expired) to help ensure that there was adequate time.

3.2.3. Self-report instruments

Several self-report instruments were used in this research, including a custom Team Survey as well as six published and validated instruments. These are included in Table 1:

Table 1. Instrument names, sources, status, and method of administration.

Instrument Name	Source	Status	Admin Method
Teaming Survey	Mile Two	Custom	Paper
National Aeronautics and Space Administration Task Load Index (NASA TLX)	Hart, 2006	Published	JADE
Karolinska Sleepiness Scale (KSS)	Åkerstedt, 2012	Published	JADE
Epworth Sleepiness Scale (ESS)	Johns, 1991	Published	JADE
Maslach Burnout Inventory (MBI)	Maslach et al., 2018	Published	Paper
State Trait Anxiety Inventory (STAI)	Spielberger, 2020	Published	Paper
Emotional Regulation Questionnaire (ERQ)	Gross & John, 2003	Published	Paper

The Team Survey was only administered to those in the Paired Team condition and participants were asked to complete this form independently. It asked participants to describe and rate aspects of their experience working with a partner on the scenario.

Descriptions of the NASA TLX, KSS, ESS, MBI, STAI, and ERQ are described throughout Section 0.

3.2.4. Laboratory setup

The experiment was administered using the JADE software platform presented on consumer laptop computers running Microsoft Windows 11 Professional (Microsoft, Redmond, Washington). Participants each had access to a computer provided to them by the research team. Those in the Paired Team condition did not share computers until it came time to write their joint IR. The Pupil Core eye trackers and microphones used wires to connect to the computers. The Polar H10s connected wirelessly to the computers using Bluetooth Low Energy technology. The Omron blood pressure cuff was a standalone unit and could be moved around the room as needed.

The laboratory itself was a temporarily repurposed ‘business center’ for Airmen that contained several computers and a printer that were not used in the experiment. A whiteboard was used as a divider between two desks to provide as much separation and privacy as possible when gathering data from multiple participants and groups simultaneously. The business center was closed for normal use during the two-week data collection period to minimize interruptions. See Figure 1.



Figure 1. Laboratory setup for individual and two-person group conditions

3.3 Experimental Design

In this section, we describe our primary independent variables and the mechanisms for manipulating them. Table 2 provides a quick overview. Longer descriptions follow thereafter.

Table 2. The independent variables and brief descriptions of their manipulation.

Independent Variable	Manipulation
Fatigue level	Participants participated in both a Low Fatigue state (i.e., had an 8+ hour break since their last shift) and in a High Fatigue state (i.e., within an hour of ending a shift).
Task Difficulty	The Hard task scenario required participants to read more Intelligence documents, consider more alternatives, and combine two pieces of a clue spread across two documents to interpret properly. The Easy scenario required less reading, fewer options, and did not divide clues amongst the documents.
Teaming level	Participants were either assigned to complete the task working as an Individual or in a Paired Team.
Time	Repeated administrations of measurement at multiple points during the experimental session.

3.3.1. Fatigue

Fatigue was manipulated using two different testing periods based on participant work schedule. All participants experience both levels of Fatigue over the course of two separate research sessions. The High Fatigue condition was administered at the end of a work shift when fatigue level was expected to be highest due to the time and effort expended during a shift. The Low Fatigue condition was administered prior to a work shift or on a non-working day, with the expectation that the analysts would have been relatively more rested. Specifically, a minimum of eight hours were required to have elapsed since the end of their most recent shift. This manipulation was coordinated with the DCGS leadership to ensure that participation in the study would have minimal impact to readiness and operational performance.

3.3.2. Task difficulty

Task Difficulty was manipulated using two different scenarios with two Difficulty levels, Easy Michigan vs Ohio scenario (MVO) and Hard (Califon). These two scenarios were designed by coauthor, RF, from PatchPlus Consulting, who is a former USAF Intelligence Officer (Lt Col) and has 32 years of experience operating, managing, and training within multi-domain planning, command & control (C2), ISR, and DCGS. The specific scenarios used were selected based on findings from a prior study (AFRL-RH-WP-TR-2023-0019) that indicated that these specific scenarios would best manipulate Difficulty as intended.

Each Intelligence document was categorized into one of three types: (1) guidance; (2) diagnostic; and (3) noise.

1. Guidance documents were made available to the participant(s) from the onset, such as RFI, Threat to Joint Operations (TTJO), and Target Systems Analysis (TSA), and contained background information and instructions.
2. Diagnostic documents contained relevant information that provided crucial details needed to identify the correct solution for each scenario.
3. Noise documents contained information that wasn't relevant to the scenario or was misleading.

In the Hard scenario there were a total of three guidance, seven noise, and two diagnostic documents; the Easy scenario contained two of each type. The Easy scenario challenged participants to determine the most likely port of origin from which surface-to-air-missile (SAM) components were being illicitly shipped between two fictitious nation-states. The Hard scenario required analysts to assess the most likely site to which a notional adversary had relocated a SAM weapon system to fill a gap in their air defense coverage.

Both scenarios contained several background documents that outlined the scenario and the task instructions. Specifically, the scenarios included the following: (1) TTJO or TSA; (2) Air Operations Directive (AOD) or Fragmentary Order (FRAGO); and (3) RFI. Each scenario also included operationally relevant Intelligence products that replicated a variety of Intelligence report types from a range of Intelligence specialties, such as human Intelligence (HUMINT), imagery Intelligence (IMINT), open-source Intelligence (OSINT), and signals Intelligence (SIGINT). MVO, the Easy scenario, had fewer options to consider, fewer additional reports (allowing more reading time), and treated the information in each document as a complete whole. California Scenario (Califon), on the other hand, split an important clue across two documents requiring participants to properly interpret the information available (e.g., Clue 1 and Clue 2 were meaningless on their own, but became important in conjunction). Califon also had more options to consider and more Intelligence documents to act as 'distractors or noise' to decrease the amount of time participants could spend per document. All participants experienced both levels of Task Difficulty.

3.3.3. Teaming level

Teaming level was manipulated by recruiting/assigning participants into the Individual Team or Paired Team condition. The purpose of this manipulation was to compare performance of an individual analyst against performance of a pair of analysts. The Individual and Paired analyst conditions were identical in all respects, except that in the Paired condition analysts worked

together to write a single, joint IR. Paired Teams also had their vocal interactions recorded with the attached Lavalier microphones.

Balancing the design required a minimum of 36 Intelligence operators to serve as our participants. This included 12 operators in the Individual condition and 24 operators to create a total of 12 Pairs. There were four more volunteers than needed for the balanced design. The four extra participants were assigned to the Individual Team condition. The Pairs were self-selected and most of the Pairs knew each other ‘Well’ or ‘Very Well’ prior to participation in the experiment as reported in their Team Survey responses (see Appendix C).

3.3.4. Time

Multiple measurements of some metrics were taken over the course of the experimental session. Thus, time refers to when a measurement was taken. This applies to the KSS, Psychomotor Vigilance Task (PVT), and some physiological measures.

3.4 Dependent Variables

This Section describes the dependent variables used in this experiment, including how they are measured and rationale for use.

3.4.1. Long- and short-term Fatigue

The *Epworth Sleepiness Scale* (ESS; Johns, 1991) is an 8-item questionnaire, based on a 4-point Likert scale that is used as a subjective measure to ascertain an individual’s general level of sleepiness. Each question required the participant to rate the chances of dozing off or falling asleep while participating in daily tasks. The ESS is a long-term fatigue measure. Higher ESS scores correspond to higher amounts of daytime sleepiness. This questionnaire was administered once at the beginning of each research session via JADE.

The KSS (Åkerstedt, 2012; Gillberg et al., 1994) is a short-term measure of fatigue that assesses an individual’s situational feelings of alertness/drowsiness from within the preceding 15 minutes. The KSS is a 9-point Likert scale, where subjects are asked to self-report their current state from 1 (extremely alert) to 9 (extremely sleepy). Higher scores tend to increase with prolonged bouts of wakefulness and are strongly correlated with time of day. Given that the KSS is sensitive to fluctuations over time, the KSS was administered twice per session or four times total (pre- and post-task for each Fatigue condition) via JADE. This subjective rating is most directly relevant for assessing the Fatigue manipulation.

The PVT (Grant et al., 2017) is an objective measure designed to measure sustained attention, alertness/fatigue, and psychomotor speed. It is also an important manipulation check. The test measures an individual’s reaction time in response to a visual stimulus that occurs at random, interstimulus intervals. This test is widely used as it is not subject to a learning effect. The standard duration of the PVT is 10 minutes (PVT-10); however, due to time constraints, a 3-minute version was used (PVT-3), which has demonstrated acceptable properties in terms of sensitivity to fatigue (Grant et al., 2017). The output measures of PVT were reaction time and accuracy. As with the KSS, the PVT is also sensitive to changes over time. As such, the PVT was also administered four times per person (pre- and post-task for each Fatigue condition) via JADE. It is useful for both long- and short-term measurements of Fatigue.

3.4.2. Intelligence Task Performance

The participants' primary task was to generate a final Intelligence Report to address the RFI that was presented at the beginning of each scenario. Each report was given an Intelligence Task Performance (hereafter referred to as 'Performance') score based on six Analytic Standards derived from Intelligence Community Directive 203 (ICD 203) Intelligence Community Analytic Standards (Office of the Director of National Intelligence, 2015). Table 3 lists the six standards, along with the type of measure the standard reflects, and number of raters assessing each standard in this experiment.

Table 3. ICD 203 Standards, the type of measure, and the number of raters conducting the assessment in this experiment.

Standard	Description	Type	No. Raters
Sources	Describing the quality and credibility of underlying sources, data, and methodologies	Process	4
Uncertainty	Expressing and explaining uncertainties associated with major analytic judgments	Process	4
Assumptions	Distinguishing between underlying Intelligence information, assumptions, and judgments	Process	4
Alternatives	Incorporating analysis of alternatives	Process	4
Logic	Using clear and logical argumentation	Process	4
Accuracy	Correctness (i.e., right or wrong) & Clarity (i.e., degree of ambiguity)	Outcome	1

As seen in Table 3, there are two types of measures - process and outcome. In operational settings there may never be a 'correct' answer to an Intelligence question, or also likely, that the answer is probabilistic. Given the uncertainty of an answer, the ICD 203 Standards define how certain aspects of the analytical process should be performed to ensure that the highest quality Intelligence is produced to give decision-makers the best chance of success. The first five Standards in Table 3 reflect the quality of the participants' processes when analyzing the materials and drafting their IRs. Four Subject Matter Experts (SMEs), neither involved in constructing the scenarios nor aware of the intended solutions to the scenarios, rated the IRs on these five process Standards (each scored from 0–30 points) using a rubric.

Given that the experimental scenarios were written with a correct solution in mind, it was appropriate to evaluate Report Accuracy as an outcome measure. Accuracy was function of whether the participant identified the correct solution and the clarity of their conclusion. For example, a participant would get full credit for Correctness if they correctly identified the intended solution. However, they would only get partial credit on Clarity if the participant's solution was qualified or 'hedged' in any way (e.g., offering a solution that identified two possible solutions). Accuracy was scored as an average of Correctness (binary; 0 points = incorrect solution, 30 points = correct solution) and Clarity (ternary; 0 points = unclear, 15 points = mildly clear, 30 points = unambiguous). Given the desire to keep the other four SMEs naïve with respect to the correct answer while assessing participant IRs, Accuracy was assessed by only one rater – the scenario designer and co-author of this report (RF).

Three performance levels. As part of a post-hoc, exploratory data analysis process, participants were divided into three performance levels – top, middle, and bottom. The top and bottom performers consist of the 16.5% highest and lowest scores, respectively. Middle performers consist of the remaining 67%.

3.4.3. Time spent viewing Intelligence documents

The JADE software allows us to track the interactions with the document database to determine the time each of the different documents were open. This allowed analysis of the time viewing each of the different types of documents as a function of the experimental manipulations.

3.4.4. Well-being

As described in the introduction, one of the primary motivators for this study was to understand the negative psychological impact on Intelligence operators. Both Bryant-Lees et al. (2021) and Chappelle et al. (2019) measured burnout using the *Maslach Burnout Inventory-General Survey* (MBI-GS; Maslach & Leiter, 2016; Schaufeli et al., 2002) in their research. Given the MBI's role in prior research in this area, it was logical to include here to assess the current state of mental health/burnout within the domain. All three well-being instruments described in this subsection were administered using paper and pencil.

The MBI-GS (or MBI) is a 16-item questionnaire that measures burnout on three scales: exhaustion, cynicism, and professional efficacy. Participants were asked to rate how often they experience situations related to their work on a six-point Likert scale. Cutoff thresholds for each subcomponent were determined by values cited in Chappelle et al. (2019): exhaustion ≥ 20 ; cynicism ≥ 20 ; and professional efficacy ≤ 12 .

Participants also were asked to complete the *State-Trait Anxiety Inventory-Adult* (Spielberger, 2020). The STAI measures "subjective feeling of tension, apprehension, nervousness, and worry, by activation or arousal of the ANSs" (p. 6). A total of 40 items that contained statements such as, "I feel calm", which participants were asked to rate on a 4-point Likert scale from "not at all" to "very much so." The STAI was scored on two sub-scales. The first is the T-scale which measures *Trait Anxiety* which is a relatively stable attribute of individuals that reflects their "tendency to perceive stressful situations as dangerous or threatening and to respond to such situations" (p. 7). The S-scale measures *State Anxiety*, which is expressed in response to a specific situation. The commonly used cutoff score of 40 is considered indicative of clinically significant levels of anxiety for both subscales.

The STAI instrument was included in this study due to the potential relation of anxiety to burnout and stress and the potential association with the physiological measures used in this experiment which index activation of the ANS.

The ERQ (; Gross & John, 2003) is a ten-item questionnaire on a 7-point Likert scale that evaluates an individual’s capacity to regulate their emotion via cognitive reappraisal and expressive suppression. Gross & John (2003) define *cognitive reappraisal* “[as] a form of cognitive change that involves construing a potentially emotion-eliciting situation in a way that changes the emotional impact (p. 349).” *Expressive suppression*, on the other hand, is described as “a form of response modulation that involves inhibiting ongoing emotion-expressive behavior” (p. 349).

Scores reflect the participants’ tendency to utilize cognitive reappraisal or employ expressive suppression as strategies for coping with stress. Gross & John (2003) found that these two different strategies have important implications both for individual health and for social relations. For example, people who rely heavily on cognitive reappraisal take a more optimistic attitude and “have fewer depressive symptoms, and greater self-esteem, life satisfaction, and every other type of well-being” (p. 360). Conversely, they report that those who rely heavily on expressive suppressions “score lowest in the domain of positive relations with others; they also have lower levels of self-esteem, are less satisfied with life, and have more depressive symptoms” (p. 360). These findings suggest that assessment of these approaches for coping with emotions may be an important individual difference to consider when selecting people to work on AETs in contexts that may elicit strong emotional responses.

Table 4. Summary of the three self-reported questionnaires and associated constructs.

Instrument	Constructs
Maslach Burnout Inventory (MBI)	Exhaustion Cynicism Professional Efficacy
State-Trait Anxiety Inventory (STAI)	State Anxiety Trait Anxiety
Emotion Regulation Questionnaire (ERQ)	Cognitive Reappraisal Expressive Suppression

3.4.5. Subjective workload

NASA TLX. Subjective workload was assessed after each scenario using the modified NASA TLX (Hart, 2006; Hart & Staveland, 1988) via JADE. Composite scores were computed as the sum of the six sub-components (i.e., physical demand, mental demand, temporal demand, frustration level, effort, and performance). Higher values indicate higher levels of subjective workload. This was included as part of a manipulation check of Task Difficulty.

3.4.6. Difficulty Survey

Scenario difficulty was also assessed at the conclusion of each scenario via the Difficulty Survey. In this survey, participants were asked to rate the scenario difficulty in terms of four categories (easy, moderately easy, hard, extremely hard). They were also asked to provide written feedback about their rationale and experience with the scenario.

3.4.7. Heart rate (HR) and heart rate variability (HRV)

HRV metrics are derived from the inter-beat-intervals (IBI) or time between adjacent heart beats and reflect the neurocardiac regulation of the ANS. The ANS is composed of two branches, the parasympathetic and sympathetic branches, which predominate during rest and arousal, respectively. The duality between these two branches exists on a continuum and is constantly in flux, enabling humans to rapidly cope and adapt to their environment, stimuli, workload, etc.

HRV metrics are generally analyzed in three different domains: time, frequency, and nonlinear. These measures reflect the variability, power, and unpredictability of the IBI time series, respectively. Generally, HRV is measured longitudinally (over a 24-hour period) or within short-term (five-minute segments). Short-term HRV is dictated by two processes: (1) autonomic changes resulting in modulation of the sympathetic and parasympathetic branches; (2) regulatory mechanisms associated with changes in respiration via the vagus nerve (respiratory sinus arrhythmia), blood pressure (baroreceptor reflex), and changes in vascular tone. Table 5 below provides an overview of the various HRV frequency domain metrics that were used. Normalized units (nu) were used in all frequency domain metrics to remove variability within and across individuals. Frequency domain measures were obtained using a Lomb Scargle Periodogram transformation (Fonseca et al., 2013).

Table 5. Identification and categorization of heart rate variability metrics related to cognitive load and fatigue.

Metric	Domain	Definition	ANS Relevance
HF* (nu)	Frequency	Relative power of high frequency band (0.15–0.4 Hz) in normalized units; HF/(LF + HF)	Reflects parasympathetic activity and respiratory sinus arrhythmia; lower values correlate with stress and anxiety.
LF* (nu)	Frequency	Relative power of low frequency band (0.04–0.15 Hz) in normalized units; LF/(LF + HF)	Produced by both the parasympathetic and sympathetic nervous systems and baroreflex activity.
LF/HF	Frequency	Ratio between LF and HF bands.	Measures the balance/ratio of the sympathetic and parasympathetic nervous systems. Lower LF/HF reflects parasympathetic dominance when energy is conserved. Higher LF/HF is indicative of sympathetic dominance related to fight-or-flight behaviors.

*Note: HF=High Frequency; LF=Low Frequency

At the start of each experimental session participants were fitted with a Polar H10 ECG wearable device that outputs IBI data. This device is worn as a chest strap around the sternum which allows direct contact of the sensors with the skin. This device is nonintrusive with respect to the cognitive work and many of the participants noted they completely forgot they were wearing them.

The data from the Polar H10 was segmented into 4 periods:

1. A 5-minute baseline at the start of an experimental session
2. A 3-minute anticipatory baseline immediately prior to beginning the analysis task
3. The time while participants are doing the analysis task and writing the Report
4. A 3-minute recovery baseline immediately following submission of the Report

As is typical of physiological research, the HRV data collected during this experiment was examined using both time domain and frequency domain methods. However, only frequency domain results are presented within the main body of this report, since these results have the most significant implications for the purposes of this research. The three frequency domain metrics were obtained using the relative power in the low frequency band in nu (LF), the relative power in the high frequency band in nu HF, and the ratio LF/HF. These metrics are useful for making inferences about the balance between sympathetic and parasympathetic activation. The sympathetic nervous system is typically associated with high levels of alertness, and the parasympathetic nervous systems is associated with relaxation. Higher LF/HF ratios would reflect relatively higher activation in the sympathetic nervous system which is associated with an excitatory or arousal response. This could reflect an activation of cognitive resources to deal with a complex or difficult situation.

3.4.8. Blood pressure

It is well substantiated in the literature that the chronic effects of stress result in increased blood pressure (Gasperin et al., 2009; Vrijkotte et al., 2000). Blood pressure is reported in two numbers – systolic and diastolic – which represent the pressure of blood exerted against arterial walls during the contraction and rest phase of the cardiac cycle, respectively (DeMers & Wachs, 2023). The inclusion of blood pressure as a measure in this present study was not intended to serve as any form of a clinical assessment on the participant, but rather to draw conclusions surrounding the impacts of the experimental manipulations on the participants’ blood pressure and relations to the well-being questionnaires associated with burnout and stress. Blood pressure was measured at three distinct time points:

- Three times during the 5-minute resting baseline at the start of the experimental session
- One time during a 3-minute anticipatory baseline immediately prior to the beginning of the task
- One time during a 3-minute “recovery” baseline immediately following submission of the Report

3.4.9. Eye metrics

Eye tracking is a thoroughly studied method for assessing cognitive and perceived workload. Oculo-based metrics can be categorized into pupillometric or fixation-based measures. Increased pupil dilation and dilation rates are associated with increased cognitive load and have been identified as the best eye-based metrics to assess mental effort and load due to sensitivities to within-task, between-task, and between-individuals changes in processing load (Mahanama et al., 2022). Behavioral metrics such as fixation rate / duration, saccade rate / duration, and blink count / rate / duration have been related to task engagement and fatigue, which are known to affect cognitive load (Di Stasi et al., 2012; Schleicher et al., 2008; Tansakul & Tangamchit, 2015). Table 6 provides a breakdown of all eye tracking metrics included in the analysis.

Table 6. Identification and categorization of eye tracking metrics related to cognitive load and fatigue.

Metric	Definition	Fatigue Relevance
Pupil Diameter	The diameter of the pupil, often ranging from 4-6 mm	Larger pupillary dilations are associated with increased cognitive load and increased mental effort
Fixation Duration	Fixations are the most common type of eye movement and occur when the field of vision is held in one place	Fixation duration indicates resources required to extract information from the relevant source of fixation and reflects information processing load and engagement; Fixation duration increases with fatigue and workload
Saccade Duration	Saccades are rapid movements of the eye that result in abrupt changes in the point of fixation; Saccades can last from 20-200 ms	Saccade duration increases during fatigue states
Saccade Velocity	Change in degrees per millisecond	Fatigue decreases the saccadic peak velocity and demonstrates increased latency
Saccade Amplitude	The distance traveled by the eye in degrees between two fixation points	Higher saccadic amplitudes suggest that the eye has traveled a greater distance during the saccade
Blink Rate	Reported in blinks per minute	Increased blinks during High Fatigue
Blink Duration	Duration in milliseconds for when a blink occurs	Increased blink duration during High Fatigue

Immediately prior to the start of the scenario, participants were fitted with a set of Pupil Core eye tracking glasses. Each pair of glasses contained two cameras that monitored each of the participant’s pupils and one world view camera that monitored the visual scene. Glasses were calibrated and data were collected within the included Pupil Capture software (Pupil Labs GmbH, Berlin, Germany).

Eye tracking metrics were offloaded from the Pupil Capture software via the manufacturer’s Pupil Player software (Pupil Labs GmbH, Berlin, Germany). The exported gaze positions, pupil data, and blink data were then ingested into MATLAB 2023A (MathWorks, Natick, MA) for further analysis. All the metrics provided in Table 6 (above) were obtained via Mile Two developed MATLAB scripts across the entire scenario and segmented by different types of Intelligence document types (i.e., noise, guidance, diagnostic).

3.5 Procedure

Figure 2 shows the general sequence of activities. In an initial phase (Setup and baseline), participants completed several questionnaires and were fitted with the physiological sensors. Baseline physiological measures were made for BP and ECG during this period. Additionally, a PVT task was administered. Note that during the initial experimental session, participants were given a quick tutorial to demonstrate the JADE UI and overview of the experimental task. This overview was not repeated for the second session. The second phase (Primary Task) involved the actual experimental task. Participants were given 52 minutes to review Intelligence data and produce the Report. With the Report complete, the final phase (post-task) involved the administration of additional subjective measures such as the NASA TLX, a second PVT and KSS session, and feedback questionnaire. The initial experimental session lasted approximately 2-hours. The second session tended to last approximately 70-minutes.

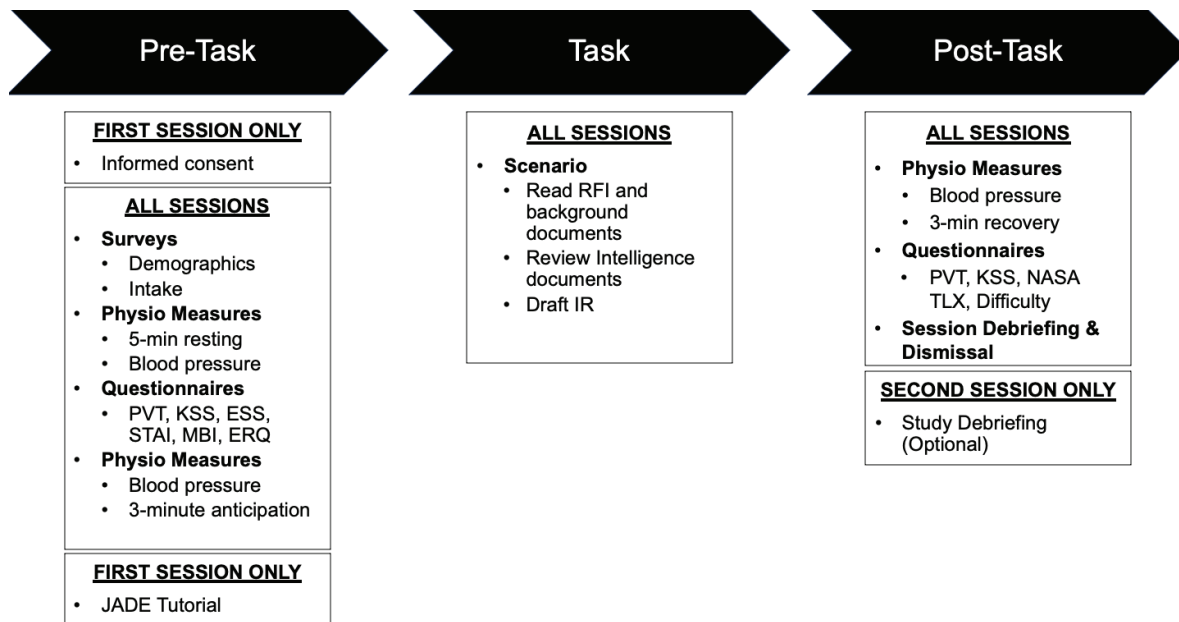


Figure 2. Diagram illustrating the procedure and sequence of activities.

4.0 RESULTS & DISCUSSION

This section describes the results in the following order:

- (Section 4.1) Responses to the pre-task questionnaires related to well-being to determine how our population of analysts compared to the results from prior studies that have suggested that burnout and fatigue may be a concern for analysts working in DCGSs (Bryant-Lees et al., 2021; Chappelle et al., 2019). Baseline blood pressure results are also reported as a physiological measure of general health and well-being.
- (Section 4.2) Responses from the sleep questionnaires and PVT tests to gauge the impact of the Fatigue manipulation.
- (Section 4.3) Responses to subjective measures of workload to assess the impact of the Difficulty manipulation.
- (Section 4.4) Intelligence Task Performance results.
- (Sections 4.5 & 4.6) HRV and eye tracking physiological results.
- (Section 4.7) Physiological Response × Performance Level.
- (Section 4.8) Time Spent Viewing Documents in JADE.
- (Section 4.9) Results and case study related to the Team Survey.

This research is an exploratory field study. The emphasis in the design has been on external (ecological) validity, and a major motivation was to explore the practical feasibility of using physiological measures as evidence for assessing performance during Intelligence analyst DCGS operations. Thus, we use a liberal cutoff of $p = 0.1$ as a criterion for reporting effects. For this reason, results should be considered as tentative evidence that will need to be explored under more controlled conditions before drawing firm conclusions about theoretical or practical implications.

4.1 Well-being

To follow up on prior research, participants were asked to respond to three instruments related to their overall mental health before beginning the experimental task (see Table 7). In addition, baseline blood pressures were taken as a potential physiological indication of high levels of stress.

Table 7. Test-retest correlation coefficients (Pearson’s *r*) for the three well-being surveys.

Instrument	Construct	<i>r</i>
Maslach Burnout Inventory (MBI)	Exhaustion	0.83
	Cynicism	0.77
	Professional Efficacy	0.81
State-Trait Anxiety Inventory (STAI)	State Anxiety	0.54
	Trait Anxiety	0.58
Emotion Regulation Questionnaire (ERQ)	Cognitive Reappraisal	0.75
	Expressive Suppression	0.62

In this experiment, participants were required to come into the lab for two experimental sessions - once before a shift (low Fatigue) and once immediately following a shift (High Fatigue). Given these instruments were all designed to measure relatively enduring personal attributes, we calculated test-retest reliability of responses from the two administrations using the Pearson’s *r* statistic. The correlations for each component of each instrument are included in the table above. There appears to be an acceptable level of agreement between the two administrations of each instrument indicating that participant responses were similar across both administrations.

4.1.1. Maslach Burnout Inventory

The MBI is defined by three subscales: Emotional Exhaustion, Cynicism, and subjective Professional Efficacy. High scores on Emotional Exhaustion and Cynicism, along with low scores on professional efficacy are considered concerning for burnout and the different patterns of scores on the subscales yields information on the nature of the burnout (Leiter & Maslach, 2016). The results of this scale for the current study are examined using the scoring norms from previous research in similar groups whereby “concerning” scores on Emotional Exhaustion and Cynicism subscales are scores greater than or equal to 20 and a “concerning” score on Professional Efficacy is a score of 12 or below.

Figure 3 shows the results from the current study within the context of the results from previous studies (Bryant-Lees et al., 2021; Chappelle et al., 2019). Results from Chappelle et al., (2019) were from a sample of 1436 Active-Duty DCGS Intelligence operators. The next four results are from Bryant-Lees et al., (2021) the combined measure included pilots, sensor operators, and Intelligence operators from AFSOC and ACC. Also, included are the results for ACC operators only: drone pilots (*n* = 252), sensor operators (*n* = 91), and Intelligence operators (*n* = 70). For the current study, *n* = 40.

Responses from the current study suggest that with respect to responses to the Maslach Burnout Inventory our participants showed patterns very similar to the patterns observed in previous investigations. It appears that the Exhaustion and Cynicism components of Burnout continue to be high despite:

- Interventions in response to the earlier concerns about burnout and stress in DCGSs through the introduction of Airmen Resiliency Training (ART)
- Changes in the mode of operations within the DCGS (e.g., increasing use of an AET framework)
- Changes in the tempo and nature of missions.

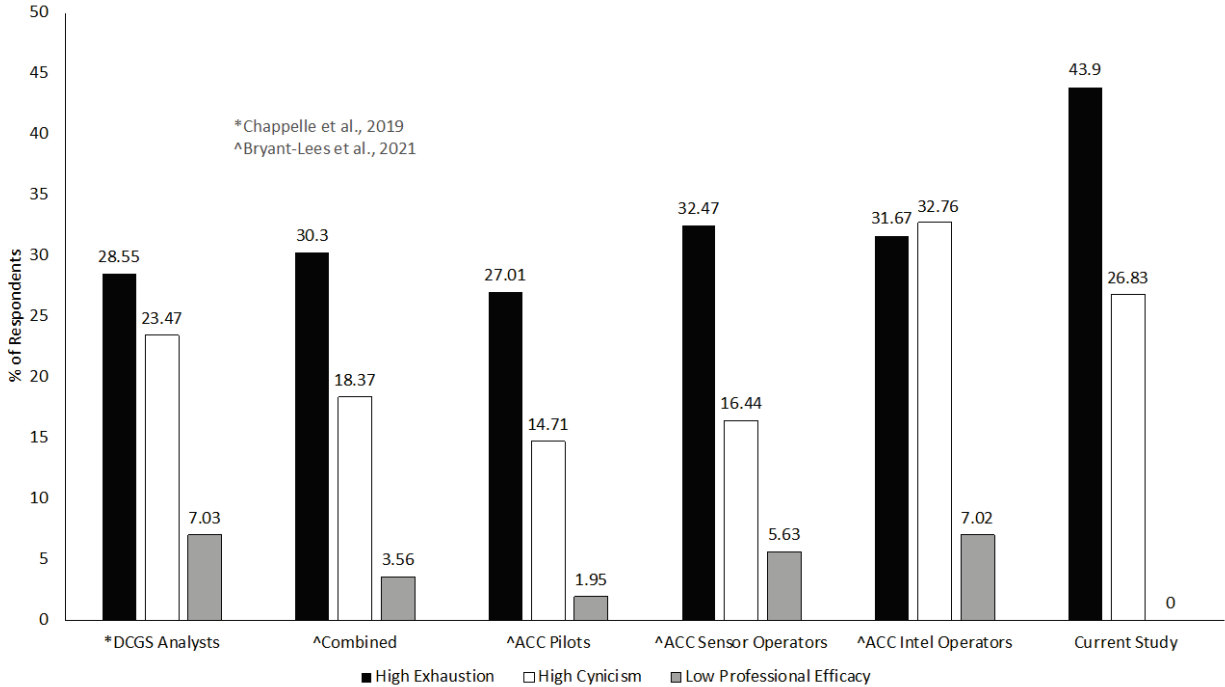


Figure 3. The percentages of respondents who reported concerning levels on the three components of the MBI (i.e., 20 or above on the Exhaustion and Cynicism scales; and 12 or below on the Professional Efficacy scale).

4.1.2. State-Trait Anxiety Inventory

Overall, the scores for the participants in this study were consistent with “low or no anxiety” classification according to this measure on both state and trait anxiety scales. Figure 4 shows the S- and T-Anxiety scores as a function of Fatigue level and Team condition. The lines show norms obtained from a sample of 1,387 Working Adult Males (Spielberger, 2020). The dotted line is the norm for State Anxiety and the solid line is the norm for Trait Anxiety. Note that the standard deviation for the norms were 10.40 for State Anxiety and 9.19 for Trait Anxiety. Although the levels of State Anxiety appear to vary as a function of Fatigue level, there was no main effect of Fatigue found in a two-factor analysis of variance (ANOVA). This suggests that being more tired does not increase the experience of anxiety or change the ability to regulate anxiety. Overall, the results indicate that the mean S- and T-Anxiety scores for our participants were consistent with observations in a comparable population.

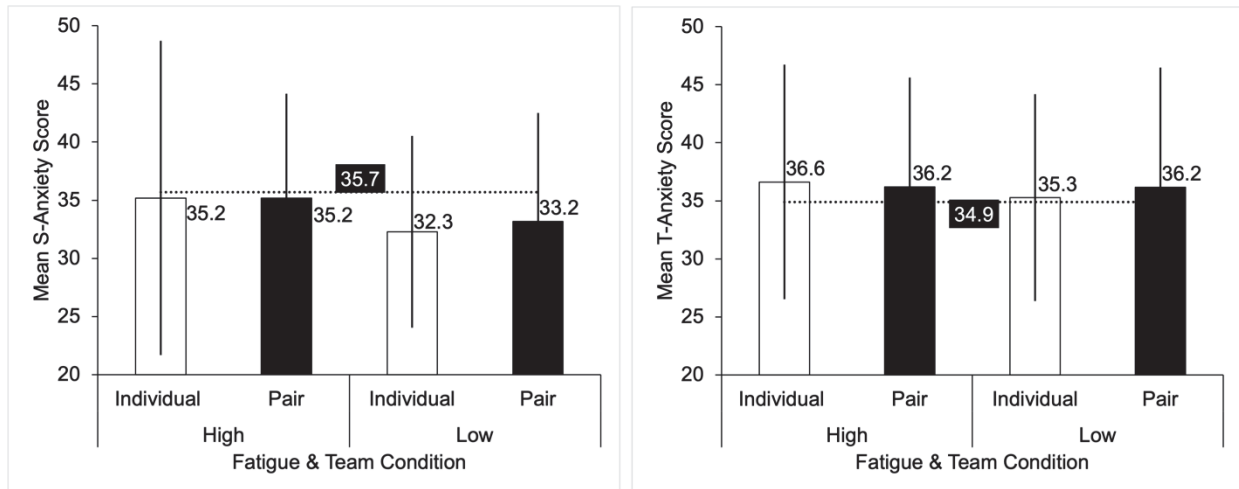


Figure 4. Mean S-Anxiety (left) and mean T-Anxiety scores (right) as a function of Fatigue and Team condition.

4.1.3. Emotion Regulation Questionnaire

The Mean and Standard Error for Cognitive Reappraisal scores was $M=30.2$ ($SE=1.0$). For Expressive Suppression, it was $M=15.6$ ($SE=0.7$). These scores were similar to those found in a study of 1,048 participants (across three samples) from the general community by Preece et al. (2020), where the approximate mean Cognitive Reappraisal score was $M=28.9$ ($SE=0.4$) and the approximate mean Expressive Suppression score was $M=15.9$ ($SE=0.3$).

These results, shown in Figure 5, suggest that the sample of analysts who participated in the present experiment included many people who employed a cognitive reappraisal coping strategy. This is typically considered the healthier coping strategy and this result is consistent with the optimism and the positive and enthusiastic engagement that we observed among most of our participants. This indicates that overall, participants in our sample are utilizing healthy strategies to deal with the stresses that they may be experiencing in their work.

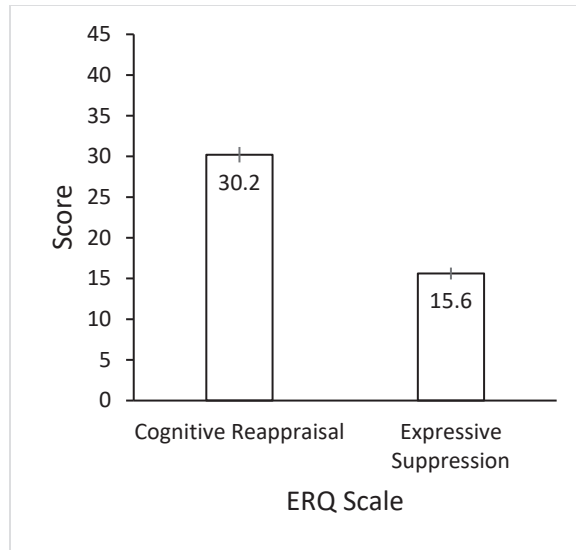


Figure 5. ERQ Cognitive Reappraisal and Expressive Suppression subscale means.

4.1.4. Baseline blood pressure

In addition to the questionnaires baseline blood pressure was taken before the start of this experiment as a general gauge of the overall health of the participants. Blood pressures were measured three times during a five-minute resting baseline period and the average of these measurements was taken.

Figure 6 and Figure 7 show the results as a function of the Fatigue and Team manipulations. The data show that the blood pressures were generally in the normal range for adults (i.e., Systolic < 120; Diastolic < 80). The results indicate that blood pressures were slightly elevated in the High Fatigue conditions (at the end of a work shift), relative to the Low Fatigue conditions (after or on a rest day). However, this difference was only marginally significant when assessed in a two-factor ANOVA [Systolic: $F(1,72) = 3.30, p = 0.07$; Diastolic: $F(1,72) = 3.03, p = 0.08$].

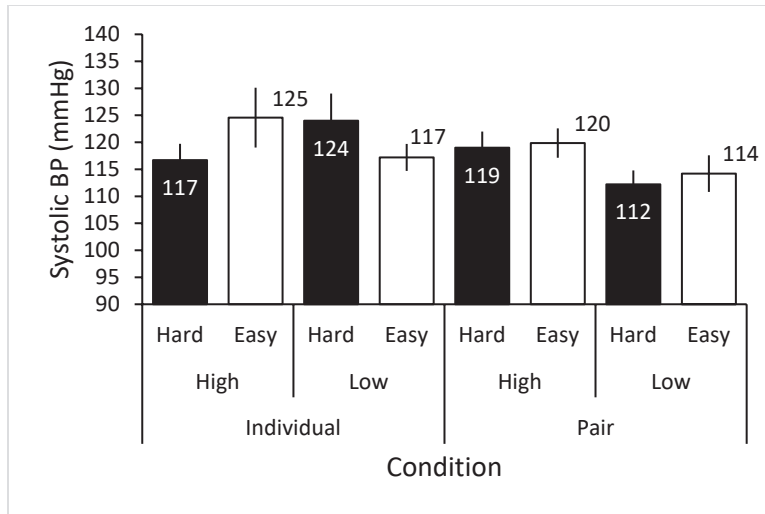


Figure 6. Mean baseline systolic blood pressures measured prior to the experimental sessions as a function of Team, Fatigue, and Difficulty.

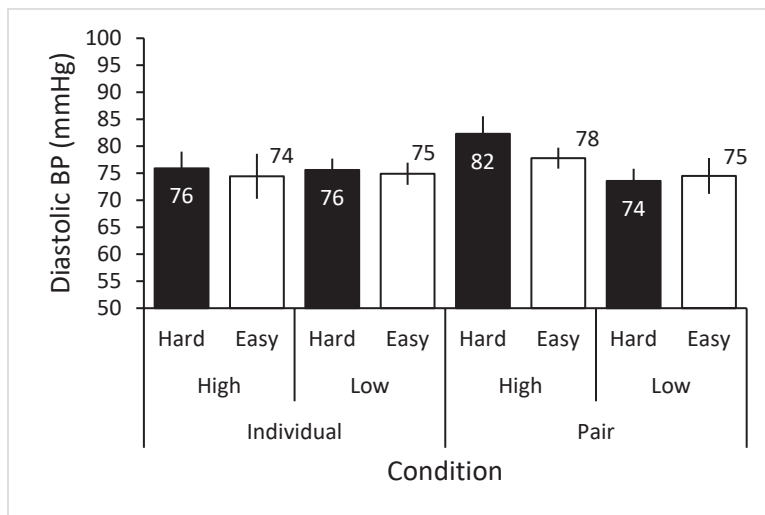


Figure 7. Mean baseline diastolic blood pressures measured prior to the experimental sessions as a function of Team, Fatigue, and Difficulty.

4.1.5. Well-being summary

Results for the MBI indicate that there are still high percentages of Airmen who show high levels of Exhaustion and Cynicism. However, none of the Airmen in our sample showed levels of professional efficacy below the critical levels identified in previous studies. Also, most of the Airmen in this sample showed that they used the healthier strategy (cognitive reappraisal) for coping with stress and anxiety. State and Trait anxieties and blood pressures seemed to be within normal ranges. Thus, this indicates that burnout is still a potential problem, but that, for the most part, the Airmen are managing the stresses effectively.

4.2 Fatigue Manipulation

The endemic prevalence of sleepiness in the military is largely affected by the demanding nature of the duties that include, but are not limited to, shift changes, long shift hours, and insufficient sleep durations. There were three measures that served to assess fatigue: the ESS (subjective daytime sleepiness), KSS (current subjective sleepiness), and the PVT (reaction time / time on task). The ESS was administered once at the beginning of the scenario it measures general sleep habits. The KSS and PVT were administered before and after the scenario. These two measures are most relevant for assessing the impact of the fatigue manipulation.

4.2.1. Epworth Sleepiness Scale (ESS)

Figure 8 shows the mean ESS scores as a function of the Fatigue and Team independent variables. Error bars represent standard errors. Note that the clinical cutoff for “excessive daytime sleepiness” is ≥ 10 , which is represented by the dotted line. A paired t -test was performed to compare mean ESS composite score on Fatigue state. There was no significant difference between the means of High vs Low Fatigue on ESS composite score [$t(39) = 1.38, p = 0.18$]. The absence of a Fatigue effect here is not surprising since the ESS is designed to assess general patterns of sleep behavior, rather than momentary feelings of sleepiness that would be impacted by the manipulation.

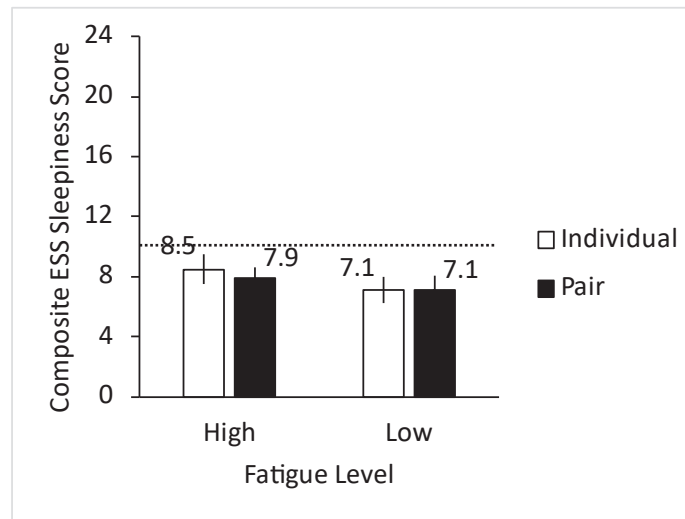


Figure 8. Mean ESS scores as a function of Fatigue and Team conditions.

4.2.2. Karolinska Sleepiness Scale (KSS)

The responses to the KSS were more important in assessing the Fatigue manipulation because this measure focused on current feelings of sleepiness within the past 15 minutes. The responses to the KSS were assessed using a four-factor ($2 \times 2 \times 2 \times 2$) ANOVA. This included two levels of Fatigue, two levels of Difficulty, two levels of Team, and two different times (Pre- and Post-Experimental Task). As seen in Figure 9, there was a significant difference as a function of the Fatigue levels [$F(1, 142) = 25.54, p < 0.001$]. Mean KSS score for the High Fatigue = 4.7 and for Low Fatigue = 3.1, which are categorized as “neither alert nor sleepy” and “alert”, respectively. No other main effects or interactions were significant. This result helps to validate the Fatigue

manipulation. Participants showed higher levels of sleepiness as indexed by the KSS instrument in the High Fatigue conditions.

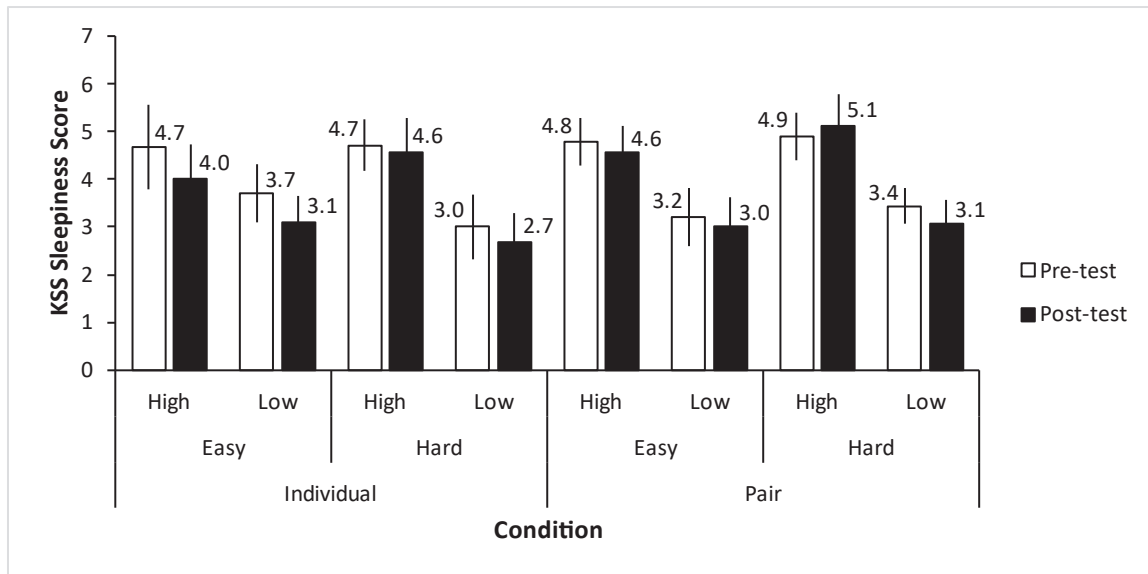


Figure 9. Mean pre- and post-test KSS sleepiness scores as a function of Team, Difficulty, and Fatigue.

4.2.3. Psychomotor Vigilance Task (PVT)

The PVT assessed alertness/attention via the measure of reaction time. There is extensive literature supporting the sensitivity of PVT to fatigue. The results for the PVT were evaluated using a 4-way ($2 \times 2 \times 2 \times 2$) ANOVA with the independent variables of Fatigue, Difficulty, Team, and Time (pre- vs post-task).

There was a significant main effect for Fatigue [$F(1,142) = 6.51, p = .01$]. As with the KSS scores, this provides additional validation for the Fatigue manipulation. Participants were slower in the High Fatigue conditions (313.8 milliseconds [ms]) than in High Fatigue conditions (300.6 ms). There was also a significant three-way interaction between Fatigue, Difficulty, and Team [$F(1, 142) = 5.07, p = .02$] as shown in Figure 10.

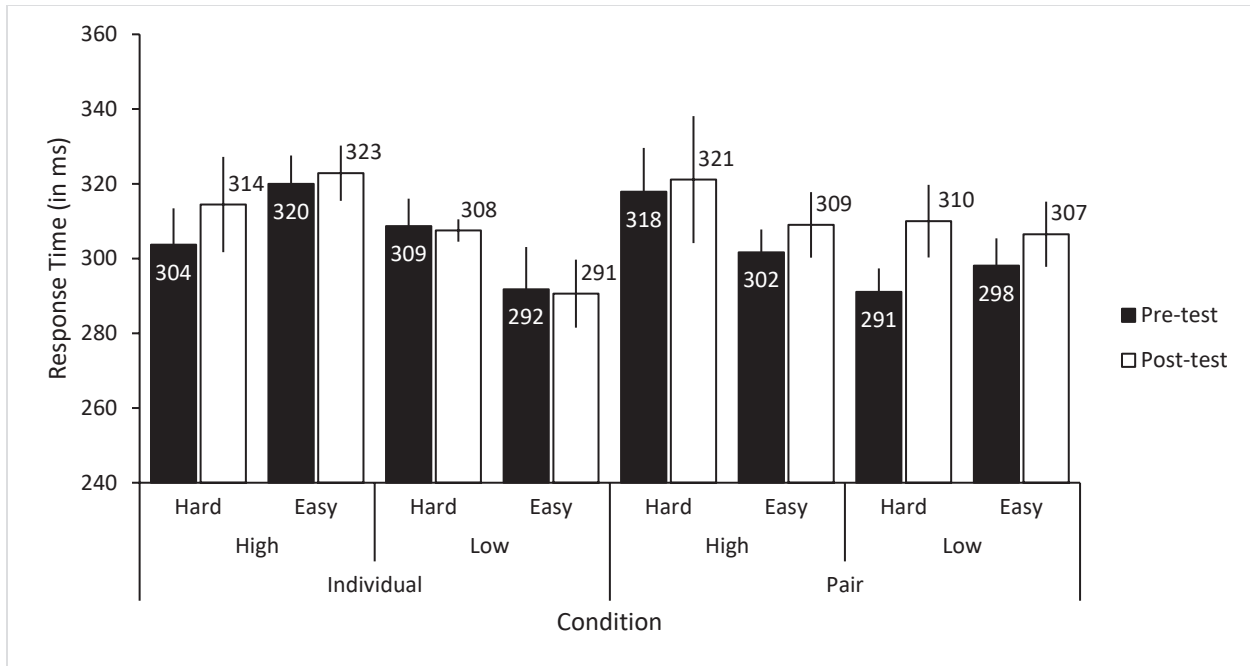


Figure 10. Mean pre- and post-test PVT response times (in ms) as a function of Team, Fatigue, and Difficulty.

4.2.4. Fatigue summary

Subjective reports using the ESS suggested that the overall sleeping patterns of the participants were not different as a function of the experimental conditions. However, subjective reports of immediate feelings of sleepiness from the KSS and objective performance obtained using the PVT test indicated that participants in the High Fatigue condition were more fatigued than in the Low Fatigue condition. This suggests that the manipulation of Fatigue was effective.

4.3 Difficulty Manipulation

Two subjective instruments were used to gauge the impact of the Difficulty manipulation – the NASA TLX and a Difficulty questionnaire. Participants responded to these questionnaires at the end of each experimental session.

4.3.1. Modified NASA TLX results

The composite workload scores for the NASA TLX were analyzed using a 3-way ANOVA with two levels of Fatigue, two levels of Difficulty, and two levels of Team. There were no significant main or interaction effects. However, there was a marginal effect for Difficulty [$F(1,70) = 3.25$, $p = 0.07$]. As can be seen in Figure 11, the mean composite scores for the Easy Scenarios ($M=18.2$) tended to be lower than for the Hard Scenarios ($M=20.2$). This provides partial support that the manipulation of task Difficulty was effective.

When removing two outliers from the dataset, there was a significant main effect for Difficulty [$F(1,66) = 4.60$, $p = 0.036$], the Hard scenario ($M=20.47$) had a higher subjective workload than the Easy scenario ($M=17.39$). The two outliers were both participants from the Individual Team condition; one individual had a self-reported reading comprehension deficiency, and the other

individual had a performance score of 0.5 and 2.79 for the Hard and Easy scenario, respectively (average scores were 16.46 and 17.55 for Hard and Easy, respectively).

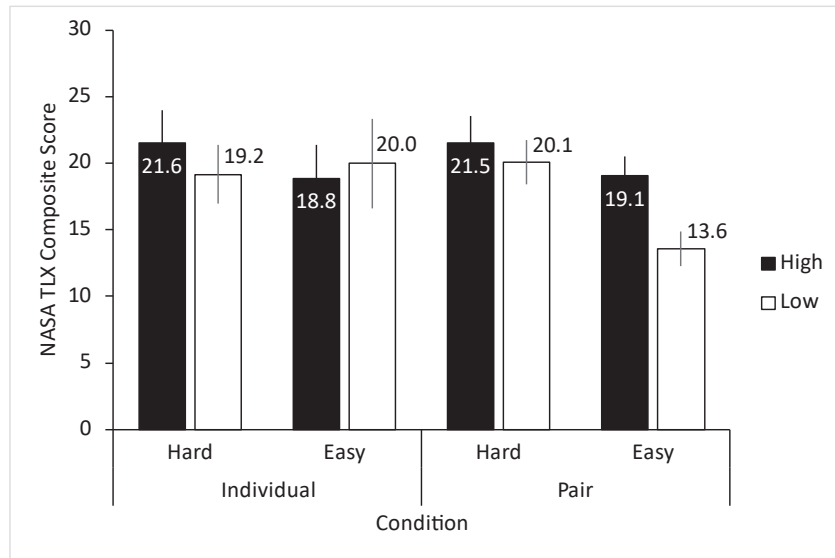


Figure 11. Mean composite TLX ratings as a function of Difficulty, Fatigue, and Team variables.

4.3.2. Global team workload

In addition to evaluating team workload from the perspective of each individual within a Paired Team, an aggregate, global team workload score was taken for the participants in the Paired condition. Funke et al. (2012) suggest the use of the global team workload score, when the goal is to compare workload across varying task demands and team characteristics. A 2-way ANOVA was conducted using the aggregate team workload score evaluated against Fatigue state (High or Low) and Difficulty (Hard or Easy). The main effect of Difficulty ($F(1,19) = 4.32, p = 0.051$) was marginally significant. There were no other significant main effects or interaction effects.

4.3.3. Difficulty questionnaire results

In addition to the TLX, participants also rated difficulty on a 4-point scale (Easy, Moderately Easy, Hard, Extremely Hard) at the end of each experimental session. As can be seen in Figure 12, no one rated the Califon (Hard) Scenario as *Easy*, but nine participants rated the MVO (Easy) Scenario as *Easy*.

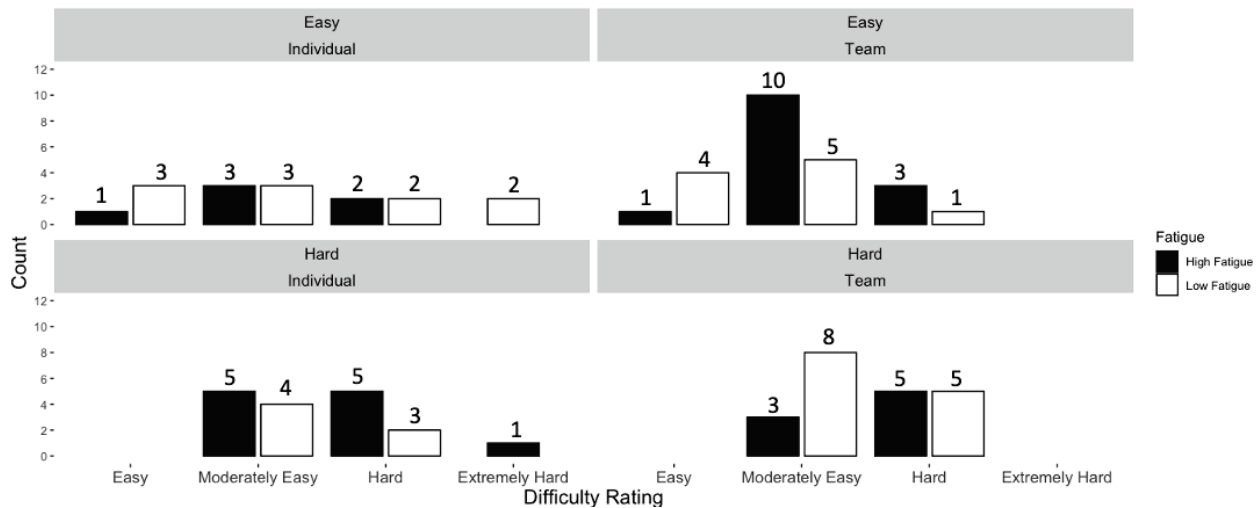


Figure 12. Participants’ ratings of subjective scenario difficulty, plotted as a function of Difficulty/Scenario (Califon Hard vs MVO Easy) and Individual vs. Team conditions.

One participant rated the Califon (Hard) Scenario as *Extremely Hard*, and two participants rated the MVO (Easy) Scenario as *Extremely Hard*. One participant who rated the MVO Scenario as *Extremely Hard* indicated having a reading disability:

“For me, there wasn’t enough time to read through all the information to grasp the scenario/intelligence question to start the analyst process. Given another 30 mins I would have been able to complete the report. Is a person’s level of reading comprehension considered in this study?”

The second participant who rated the MVO Scenario as *Extremely Hard* indicated that the nature of the scenario was different than their area of specialization:

“Realistic, but not representative of my line of work as [a full motion video] Analyst so it was quite challenging for me to accomplish in a timely manner and didn’t feel well-equipped and well trained for this kind of work.”

A similar comment was made by the lone participant who rated the Califon Scenario as *Extremely Hard*.

4.3.4. Difficulty summary

The two scenarios used in this study were chosen to reflect two different levels of cognitive demand or difficulty. The Califon Scenario was intended to be more difficult than the MVO Scenario. However, the subjective impression of difficulty depends both on the nature of the scenario and the skills of the participants making the judgement. As both the composite TLX scores and the difficulty ratings indicate, there was significant variability in the assessment of difficulty. This is at least in part due to the diversity among the population of participants, which included people with various specializations (e.g., all source analysts and full motion video analysts) and different levels of experience (e.g., rank, and years in the DCGS).

Overall, the subjective reports of workload and difficulty indicate that the manipulation of Difficulty did have an impact. The trends all indicate that the Califon Scenario was generally considered more demanding than the MVO Scenario.

4.4 Intelligence Task Performance

As mentioned in Section 0, there were two measures used to assess the quality of participants' Report: (1) correctness (2) adherence to the guidelines set forth by the ICD 203 Analytic Standards. Correctness was evaluated by the Scenario Design SME. Four other different SMEs graded each Report against the ICD 203 rubric. The judgement of adherence to the ICD 203 Analytics Standards was largely subjective; therefore, scores were averaged across the 4 SMEs and then combined with the correctness score to yield a final, total score. The total score was used to bin participants into a performance rank distribution of Top (16.5%), Middle (66%), and Bottom (16.5%) performers.

4.4.1. Correct responses

Thirteen out of 14 Individuals and all 12 Pairs came to the same conclusion as intended by the SME for the Easy Scenario (MVO). For the Hard Scenario (Califon), only a single person identified the intended solution. So, in terms of getting the correct answer to the scenarios, these results indicate that performance did vary as a function of the task Difficulty manipulation.

4.4.2. SME ratings

The final SME-rated scores for each Report were determined by taking the average of the ratings across the five SMEs and six criteria listed in the Methods section of this report. Thus, a maximum possible score was 30 points. Because of the subjective nature of the ratings, Intraclass Correlations (ICC) were computed to assess the reliability of the first four SME ratings. Table 8 shows the ICCs computed for each of the five criteria. According to (Cicchetti, 1994) correlations less than 0.40 indicate poor agreement among raters, and correlations between 0.40 and 0.59 indicate fair agreement. As can be seen in Table 8, there is fair agreement on four criteria (i.e., Alternatives, Logic, Sources, & Uncertainty) and poor agreement for only one (i.e., Assumptions).

Table 8. Intraclass Correlations (ICC) of SME ratings of Performance on the Process measures

Scenario	ICD 203 Standard				
	Alternatives	Assumptions	Logic	Sources	Uncertainty
Califon	0.51	0.35	0.43	0.51	0.48
MVO	0.55	0.35	0.45	0.55	0.51

The scores were analyzed using a 3-way (2×2×2) ANOVA with Fatigue, Difficulty, and Team. There were no significant main effects or interactions. Figure 13 shows the interaction between the Fatigue and Team variables. Although the effect was not significant, the impact of fatigue is in the expected direction with the reports generated from the High Fatigue conditions being lower than for the Low Fatigue condition [$F(1,47) = 0.62, p = 0.44$]. Note that although there was no main effect for Teams [$F(1,47) = 1.55, p = 0.22$], the lowest ratings were received by

Teams in the High Fatigue condition. This is contrary to conventional wisdom that suggests that ‘two heads are better than one.’

Like the TLX data, performance was also analyzed after removing the two outliers. There was a significant main effect of Teaming on performance [$F(1,44) = 5.57, p = 0.02$], where Individuals had higher total scores ($M=17.4$) than Teams ($M=15.6$). There were no other main effects or interaction effects.

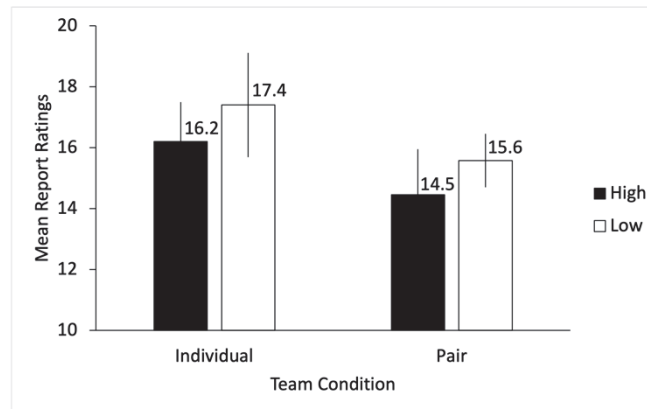


Figure 13. Mean Report ratings as a function of Fatigue and Team.

The only independent variable that had impact on the SME scores of the IRs was the Team manipulation. Individuals performed better than Pairs. However, because of the low levels of agreement among the SMEs, any conclusions must be considered conservatively. That said, there appeared to be a clear difference between the Easy and Hard Scenarios in terms of the participants’ ability to draw a conclusion that was in line with the intended solution.

4.4.3. Intelligence task performance summary

Two measures of performance were used. The first measure was simply whether the Report came to the same conclusion as intended by the SME who created the scenarios. In essence, did they arrive at the correct solution? For this measure, there was a clear difference as a function of difficulty level with participants doing much better for the Easy scenario. The second measure was determined based on rating of SME who used a rubric for quality analysis. This measure is relevant because in operational conditions it may not be possible to identify an objectively correct conclusion. Thus, this measure reflected the quality of the analysis and logic, independent of whether it was correct or not. However, there were relatively low levels of agreement among the SMEs who rated performance. This may explain why there were no main effects on the SME scores as a function of difficulty. However, there was one surprising result - individuals performed better than teams. This suggests that creating a quality joint report may have been more difficult than when an individual had full control over the report (i.e., one author is better than two). To this end, it is important to recognize that the scenarios were originally intended for research with individuals, not pairs. We acknowledge that more complex scenarios (i.e., more intelligence documents, greater time constraint, etc.) may result in a fairer test of the potential benefits of teaming.

4.5 Heart Rate Variability

Numerous aspects of the HRV data were examined using both time domain and frequency domain metrics as is typical of physiological research. However, as mentioned earlier, this report focuses solely on the frequency domain, specifically at normalized, relative power in the high- and low-frequency ranges (HF and LF).

4.5.1. High & low frequency normalized unit metrics

HF and LF normalized metrics were assessed via a 3-way ($2 \times 2 \times 2$) ANOVA using values averaged across the entire scenario for Fatigue, Task Difficulty, and Team. As shown below in Figure 14, average baseline corrected HF and LF were net positive and net negative, respectively. Note that normalization of LF and HF makes the two variables linearly associated [$HF = HF \times 100 / (LF+HF)$; $LF = LF \times 100 / (LF+HF)$]; therefore, it is expected that the results of these two variables are the same in magnitude, but opposite in direction.

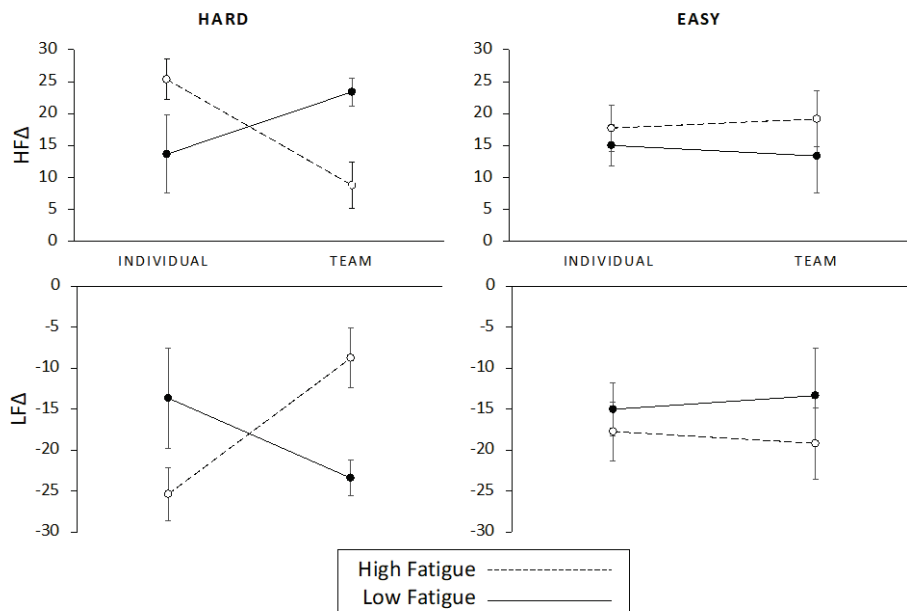


Figure 14. Baseline Normalized HF and LF (in nu; top and bottom, respectively) averaged throughout the entire scenario across Fatigue, Difficulty, and Team.

The ANOVA resulted in a significant three-way interaction between Fatigue, Difficulty, and Team for both baseline normalized HF [$F(1,45) = 5.40, p = 0.02$] and baseline normalized LF [$F(1,45) = 5.40, p = 0.02$], as shown in Figure 14. For the Easy level of Difficulty, there were higher levels of HF (and low levels of LF) in the High Fatigue condition for both the Individual and Pair levels of Teams. In essence, a main effect for Fatigue was found in this level of Difficulty. However, for the Hard level of Difficulty, there was a crossover effect. The impact of Fatigue for the Individual level of Teams was the same as for the Easy Difficulty conditions – HF was higher in the High Fatigue condition. But for the Team condition higher levels of HF (and low levels of LF) were found for the Low Fatigue condition. The two-way interactions

between Fatigue and Team were also significant for HF [$F(1,45) = 4.50, p = 0.04$] and LF [$F(1,45) = 4.50, p = 0.04$].

Thus, the pattern of ANS activation as a function of Fatigue seems to be different for the Pairs when working on the Hard Scenario than for any of the other three combinations of conditions. In all the other conditions the mean HF activation is higher and LF activation is lower in the High Fatigue Condition. However, in contrast, when the Pairs were in the High Difficulty condition HF activation was lower for the High Fatigue condition and LF was higher. Since a high LF/HF ratio is typically associated with activation of the sympathetic nervous system, which may reflect the recruitment of resources to deal with a cognitively challenging situation – this may be an indication that the need to coordinate with a teammate may increase the need for cognitive resources. In essence, this suggests that the costs of collaboration outweighed the benefits of having two heads for this specific task. Note also that Teams in the High Fatigue condition received the lowest report ratings from the SMEs.

4.5.2. Heart rate variability summary

Frequency analysis of HRV was sensitive to the experimental manipulations. For individuals, HF power was higher for the High Fatigue condition for both Hard and Easy scenarios. However, a different pattern was observed for teams. Similar to individuals, the HF power was higher for the Easy Fatigue condition, but in the Hard Scenario condition the LF power was higher. This pattern suggests that the teams had higher activation of the sympathetic nervous system in the most stressful session (High Fatigue and Hard Scenario). One possible interpretation is that the social demands of collaboration add to the stress levels experienced – pushing them over a tipping point in terms of the need to recruit more resources.

4.6 Eye Tracking Metrics

Numerous metrics from the eye tracking data were examined across the entire scenario and segmented across Intelligence document types (guidance, diagnostic, and noise). However, only significant results or results approaching significance were included in this report.

4.6.1. Averaged across the entire scenario

Each eye tracking metric was evaluated using a 3-way ANOVA, with two levels of Difficulty, two levels of Fatigue, and two levels of Team. There was one significant result and several results that approached significance.

4.6.1.1 Blink rate

The manipulation of Fatigue was influenced by the time of experimentation in relation to participant's shift, but also, time on task. There was a marginal main effect of Fatigue [$F(1,52) = 2.95, p = 0.09$]; the average blink rate was higher in High Fatigue (10.2 ± 6.3 blinks / min) condition than Low Fatigue (7.8 ± 3.9 blinks / min) conditions. This finding coincides with reports in literature suggesting that more fatigued individuals tend to have a higher blink rate.

There was a marginal main effect of Team on Blink Rate [$F(1,52) = 3.31, p = 0.07$]; the average blink rate was higher in Pairs (10.2 ± 5.5 blinks / min) than Individuals (7.7 ± 5.0 blinks / min). Variations in blink rate across the Team condition could be related to the communication and interaction of Pairs compared to Individuals.

4.6.1.2 Saccadic amplitude

There was a marginal main effect of Fatigue on saccadic amplitude [$F(1,58) = 3.09, p = 0.08$]. The average saccadic amplitude was higher during conditions of Low Fatigue (17.8 ± 4.2 mm) than in conditions of High Fatigue (15.9 ± 4.7). Findings on amplitude coincide with previous research, that suggests that amplitude decreases with time on task (Chen et al., 2022). This pattern of results reflects underlying physiology related to the control of attention. The superior colliculus is a cortical structure that plays a significant role in saccade and attentional control and whose downstream pathway helps to initiate eye movements, such as blinks or saccades and is influenced by factors such as fatigue (Chen et al., 2022; Peel et al., 2017).

4.6.1.3 Pupil diameter

There was a marginal main effect of Team on pupil diameter [$F(1,52) = 3.26, p = 0.08$] where Pairs (2.94 ± 0.93 mm) had a smaller average pupil diameter than Individuals (3.29 ± 0.51 mm). This size difference could be attributed to the information load sharing, where Pairs were permitted to work collaboratively on all aspects of the scenario. Furthermore, there was a significant effect for Team on fixation duration [$F(1,58) = 5.98, p = 0.02$].

4.6.1.4 Fixation duration

The average fixation duration was higher in Pairs (111.9 ± 16.2 ms) than Individuals (101.3 ± 18.1 ms). This might reflect interactions between participants to direct attention. Pairs could call out specific aspects within the document that their partner should pay attention to, thus increasing the average dwell time.

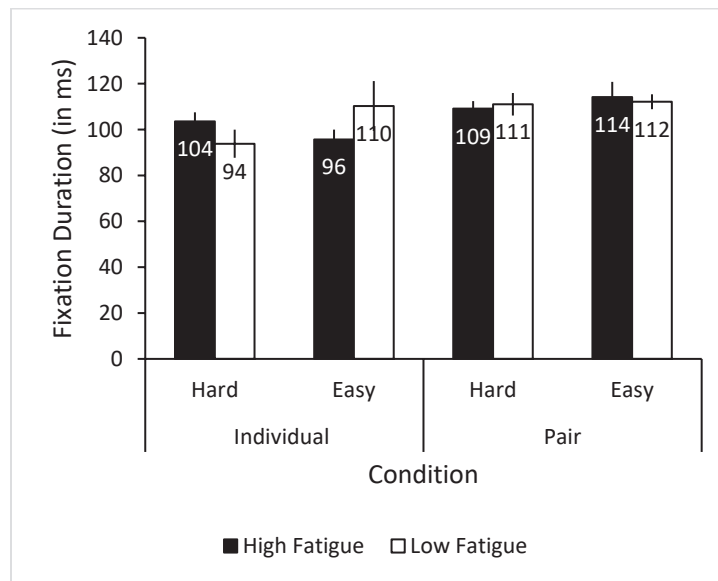


Figure 15. Mean fixation duration (in ms) as a function of Fatigue, Team, and Difficulty.

4.6.2. Eye tracking metrics by Intelligence document type

In addition to evaluating changes in eye tracking measures across the entire scenario, each eye tracking metric was evaluated as a function of the Intelligence document type. As mentioned previously, each type of Intelligence document was categorized as diagnostic, guidance, or noise.

Three separate, 3-way ANOVAs were conducted on each of the types as a function of Difficulty, Fatigue, and Team. Table 9 provides a breakdown of all the significant results or results approaching significance. As shown below, Team was the primary manipulation driving the significant effects even though the documents for the Individuals and Pairs were identical with the exception that Pairs collaborated to make sense of the documents and to create the final Report.

There were three significant main effects on pupil diameter as a function of the Team manipulation across all three types, where diameter was consistently lower in Pairs than Individuals. Increased diameter is associated with increased cognitive load. The ability to distribute the load with another individual may have resulted in Pairs having consistently lower pupil diameters. The prevalence of the Team manipulation was found in all the interaction effects, suggesting that irrespective of the type of document, working collaboratively with another individual had a measurable impact on eye behavior.

The only other main effect that was marginally significant was found in the saccadic amplitude metric whilst viewing Diagnostic Intelligence documents across the Fatigue manipulation. These results were consistent when amplitude was averaged across the entire scenario (presented in Section 0); namely, amplitude was higher in a state of Low Fatigue than High Fatigue.

Table 9. Eye tracking metrics broken down by Intelligence document type that were significant or approached significance.

Type	Eye Tracking Metric	Manipulation Variable	Statistic	Means \pm <i>SD</i>
Diagnostic	Pupil Diameter	Team	$F(1,53) = 7.63$ $p = 0.008^{**}$	<ul style="list-style-type: none"> Individual: 3.42 ± 0.57 mm Pair: 2.88 ± 0.82 mm
	Saccade Duration	Fatigue State \times Team	$F(1,53) = 2.82$ $p = 0.01^*$	<ul style="list-style-type: none"> Individual \times High: 24.92 ± 0.58 ms Individual \times Low: 24.44 ± 0.73 ms Pair \times High: 24.4 ± 0.8 ms Pair \times Low: 24.6 ± 0.93 ms
	Saccade Duration	Fatigue State \times Difficulty \times Team	$F(1,53) = 3.52$ $p = 0.06$	<ul style="list-style-type: none"> Individual \times High \times Hard: 24.99 ± 0.68 ms Individual \times Low \times Hard: 24.18 ± 0.74 ms Individual \times High \times Easy: 24.83 ± 0.45 ms Individual \times Low \times Easy: 24.68 ± 0.71 ms Pair \times High \times Hard: 24.17 ± 1.01 ms Pair \times Low \times Hard: 24.83 ± 0.61 ms Pair \times High \times Easy: 24.61 ± 0.51 ms Pair \times Low \times Easy: 24.4 ± 1.14 ms
	Saccade Amplitude	Fatigue State	$F(1,53) = 3.18$ $p = 0.08$	<ul style="list-style-type: none"> High: 14.85 ± 5.63 degrees Low: 17.54 ± 5.6 degrees
	Pupil Diameter	Team	$F(1,53) = 6.76$ $p = 0.01^*$	<ul style="list-style-type: none"> Individual: 3.24 ± 0.54 mm Pair: 2.76 ± 0.81 mm
Guidance	Fixation Duration	Team	$F(1,53) = 6.45$ $p = 0.01^*$	<ul style="list-style-type: none"> Individual: 87.43 ± 13.1 ms Pair: 96.19 ± 13.46 ms
	Pupil Diameter	Team	$F(1,53) = 8.00$ $p = 0.007^{**}$	<ul style="list-style-type: none"> Individual: 3.31 ± 0.51 mm Pair: 2.77 ± 0.82 mm
Noise	Fixation Duration	Difficulty \times Team	$F(1,51) = 3.27$ $p = 0.08$	<ul style="list-style-type: none"> Individual \times Hard: 95.01 ± 17.44 ms Individual \times Easy: 87.18 ± 20.08 ms Pair \times Hard: 89.56 ± 11.8 ms Pair \times Easy: 98.82 ± 19.72 ms

4.6.3. Eye metrics summary

Blink Rate and Saccadic Amplitude were sensitive to the Fatigue manipulation. Blink Rate was higher and Saccadic Amplitudes were lower in the High Fatigue Condition. Multiple eye metrics were sensitive to the Team manipulation. Pupil diameters were smaller, and Fixation Rates were longer for Pairs. The Team factor was also involved in numerous higher order interactions shown in Table 9.

There were no significant main effects or interactions as a function of Difficulty. Interestingly, a prior study found that measures related to eye movement dynamics (fixation and saccade) dynamics were influenced by Fatigue due to time on task, irrespective of task Difficulty (Di Stasi et al., 2013). This finding supports other results from this current study and is also corroborated by the results of the first experiment.

4.7 Physiological Metrics by Performance Level

In addition to evaluating changes in physiology associated with cognitive workload and fatigue, understanding how physiology changes in accordance with performance level or expertise is also of interest. The identification of key metrics or patterns associated with performance decrement or deviations from optimal performance are just as useful as characterizing the impact of cognitive load itself. Namely, modulation of an individual's ability to appropriately allocate resources in response to varying degrees of operational workload demands, can provide a representation of individual state performance and regulation capacity (Pezzulo et al., 2015). Comparing the dynamical changes in resource allocation or relevant performance outcomes across time, individuals, or interventions enables the characterization of adaptation / maladaptation or operational expertise (Davids et al., 2003). Previous research suggests that an individual's psychophysiological profile while completing a task is predictive in identifying novices from experts (Benner, 1984).

This section looks at frequency domain metrics of HRV and eye-movement metrics as a function of three levels of three different levels of performance scores (Top, Middle, & Bottom).

4.7.1. HRV × Performance level

There were notable trends in HF and LF differences between High and Low Fatigue states (see Figure 16) and Hard and Easy difficulty (see Figure 17) as a function of the analysts' performance based on SME ratings (top, middle, and bottom ranks). Note that the differences (Δ) were calculated by subtracting Easy from Hard for Task Difficulty and by subtracting Low from High for Fatigue Level. These differences use the labeling conventions in Table 10.

Table 10. Labels and descriptions related to HF and LF differences for Fatigue and Task Difficulty

Label	Description
HF Δ_F	High frequency differences between Fatigue conditions
LF Δ_F	Low frequency differences between Fatigue conditions
HF Δ_D	High frequency differences between Task Difficulty conditions
LF Δ_D	Low frequency differences between Task Difficulty conditions

As shown in Figure 16 and Figure 17, Top performers had a negative, median delta HF and positive, median delta LF, in contrast to Bottom and Middle performers across scenario Difficulty and Fatigue states. This suggests that Top performers were consistently in a state of 'sympathetic activation' when completing the Hard scenario or in a state of High Fatigue. In contrast, Bottom performers were in a state of "parasympathetic dominance" whilst completing a Hard scenario or in a state of High Fatigue. These trends are comparable to previous research findings that examined differences in Fatigue state (Vicente et al., 2016) and Task Difficulty (Paskett et al., 2022; Skibniewski et al., 2015).

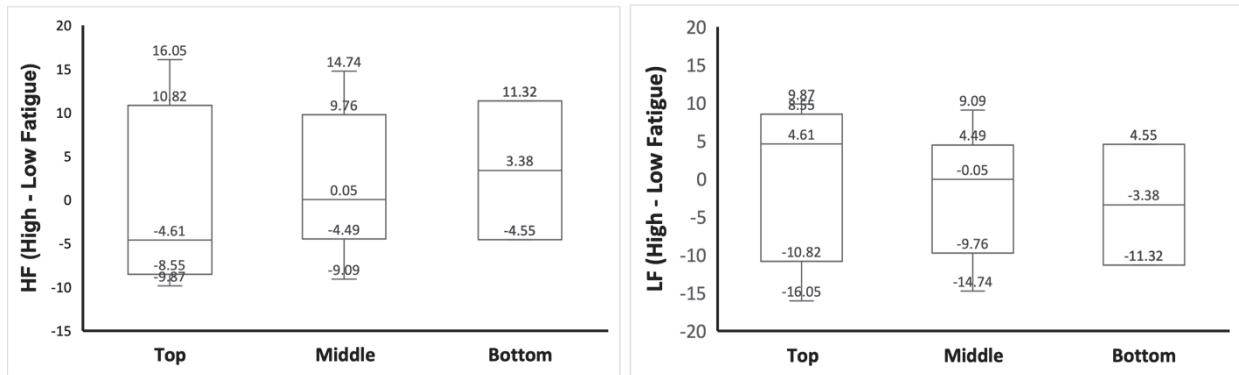


Figure 16. Differences in HF and LF (in nu) between Fatigue conditions as a function of Performance. $HF\Delta_F$ (left) and $LF\Delta_F$ (right).

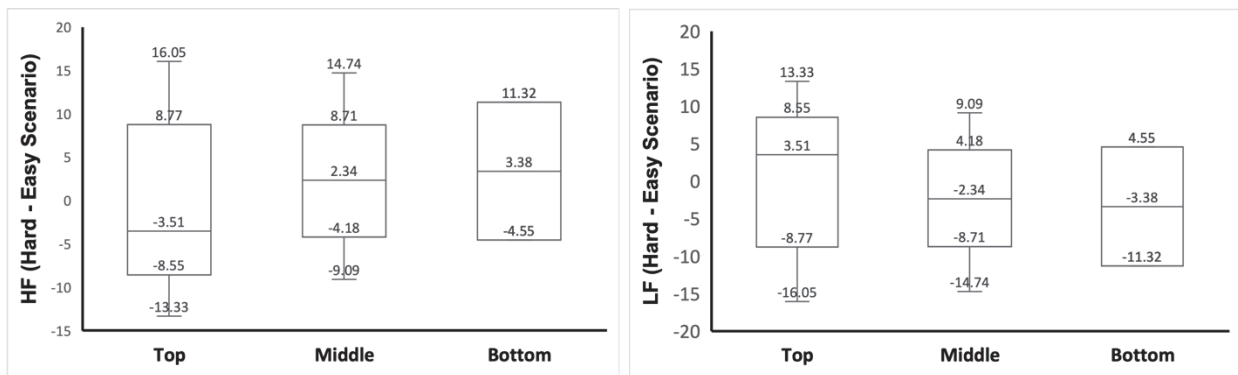


Figure 17. Differences in HF and LF (in nu) between Task Difficulty conditions as a function of Performance. $HF\Delta_D$ (left) and $LF\Delta_D$ (right).

4.7.2. Eye Metrics × Performance level

Eye tracking metrics were evaluated as differences across Fatigue State and Difficulty as a function of Performance level. The most interesting eye tracking trends were observed in blink rate (see Figure 18).

As with HRV metrics, differences in blink rate varied across Performance. The difference in blink rate across Fatigue and Difficulty in Top and Middle performers was nearly negligible compared to Bottom performers who exhibited higher blink rates in conditions of High Fatigue and the Hard scenario (see Figure 18). Normally, fatigue increases blink rate; however, when partitioned by performance, it was observed that Top performers were seemingly more immune to the effects of High Fatigue and Hard scenarios. Our interpretation of these results is motivated by the same patterns that were observed in relation to the delta HF and LF HRV (Section 0). Eye blinks are indicative of reallocation of mental resources, cognitive states, and problem solving (Paprocki & Lenskiy, 2017). Specifically, Top and Middle performers appeared to allocate

resources needed to combat the stressors associated with greater fatigue and cognitive demand, via the reduction in blink rate.

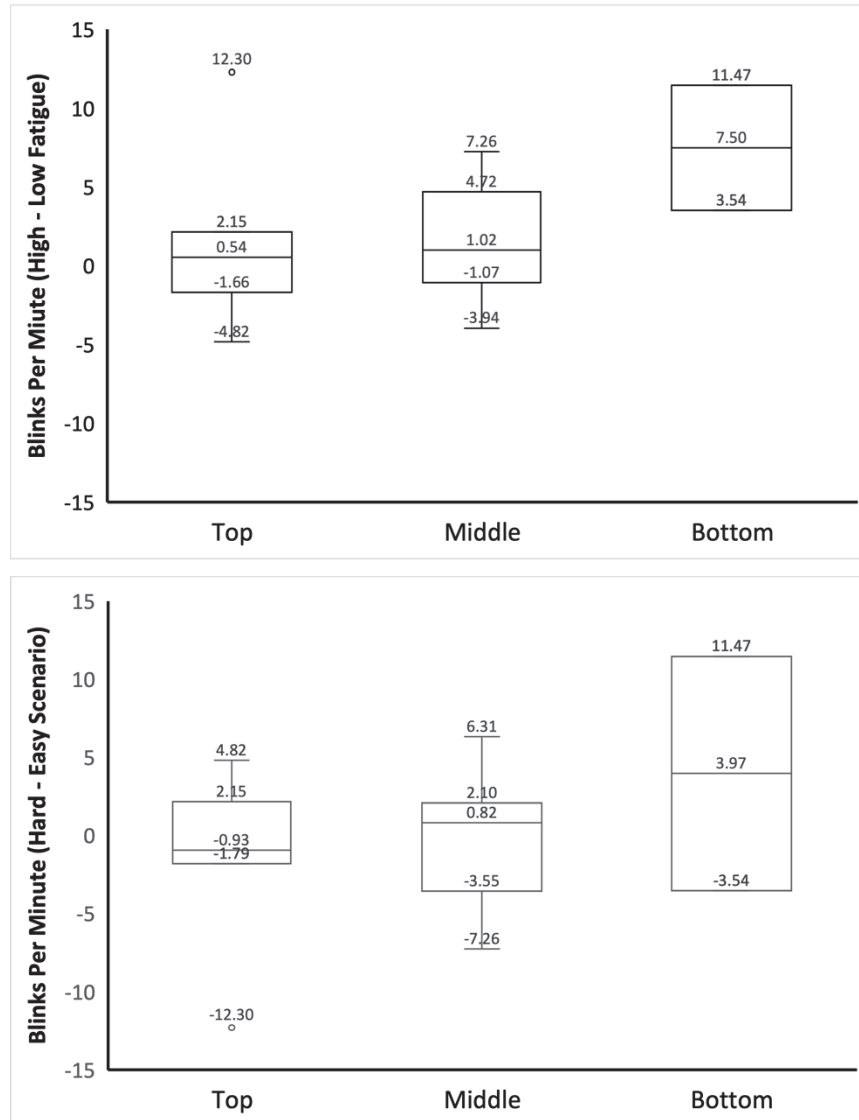


Figure 18. Differences in blink rate across rank as a function of Fatigue (top) and Difficulty (bottom).

4.7.3. Physiological measures × Performance level summary

Speculation for the trends found above coincide with theories on sleep demand (Tanaka & Tamura, 2016; Vicente et al., 2016), resource allocation (Hockey, 1997; Mandrick et al., 2016), level of engagement (Mehler et al., 2009), and challenge vs. threat distinction (Uphill et al., 2019). Perhaps, Top performers recognized that more cognitive effort had to be invested in the Hard scenario or when in a state of High Fatigue state to maintain the same level of effectiveness and efficiency in performance, at the expense of their physiology. By catapulting themselves into a hyper-functional state, physiological resources were expended at a higher rate to minimize adverse effects on task performance.

Prior studies support this conclusion. Namely, Mandrick et al. (2016) examined the impact of high mental workload on psychophysiological responses and found that when there were no significant changes in performance with Task Difficulty, it was at the expense of greater physiological cost. Performance trends found in Experiment 2 corroborate the trends found in Experiment 1 (see AFRL-RH-WP-TR-2023-0019).

4.8 Time Spent Viewing Documents in JADE

As previously mentioned, JADE provides the capability to export key interactions within the software including, but not limited to, time spent viewing Intelligence documents. Average time spent viewing documents was evaluated using a 3-way ($2 \times 2 \times 2$) ANOVA with independent variables Fatigue, Difficulty, and Team. There were two significant main effects of Scenario for the Diagnostic [$F(1,61) = 6.7, p = 0.01$] and Noise [$F(1,61) = 9.0, p = 0.004$] type. The average time spent viewing Diagnostic and Noise documents were 61.1 ± 36.1 s and 57.3 ± 25.9 s, respectively, for the Hard scenario and 86.7 ± 46.4 s and 91.8 ± 62.3 s, respectively, for the Easy scenario. This result is consistent with the fact that there were more documents to view in the Hard scenario, even though the total time for a session was the same. Furthermore, it is unsurprising that the average time spent viewing Guidance documents did not differ significantly across Difficulty because participants were forced to read the background documents prior to reading the supplementary pieces of Intelligence.

There was also a significant main effect of Team for the Guidance [$F(1,61) = 0.19, p = 0.005$] type. Individuals spent more time (279.1 ± 297.3 s) than the Teams (139.2 ± 65.4 s) reviewing each background document, respectively. Although, each individual within the Pair was required to review each background document, Pairs were permitted to collaborate during this time period.

In addition to the main effects, there were also several interaction effects that trended towards significance. The most interesting marginal interaction effect was related to the average time spent viewing Noise documents, across Difficulty \times Fatigue State [$F(1,61) = 3.67, p = 0.06$]. As shown in Figure 19 below, the impact of fatigue is much greater for the Easy Scenario than for the Hard Scenario. In the Easy Scenario participants much spent less time on the Noise documents in the High Fatigue condition.

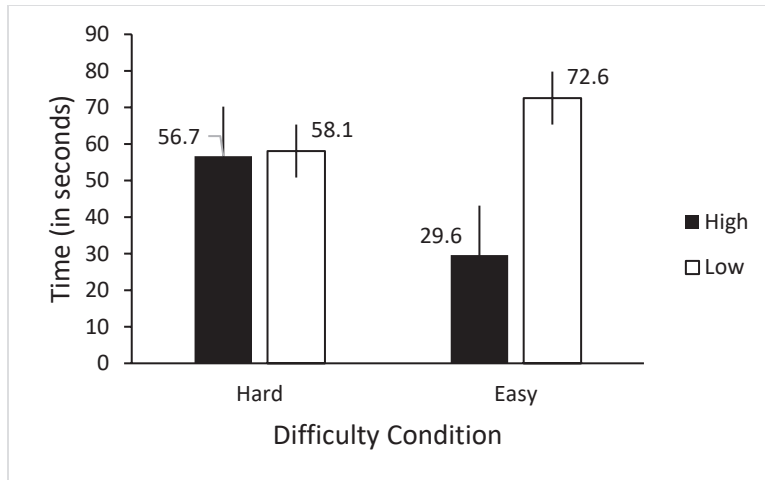


Figure 19. Mean time (in seconds) spent viewing Noise documents as a function of Fatigue and Difficulty.

4.9 Team Survey & Case Study

While participants answered several questions, only one result seemed noteworthy and that is particularly due to a conspicuous physiological response captured by the heart rate monitors. All other Team Survey results are in Appendix C.

As part of the Team Survey, participants were asked “*how satisfied they were with team performance*” and whether they would “*choose to work with the same partner again.*” The distribution of responses to these questions are shown in Figure 20. Overall, the responses showed that all but one of the Teams were Satisfied or Very Satisfied with team performance and all the participants would be willing to work with the same partner again. Reviewing the records revealed that the negative responses for both questions came from the same team.

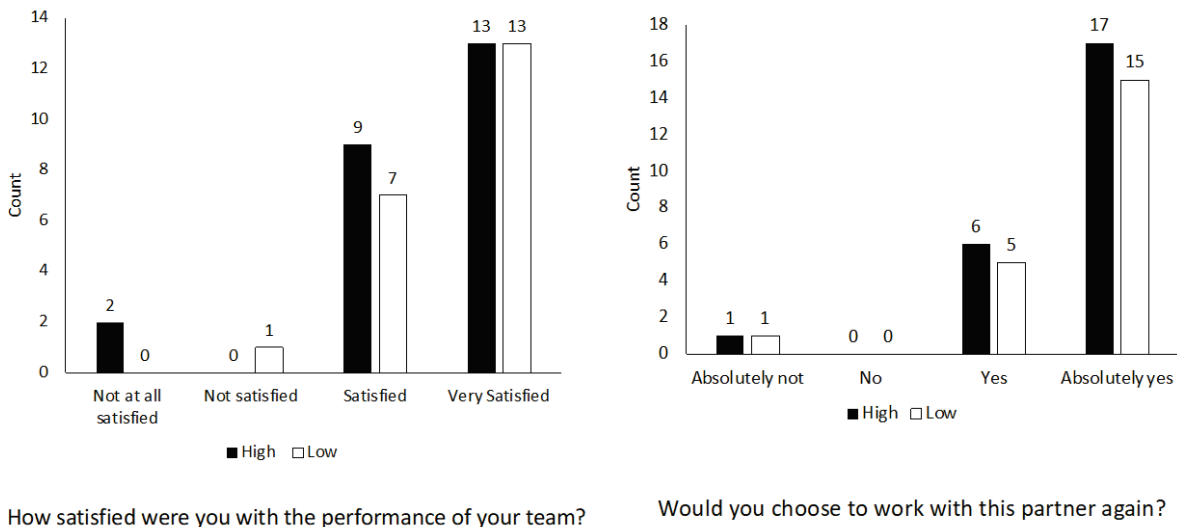


Figure 20. The distribution of responses to the Team Questionnaire items.

This team was also noticeably lower than the overall pool of participants on information sharing and listening. Furthermore, although teams were instructed to submit a single joint final report, in their first session one of the participants on this team submitted the final report without coordinating or consulting with their partner. This led to visible tension, when the second participant in the Pair realized that the report had been submitted without their input or approval.

The physiological responses for this team were examined independently to see whether there was any indication of the social difficulties or tension that this group experienced.

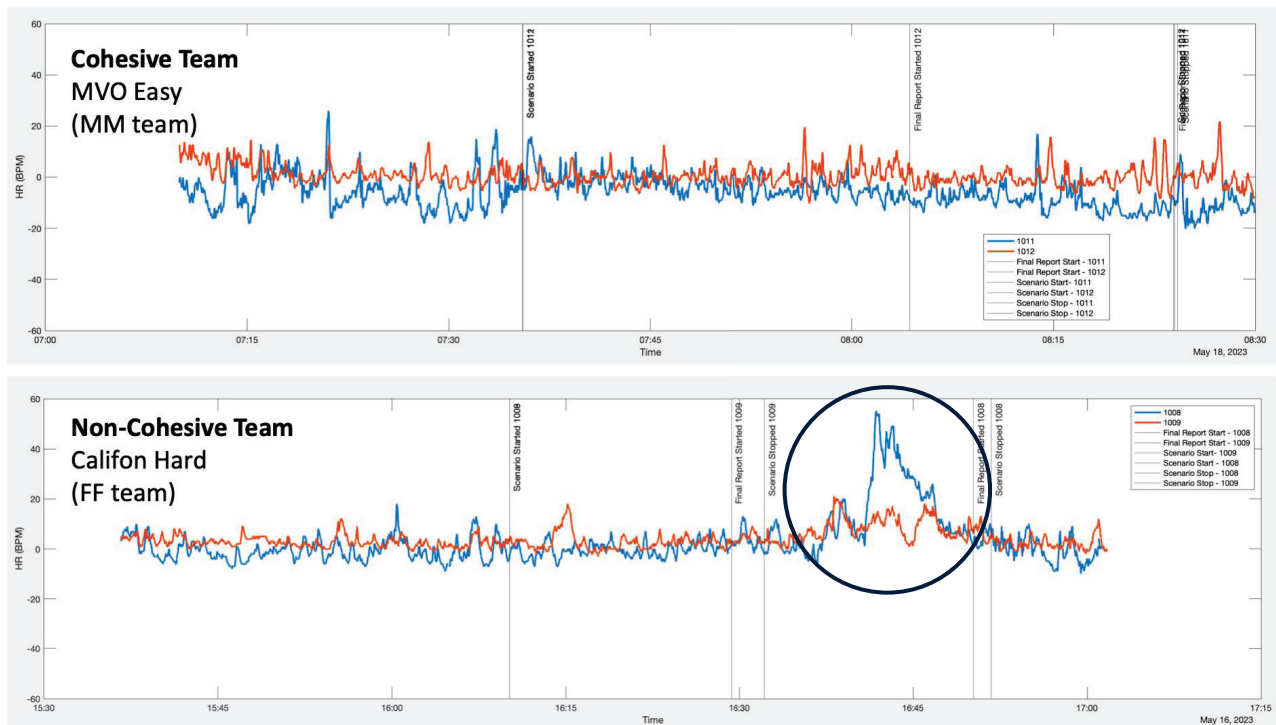


Figure 21. (Top) HR across the entire scenario for a cohesive team, (Bottom) HR across the entire scenario for a non-cohesive team.

As observed in the bottom graph of Figure 21 above, there is noticeable change in HR when it was discovered that their partner submitted the Report without any collaboration or consultation. Although this transient change in HR is not diagnostic to the underlying problem, it does provide insight that there is some stimulus that is evoking such a strong physiological response. From an operational perspective, a commander would be able to take note of this difference and investigate further. This is of particular interest when in the case of interpersonal issues that may affect team dynamics and operational performance but may not be overtly evident to a supervisor.

5.0 GENERAL DISCUSSION & CONCLUSION

This research was designed to explore the potential for using real-time physiological sensors to monitor performance of Intelligence operators working in the DGCS. To this end, the research was conducted using a representative population of Airmen who were all currently stationed at a DGCS. The scenarios and the JADE interface were designed to be representative of Intelligence work in terms of the task and tools provided. Physiological responses during the scenarios were monitored using low cost, consumer grade ECG sensors along with eye tracking glasses. The experiment was designed to explore the impact of three independent variables: Difficulty (Hard vs Easy), Fatigue (High vs Low), and Team (Individual vs Pair) configuration.

Before the experimental sessions, participants filled out several surveys to assess general mental health. Among these was the Maslach Burnout Inventory which had been used in prior studies (Bryant-Lees et al., 2021; Chappelle et al., 2019). Consistent with the previous results, the levels of Exhaustion and Cynicism were high. This indicates that the threat for Burnout remains high.

5.1 Fatigue

Two levels of Fatigue were manipulated by controlling the time of the experimental sessions relative to a participant's work shifts. In the High Fatigue condition, the session was conducted at the end of a normal eight-hour shift. In the Low Fatigue condition, the session was conducted on or after a rest day before participants started their work shift. The impact of the Fatigue manipulation was evident in both subjective reports of sleepiness on the KSS and in objective performance in the PVT.

For the indices of HRV, as with Difficulty, the interesting effects were seen in the frequency domain analyses. For Individuals, the power in the HF range was greater in the High Fatigue and Low Fatigue conditions – which coincides with the literature, suggesting that while in a state of High Fatigue, these participants were more parasympathetically dominant. However, for Teams there was a crossover effect such that HF power was greater in the Low Fatigue condition, but lower in the High Fatigue condition. As noted above – this suggests that coordination within a team increased the demands on cognitive resources. Also, there were similar differences in trends for the Fatigue effect on relative power in the LF and HF ranges for top and bottom performers. Top performers tend to have relatively higher sympathetic activation in both High Fatigue and Hard Difficulty conditions.

For eye metrics the impact of Fatigue resulted in higher blink rates and larger saccadic amplitudes in the High Fatigue condition. This pattern of results reflects underlying physiology related to the control of attention. There were also trends related to quality of performance in the blink rates that were consistent with the observations of HRV frequency domain metrics. The difference between High and Low levels of Fatigue was greater for the bottom performers than for the top performers. Again, this suggests that top performers were better able to allocate resources and control attention in when Fatigue was high.

5.2 Difficulty

Two levels of scenario Difficulty (Easy, Hard) were created and vetted by SMEs. Additionally, the Difficulty manipulation was validated in a prior study. In the present study, the impact of the Difficulty manipulation was confirmed by subjective estimates of workload and difficulty.

Additionally, participant performance in terms of reaching the intended conclusion was better for the Easy scenario.

An impact of Difficulty was observed on the indices of HRV. In Individuals for both Easy and Hard scenarios, there was more power in the HF band. But for the Pair condition, this was only true for the Easy scenario. In the Hard scenario the effect was reversed with higher power in the LF band. Since a high LF/HF ratio is typically associated with activation of the sympathetic nervous system – state of arousal – this may reflect the recruitment of resources to effectively deal with a cognitively challenging situation. Coordination and collaboration with a teammate may come at a greater physiological expense. This corroborates findings in performance metrics (Pairs in the High Fatigue condition had the lowest SME rating) and subjective workload that the costs of collaboration outweighed the benefits of having two heads for this specific task.

Additionally, there were trends that showed that the Top performers and Bottom performers responded differently as a function of Difficulty. Top performers showed increasing levels of LF power for High Difficulty relative to Low Difficulty, indicating higher sympathetic activation in the Hard Difficulty scenario. By catapulting themselves into a hyper-functional state, Top performers expended more physiological resources at a higher rate to minimize adverse effects on task performance. The opposite pattern was seen for bottom performers.

For the eye metrics, there was no evidence of differences due to the Difficulty manipulation. This is consistent with the Experiment 1 results as well as previous research (Di Stasi et al., 2013). There were however trends for the blink rate as a function of performance level such that the differences between Hard and Easy conditions was larger for the bottom performers than for the top performers. This pattern is consistent with observations of trends in the HRV frequency domain. The hypothesis is that this is attributed to the ability to better control and allocate attentional resources.

5.3 Teaming

Two levels of the Team independent variable were evaluated in this study. Participants either did the task alone in the Individual condition or participants worked as a Pair. Note that the Pairs were self-selected, and most Pairs knew their partner before participation in this experiment. Perhaps the most surprising result was that Individuals tended to perform better than Pairs. Also, the patterns of relative power in the LF and HF ranges was different for the Pairs than for Individuals. The Team factor seemed also to have complex impacts on eye metrics – impacting pupil diameter, fixation duration, and blink rate. We suspect that this reflects the impact of communications and joint sense making as team participants direct each other’s attention to specific information in the documents.

In general, the subjective responses on the Team Questionnaire indicated that the participants reported information sharing behaviors that are consistent with good teamwork. However, there was one anomalous team that experienced observable tension due to poor information sharing behaviors. This tension was evident in a transient response on the HRV response of one of the participants in this Pair.

The impact of the Teaming manipulation raises many questions for further research. First, there is reason to believe that the social dynamics for a Pair is not representative of the behavior of larger teams (Woolley et al., 2010), and it is important to note that the scenarios were designed so that individuals could accomplish the tasks alone. Thus, a direction for future research is to

examine different Intelligence tasks that require teamwork and to examine larger team sizes within the DCGS. The results also suggest that the social dynamics associated with teamwork can introduce increased stresses that may have a negative impact on performance. It will be interesting to examine the Team communications obtained during this study to see whether there are patterns of information sharing that can be used to differentiate high functioning Pairs from lower functioning Pairs. Future studies need to explore the social dynamics to better understand how cognitive work is coordinated and how information is shared.

6.0 RECOMMENDATIONS

Overall, the results provide evidence that both HRV and eye metrics are sensitive to Fatigue. There were also some subtle effects related to Difficulty, especially when examined as a function of performance levels. Thus, this seems to be a promising direction for continued research. As physiological sensing technologies become more sensitive and less invasive, it is likely that there is the potential to utilize this information effectively in making decisions about managing work in DCGSs.

We recommend that follow-on research explore performance over longer time periods – consistent with typical work shifts and longitudinal monitoring – responses to shift changes, adverse events, etc. Longitudinal monitoring of participants could be achieved with wearable devices that also monitor participant’s sleep patterns in addition to daily health monitoring. A multitude of factors can influence an individual’s current physiology that may not be objectively quantified, such as, but not limited to, emotional stress, sleep quality, consumption of medications, etc. These factors can modulate an individual’s physiology, and though not acutely noticeable, can have notable impacts on an individual’s readiness and performance capacity over time. Therefore, evaluation of longitudinal changes can provide a more holistic perspective of an operator and the impact of external stressors/factors on their physiology.

Also, we suggest that the size of teams be expanded beyond two, to be more representative of operational AETs. This would subsequently result in scenario modifications (e.g., increased number of intelligence documents / types) that would increase complexity to better suit examination of pair / team coordination and problem solving. Finally, we think that there are many interesting questions to be explored with respect to the impact of the social dynamics of teams on the quality of group problem solving. This research needs to track the communication activities as a potential important measure of the quality of team problem solving (e.g., Cooke et al., 2013). In addition to expansion of AET team size, exploration of remote collaboration could be interesting, especially within the context of communication across DCGS locations.

7.0 REFERENCES

- Åkerstedt, T. (2012). Karolinska Sleepiness Scale (KSS). In A. Shahid, K. Wilkinson, S. Marcu, & C. M. Shapiro (Eds.), *STOP, THAT and One Hundred Other Sleep Scales* (pp. 209–210). Springer New York. <https://doi.org/10.1007/978-1-4419-9893-4>
- Benner, P. (1984). From Novice to Expert: Excellence and Power in Clinical Nursing Practice. *AJN The American Journal of Nursing*, *84*(12), 1480.
- Borukhovich, K., & Morton, T. (2020, September 26). *DCGS Next Generation: Accelerating Change to Deliver Decision Advantage*. OTH. <https://overthehorizonmdos.wpcomstaging.com/2020/09/26/dcgs-next-generation-accelerating-change-to-deliver-decision-advantage/>
- Bryant-Lees, K. B., Prince, L., Goodman, T., Chappelle, W., & Thompson, B. (2021). Sources of Stress and Psychological Health Outcomes for Remotely Piloted Aircraft Operators: A Comparison Across Career Fields and Major Commands. *Military Medicine*, *186*(7–8), e784–e795. <https://doi.org/10.1093/milmed/usaa257>
- Chappelle, W. L., Prince, L. R., & Goodman, T. M. (2019). Sources of Stress and Psychological Health Outcomes Among U.S. Air Force Total Force Distributed Common Ground System Operators. *Military Medicine*, *184*(Supplement_1), 451–460. <https://doi.org/10.1093/milmed/usy398>
- Chen, J.-T., Kuo, Y.-C., Hsu, T.-Y., & Wang, C.-A. (2022). Fatigue and Arousal Modulations Revealed by Saccade and Pupil Dynamics. *International Journal of Environmental Research and Public Health*, *19*(15), 9234. <https://doi.org/10.3390/ijerph19159234>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive Team Cognition. *Cognitive Science*, *37*(2), 255–285. <https://doi.org/10.1111/cogs.12009>
- Davids, K., Glazier, P., Araújo, D., & Bartlett, R. (2003). Movement Systems as Dynamical Systems: The Functional Role of Variability and its Implications for Sports Medicine. *Sports Medicine*, *33*(4), 245–260. <https://doi.org/10.2165/00007256-200333040-00001>
- DeMers, D., & Wachs, D. (2023). Physiology, Mean Arterial Pressure. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK538226/>
- Di Stasi, L. L., McCamy, M. B., Catena, A., Macknik, S. L., Cañas, J. J., & Martinez-Conde, S. (2013). Microsaccade and drift dynamics reflect mental fatigue. *European Journal of Neuroscience*, *38*(3), 2389–2398. <https://doi.org/10.1111/ejn.12248>
- Di Stasi, L. L., Renner, R., Catena, A., Cañas, J. J., Velichkovsky, B. M., & Pannasch, S. (2012). Towards a driver fatigue test based on the saccadic main sequence: A partial validation by subjective report data. *Transportation Research Part C: Emerging Technologies*, *21*(1), 122–133. <https://doi.org/10.1016/j.trc.2011.07.002>
- Fonseca, D. S., Netto, A. D., Ferreira, R. B., & de Sá, A. M. F. L. M. (2013). Lomb-scargle periodogram applied to heart rate variability study. *2013 ISSNIP Biosignals and Biorobotics Conference: Biosignals and Robotics for Better and Safer Living (BRC)*, 1–4. <https://doi.org/10.1109/BRC.2013.6487524>
- Funke, G. J., Knott, B. A., Salas, E., Pavlas, D., & Strang, A. J. (2012). Conceptualization and Measurement of Team Workload: A Critical Need. *Human Factors*, *54*(1), 36–51. <https://doi.org/10.1177/0018720811427901>

- Gasparin, D., Netuveli, G., Dias-da-Costa, J. S., & Pattussi, M. P. (2009). Effect of psychological stress on blood pressure increase: A meta-analysis of cohort studies. *Cadernos de Saude Pública*, 25(4), 715–726. <https://doi.org/10.1590/S0102-311X2009000400002>
- Gillberg, M., Kecklund, G., & Åkerstedt, T. (1994). Relations Between Performance and Subjective Ratings of Sleepiness During a Night Awake. *Sleep*, 17(3), 236–241. <https://doi.org/10.1093/sleep/17.3.236>
- Grant, D. A., Honn, K. A., Layton, M. E., Riedy, S. M., & Van Dongen, H. P. A. (2017). 3-minute smartphone-based and tablet-based psychomotor vigilance tests for the assessment of reduced alertness due to sleep deprivation. *Behavior Research Methods*, 49(3), 1020–1029. <https://doi.org/10.3758/s13428-016-0763-8>
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85, 348–362. <https://doi.org/10.1037/0022-3514.85.2.348>
- Hart, S. G. (2006). Nasa-Task Load Index (NASA-TLX); 20 Years Later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9), 904–908. <https://doi.org/10.1177/154193120605000909>
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Advances in Psychology* (Vol. 52, pp. 139–183). Elsevier. [https://doi.org/10.1016/S0166-4115\(08\)62386-9](https://doi.org/10.1016/S0166-4115(08)62386-9)
- Hockey, R. J. (1997). Compensatory control in the regulation of human performance under stress and high workload: A cognitive-energetical framework. *Biological Psychology*, 45(1), 73–93. [https://doi.org/10.1016/S0301-0511\(96\)05223-4](https://doi.org/10.1016/S0301-0511(96)05223-4)
- Johns, M. W. (1991). A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep*, 14(6), 540–545. <https://doi.org/10.1093/sleep/14.6.540>
- Leiter, M.P., & Maslach, C. (2016). Latent burnout profiles: A new approach to understanding the burnout experience. *Burnout Research*, 3, 89-100. <http://dx.doi.org/10.1016/j.burn.2016.09.001>
- Mahanama, B., Jayawardana, Y., Rengarajan, S., Jayawardana, G., Chukoskie, L., Snider, J., & Jayarathna, S. (2022). Eye Movement and Pupil Measures: A Review. *Frontiers in Computer Science*, 3, 733531. <https://doi.org/10.3389/fcomp.2021.733531>
- Mandrick, K., Peysakhovich, V., Rémy, F., Lepron, E., & Causse, M. (2016). Neural and psychophysiological correlates of human performance under stress and high mental workload. *Biological Psychology*, 121, 62–73. <https://doi.org/10.1016/j.biopsycho.2016.10.002>
- Maslach, C., Jackson, S. E., Leiter, M. P., Schaufeli, W. B., & Schwab, R. L. (2018). *Maslach Burnout Inventory Manual* (Forth Edition). Mind Garden.
- Maslach, C., & Leiter, M. P. (2016). Understanding the burnout experience: Recent research and its implications for psychiatry. *World Psychiatry*, 15(2), 103–111. <https://doi.org/10.1002/wps.20311>
- 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>
- Office of the Director of National Intelligence. (2015, January 2). *Intelligence Community Directive 203*. <https://www.dni.gov/files/documents/ICD/ICD%20203%20Analytic%20Standards.pdf>

- Paprocki, R., & Lenskiy, A. (2017). What Does Eye-Blink Rate Variability Dynamics Tell Us About Cognitive Performance? *Frontiers in Human Neuroscience*, *11*, 620. <https://doi.org/10.3389/fnhum.2017.00620>
- Paskett, M. D., Garcia, J. K., Jones, S. T., Brinton, M. R., Davis, T. S., Duncan, C. C., Cooper, J. M., Strayer, D. L., & Clark, G. A. (2022). *Improving Upper-limb Prosthesis Usability: Cognitive Workload Measures Quantify Task Difficulty* [Preprint]. Rehabilitation Medicine and Physical Therapy. <https://doi.org/10.1101/2022.08.02.22278038>
- Peel, T. R., Dash, S., Lomber, S. G., & Corneil, B. D. (2017). Frontal Eye Field Inactivation Diminishes Superior Colliculus Activity, But Delayed Saccadic Accumulation Governs Reaction Time Increases. *The Journal of Neuroscience*, *37*(48), 11715–11730. <https://doi.org/10.1523/JNEUROSCI.2664-17.2017>
- Pentland, A. (2015). *Social Physics: How Social Networks Can Make Us Smarter*. Penguin.
- Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35. <https://doi.org/10.1016/j.pneurobio.2015.09.001>
- Preece, D. A., Becerra, R., Robinson, K., & Gross, J. J. (2020). The Emotion Regulation Questionnaire: Psychometric properties in general community samples. *Journal of Personality Assessment*, *102*(3), 348–356. <https://doi.org/10.1080/00223891.2018.1564319>
- Schaufeli, W. B., Salanova, M., González-romá, V., & Bakker, A. B. (2002). The Measurement of Engagement and Burnout: A Two Sample Confirmatory Factor Analytic Approach. *Journal of Happiness Studies*, *3*(1), 71–92. <https://doi.org/10.1023/A:1015630930326>
- Schleicher, R., Galley, N., Briest, S., & Galley, L. (2008). Blinks and saccades as indicators of fatigue in sleepiness warnings: Looking tired? *Ergonomics*, *51*(7), 982–1010. <https://doi.org/10.1080/00140130701817062>
- Skibniewski, F. W., Dziuda, Ł., Baran, P. M., Krej, M. K., Guzowski, S., Piotrowski, M. A., & Truszczyński, O. E. (2015). Preliminary Results of the LF/HF Ratio as an Indicator for Estimating Difficulty Level of Flight Tasks. *Aerospace Medicine and Human Performance*, *86*(6), 518–523. <https://doi.org/10.3357/AMHP.4087.2015>
- Spielberger, C. D. (2020). *State-Trait Anxiety for Adults(TM) Manual* (Forms Y1 and Y2). Mind Garden.
- Tanaka, H., & Tamura, N. (2016). Sleep education with self-help treatment and sleep health promotion for mental and physical wellness in Japan. *Sleep and Biological Rhythms*, *14*(1), 89–99. <https://doi.org/10.1007/s41105-015-0018-6>
- Tansakul, W., & Tangamchit, P. (2015). Fatigue Driver Detection System Using a Combination of Blinking Rate and Driving Inactivity. *Journal of Automation and Control Engineering*, *3*(6), 33–39. <https://doi.org/10.12720/joace.4.1.33-39>
- Uphill, M. A., Rossato, C. J. L., Swain, J., & O’Driscoll, J. (2019). Challenge and Threat: A Critical Review of the Literature and an Alternative Conceptualization. *Frontiers in Psychology*, *10*, 1255. <https://doi.org/10.3389/fpsyg.2019.01255>
- Vicente, J., Laguna, P., Bartra, A., & Bailón, R. (2016). Drowsiness detection using heart rate variability. *Medical & Biological Engineering & Computing*, *54*(6), 927–937. <https://doi.org/10.1007/s11517-015-1448-7>
- Vrijkotte, T. G. M., Van Doornen, L. J. P., & De Geus, E. J. C. (2000). Effects of Work Stress on Ambulatory Blood Pressure, Heart Rate, and Heart Rate Variability. *Hypertension*, *35*(4), 880–886. <https://doi.org/10.1161/01.HYP.35.4.880>

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a Collective Intelligence Factor in the Performance of Human Groups. *Science*, 330(6004), 686–688. <https://doi.org/10.1126/science.1193147>

APPENDIX A. PARTICIPANT CHARACTERISTICS

A.1. Demographics

A.1.1. Age, Sex, and Work Experience

The two tables below describe the participants by their Team condition with respect to their age, biological sex assigned at birth, and years of experience in their current job.

Table A.1. Participant age, sex, and work experience (Individual)

	Count	Age			Years of Experience		
		Mean	<i>SD</i>	Range	Mean	<i>SD</i>	Range
Female	4	26.0	4.5	22–33	5.0	3.2	2–10
Male	12	27.3	7.1	19–43	4.0	4.1	1–15
Total	16	27.0	6.5	19–43	4.3	3.8	1–15

Table A.2. Participant age, sex, and work experience (Paired)

	Count	Age			Years of Experience		
		Mean	<i>SD</i>	Range	Mean	<i>SD</i>	Range
Female	9	26.3	8.4	18–44	4.9	5.3	1–19
Male	15	25.2	4.6	19–35	2.5	2.3	1–9
Total	24	25.6	6.2	18–44	3.4	3.4	0–19

Ninety percent of participants reported having experience working within an AET construct (see Table A.3). Given that most participants reported that their units currently work using the AET format and that all participants seemed to have come from the same couple of military units, it is possible that the four ‘no prior AET experience’ responses are inaccurate. Regardless, the high percentage rate implies that most of the participant sample should generally be familiar with how AETs operate and therefore have some familiarity with how the experimental task would be structured and presented.

Table A.3. Self-reported AET experience

AET Experience	Team	
	Individual	Pair
No	3	1
Yes	13	23
Total	16	24

A.1.2. Job and Rank

With respect to participant jobs as identified by their Air Force Specialty Codes (AFSC), the most common job was 1N1, or Geospatial Intelligence Analyst (GEOINT), for both the Individual and Paired Team conditions (see Figure A.1). For the Individual condition, the second most common job was 1N4X2, or Cryptologic Analyst & Reporter (CAR). For the Paired condition, the second most common job was 1N0, or All Source Intelligence Analyst (ALL).

The two most common ranks within the Individual condition were E-4 (Senior Airman; SrA) and E-3 (Airman First Class; A1C). Within the Paired Team condition, the two most common ranks were E-3 and E-5 (Staff Sergeant; SSgt).

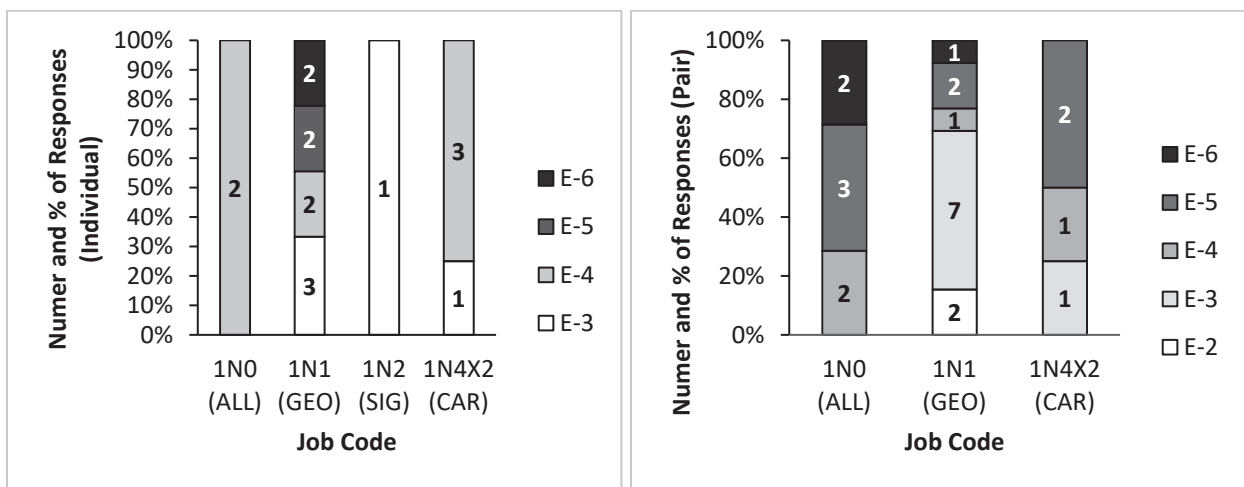


Figure A.1. AFSC jobs by rank as a percentage for participants in the Individual condition (left) and Paired Team (right).

A.1.2. Education

As seen in Figure A.2, 28 (70%) participants reported receiving some level of education or training beyond a high school diploma/GED. The most common college major was Intelligence Studies, followed by three-way tie between Criminal Justice, Cybersecurity, and Psychology, and then Computer Science and Management. Ten participants were categorized as Other in the figure below because their responses were not easily categorized with each other or with the other specified areas of study (e.g., massage therapy, athletic training, etc).

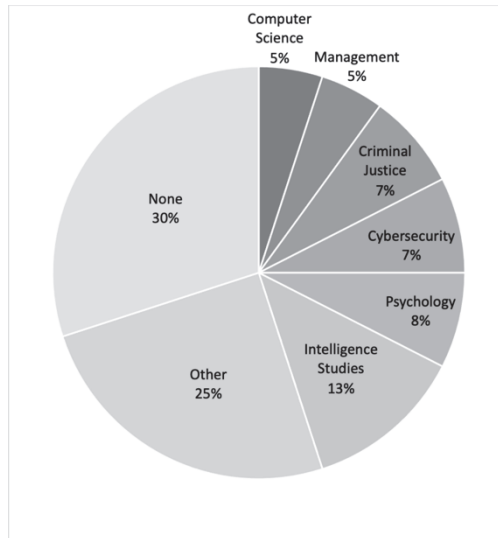


Figure A.2. Frequency of reported educational majors.

The graphs in Figure A.3 below display the distribution of educational attainment level by job.

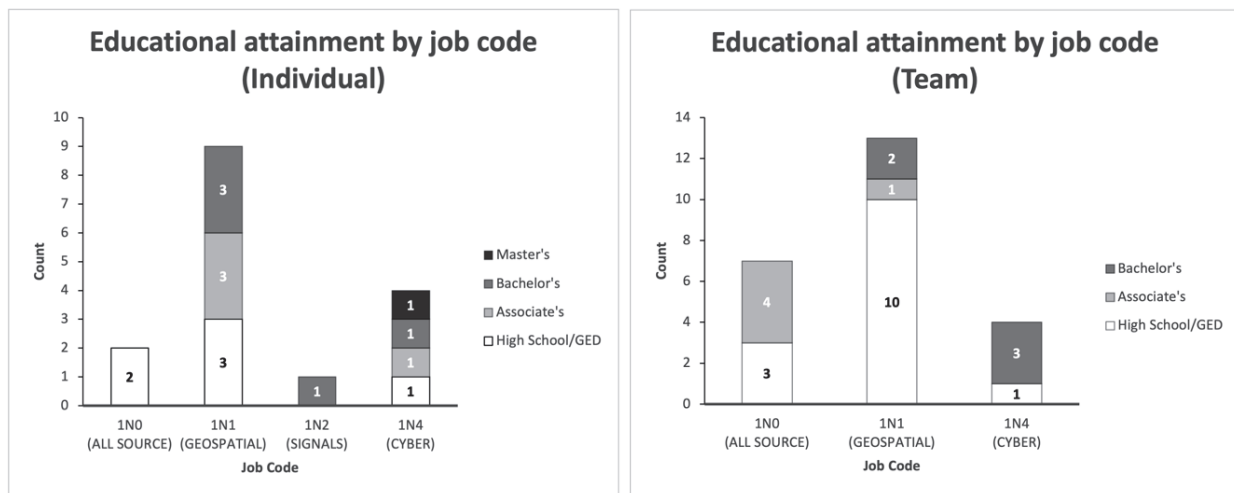


Figure 22. Educational attainment by job code for those in the Individual condition (left) and by those in the Paired Team condition (right).

A.2. Intake Questionnaire

After completing a demographical survey, participants completed an intake questionnaire during each session to provide additional context for understanding and interpreting participant performance.

A.2.1. Visual Correction

The cameras on the eye trackers require unobstructed sight to participants' eyes. As such, participants who could not see a computer screen clearly without contact lenses were excluded

from the study. For the Individual Team condition, eight of 16 participants (50%) required corrected vision. For the Paired Team condition, eight of 24 participants (33%) required correction.

A.2.2. Rest Amount and Quality

“How many hours of sleep did you get in the past 24 hours?”

Response options included: *Less than 1 hour or none, 1–3 hours, 4–6 hours, 7–9 hours, and More than 9 hours.* As seen in the two graphs in Figure A.4 below, most participants indicated having slept between 7–9 hours followed by those claiming 4–6 hours within the last 24-hours. Those that claimed getting more than 9 hours never accounted for more than 10% of the total.

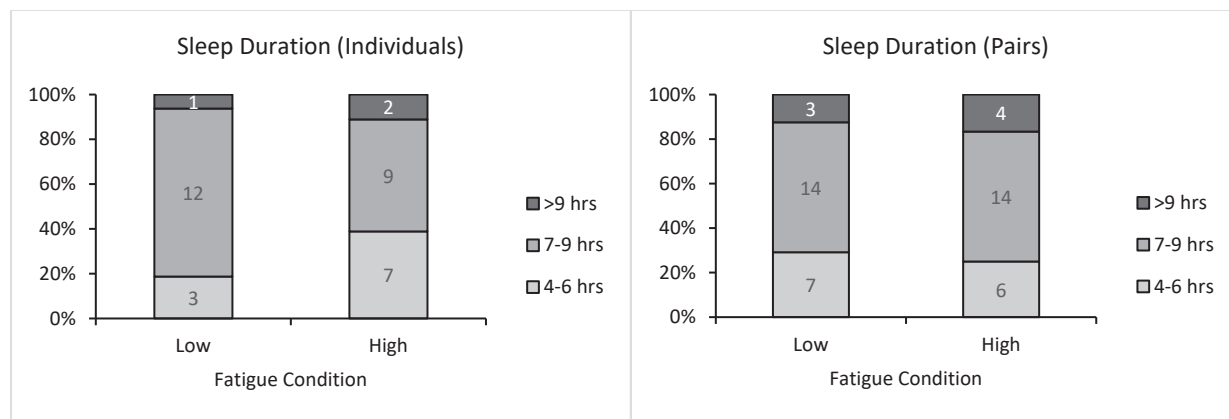


Figure 23. Reported hours slept for those in the Individual condition (left) and those in the Paired Team condition (right).

“How did you sleep last night?”

Participants were also asked to rate the quality of the sleep they received. Response options, in order of increasing quality, included Very Poor, Poor, Average, Good, and Excellent. All participants in the Individual condition rated the quality of their sleep as Average or better. Except for the 3 participants in the Paired condition who rated their sleep quality as Poor, all other Pairs also tended to rate the sleep quality as Average or better.

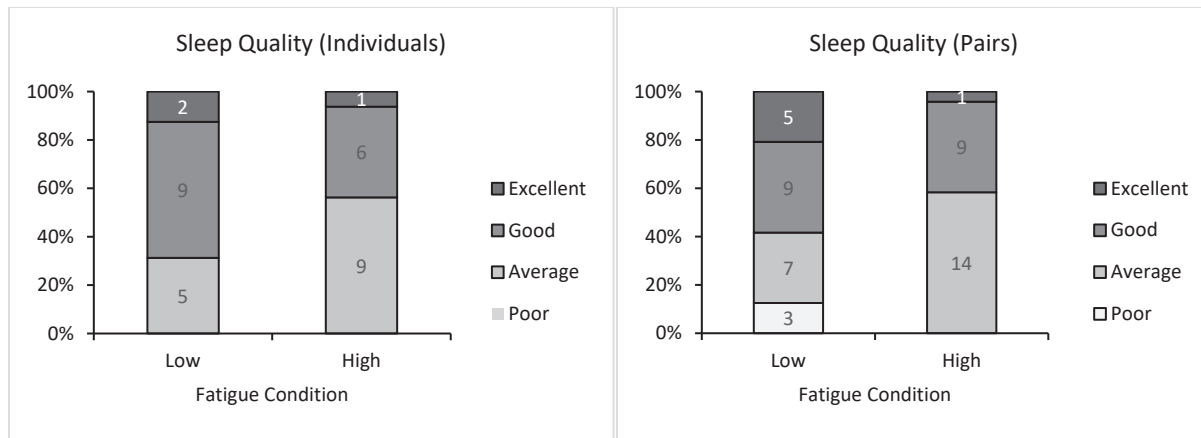


Figure A.5. Number of participants by sleep quality and fatigue condition for those in the Individual condition (left) and those in the Paired Team condition (right).

A.2.3. Physical Activity

“What level of physical activity did you experience today?”

Participants response options included: None, Minimal, Moderate, and High.

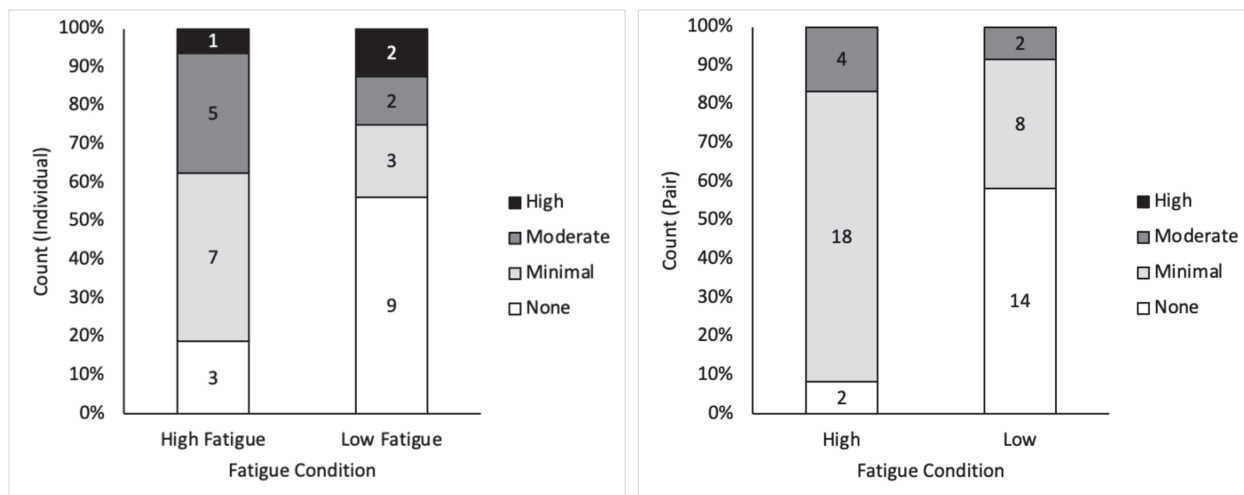


Figure 24. Self-reported physical activity in the last day for Individuals (left) and in Paired Teams (right).

“How active were you during your shift?”

Response options included: Low, Moderate, and High.

As part of a manipulation check, participants were asked to state whether they had just come off a shift. Three participants in the Individual cohort reported having come into their scheduled Low Fatigue session after having completed their shift in opposition to the instructions given by the research team during the scheduling process. It is unclear whether these three participants

worked prior to their Low Fatigue session but they were also the only three to have indicated experiencing Low or Moderate physical activity during their shift.

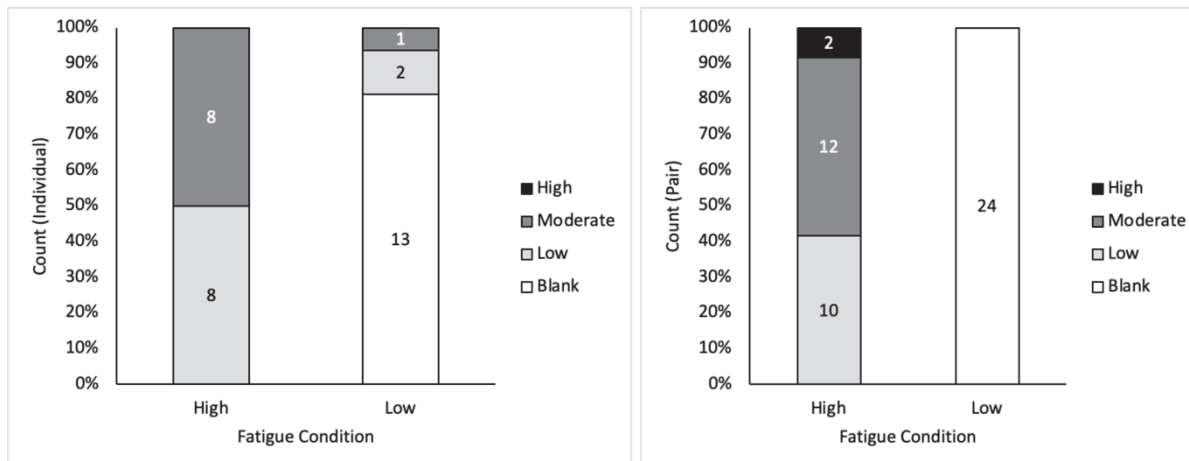


Figure 25. Self-reported physical activity during the last shift for those in the Individual condition (left) and in the Paired Team condition (right).

A.2.4. Stimulant Use

“Have you taken a prescription stimulant (Adderall, Ritalin, etc.) in the last 12 hours?”

No participants reported taking prescription stimulants prior to their participation in the study.

“How much caffeine have you consumed in the last 6 hours?”

Response options included: <100 mg, 100-200 mg, 200-300 mg, and 300+ mg. These options also included a reference to the equivalency to the number of cups of coffee and cans of the Red Bull energy drink. Red Bull was selected as a reference point because our SMEs mentioned that it was a common beverage amongst analysts within the DCGS.

Most participants also denied caffeine use but there were still many that admitted some use within the preceding six hours prior to their session. Caffeine use can be seen in Figure A.8 below. It is unclear why participants reported greater caffeine consumption in the Paired Team condition at levels that cannot be attributed simply to sample size, particularly during the Low fatigue session. We suspect that it might be due, in part, to the only three individuals (who also happened to be in the Paired Team condition) to report poor quality sleep trying to compensate by increasing their caffeine consumption.

Participants were also asked to respond to the amount of caffeine consumed that day, but their responses did not seem to correspond to the reported caffeine consumption in the last 6 hours. For example, some participants stated that they had consumed 100-200 mg within the last 6 hours but then responded that they had consumed no caffeine that day. Given the discrepant responses covering overlapping time periods, only the 6-hour period is reported here with the assumption that it would be the least open to misinterpretation about what constitutes a ‘day’ and would, therefore, be most accurate.

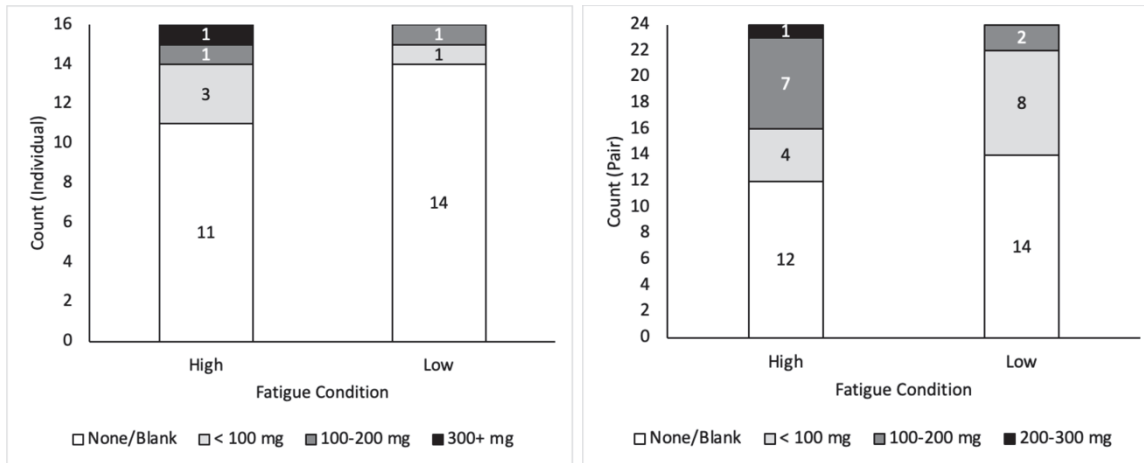


Figure A.26. Self-reported caffeine use over the last 6-hours for those in the Individual condition (left) and those in Paired Team condition (right).

APPENDIX B. JADE SOFTWARE

This section describes the JADE software as was configured for Experiment 2. It follows the sequence of screens, with only a few noted deviations, that participants would have followed to complete one study session.

Upon signing into the JADE software, participants were presented with the following screenshot of three modules to select from. Research facilitators provided instruction on the proper selection sequence.

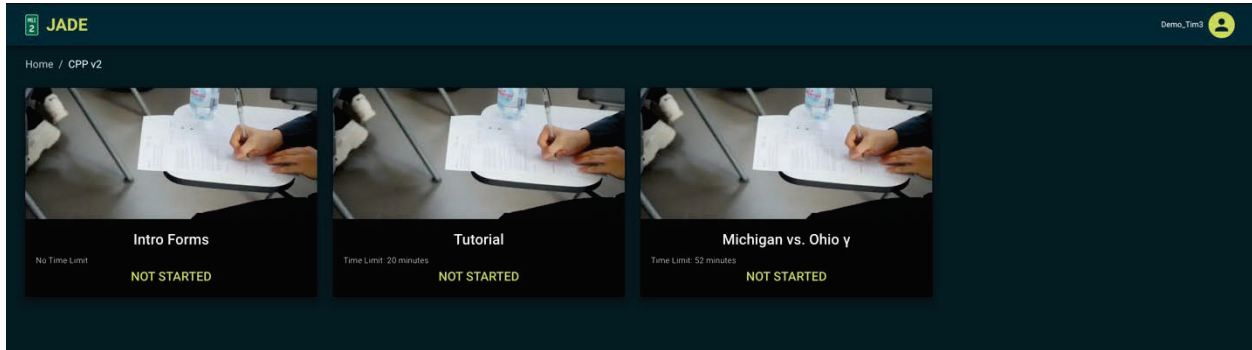


Figure B.1. Partial screenshot of the three modules used in each session.

Participants were first instructed to select the “Intro Forms” module to complete the Demographics and Intake surveys (see Figure B.2 and Figure B.3, respectively).

A screenshot of the 'Intro Forms' module in the JADE software. The breadcrumb path is 'Home / CPP v2 / Intro Forms'. The main heading is 'Please tell us a little about your background and training'. The form contains several input fields with the following data: 'Please type in your age (in years) *' with the value '24'; 'Please select the sex you were assigned at birth *' with the value 'Male'; 'Please select your highest level of educational attainment *' with the value 'Bachelor's Degree'; 'If you attended any college, please enter your most recent major *' with the value 'Intelligence studies'; 'Current or most recent rank held *' with the value 'E-4'; 'Select the service that you last served as an analyst *' with the value 'Air Force'; and 'AFSC, MOS, Rating/Designator, or other career designation *' with the value '1N0'.

Figure B.2. Partial screenshot of the Demographics survey.

Figure B.3. Partial screenshot of the Intake survey.

After completing the Demographics and Intake surveys, participants were instructed to wait for a research facilitator to fit all physiological sensors and record baseline physiological measurements (see Figure B.4).

Figure B.4. Instructions and notification to participants that initial baseline measurements (Pre-task) will be recorded shortly.

Following the baseline physiological measurements, participants completed the KSS, PVT, and ESS (see Figure B.5, Figure B.6, and Figure B.7, respectively) to complete the Intro Forms module.

Home / CPP v2 / Intro Forms

KSS

Which level best reflects your current sleepiness?? *

- Extremely Alert
- Very Alert
- Alert
- Rather Alert
- Neither Alert nor Sleepy
- Some Signs of Sleepiness
- Sleepy, but No Effort to Keep Awake
- Sleepy, but Some Effort to Keep Awake
- Very Sleepy, Great Effort to Keep Awake, Fighting Sleep
- Extremely Sleepy, Can't Keep Awake

Akerstedt, T., & Gillberg, M. (1990). Subjective and objective sleepiness in the active individual. *International Journal of Neuroscience*, 52(1-2), 29-37.

SUBMIT

Step 4 of 6

Figure 27. Screenshot of the KSS as implemented within JADE.

Home / CPP v2 / Intro Forms

PVT

For this assessment we are going to measure your reaction time to numbers presented on the screen. You will need to press the space bar as quickly as you can when you see the numbers appear on screen.

1. Visit [PVT web app](#) and follow the instructions to take your PVT test.
2. Once informed to stop your PVT by your facilitator, exit by either clicking the X in the top right corner or hitting escape on the keyboard

4701

result_1689263465619.

Step 5 of 6

Figure 28. Instructions for completing the PVT.



Figure 29. Instruction screen for the ESS.

Upon returning to the module selection screen, participants were instructed to select the one of the remaining options (see Figure B.8). For the first session, participants were instructed the Tutorial module to receive training on the task and JADE user interface. For the second session, participants skipped the Tutorial to open the designated experimental scenario.

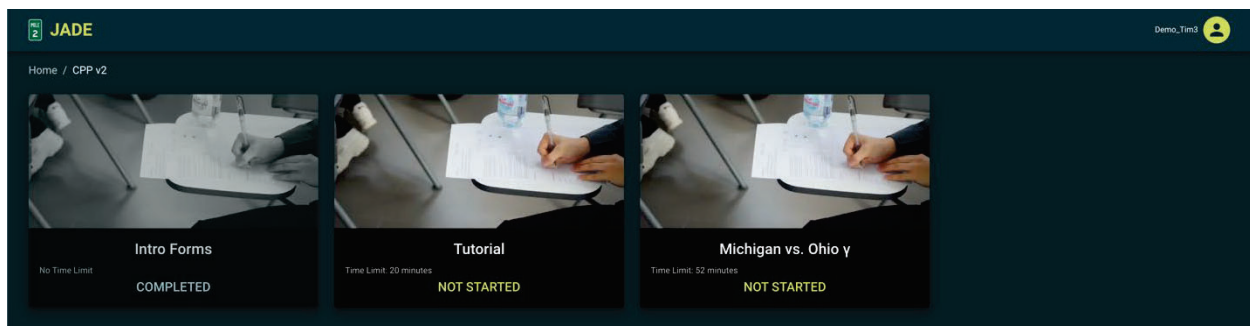


Figure 30. Partial screenshot of the module selection screen with the Intro Forms module ‘grayed out’ to indicate that it has been completed.

Given that the Tutorial module is identical to the experimental scenario (with the minor exception of a shorter duration and fewer Intelligence documents to review), only screenshots for the experimental scenario will be provided. Once selected, participants were again prompted to await a facilitator to collect pre-task, anticipatory physiological measurements (image not provided). After collecting those measurements, participants were provided with basic instructions for the experimental task (see Figure B.9) that served to reinforce the instructions previously provided.

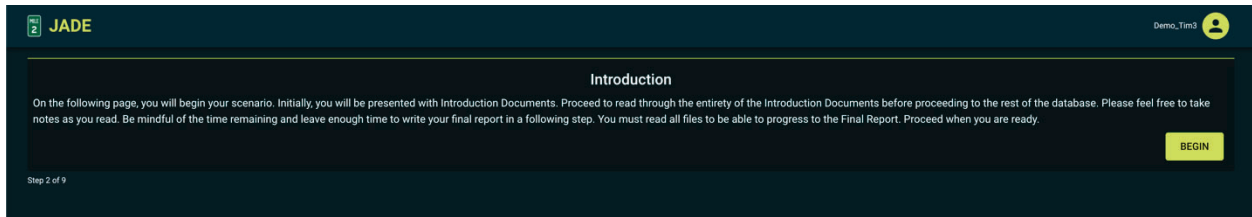


Figure B.9. Overview and basic instructions for the experimental task.

Once the participant clicked on the Begin button from the Introduction screen, participants were presented with the primary workspace within JADE (see Figure B.10). The primary workspace is initially divided into three areas: (1) Time Remaining clock (upper left); (2) Notepad (lower left); and (3) Intel Database (right side).

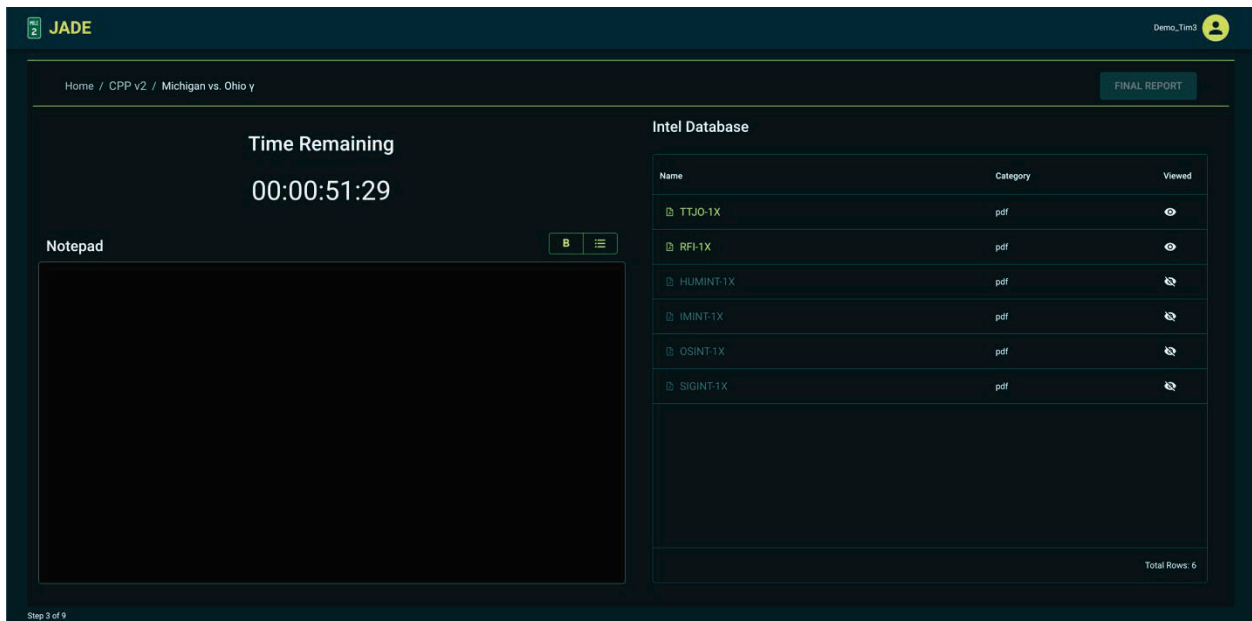


Figure 31. Screenshot of the primary workspace within JADE.

Before participants could proceed, they had to first view all background documents highlighted in the Intel Database. If a document has not been opened, the icon had a slash through it. If the document was not yet available to view, it would be ‘grayed out.’ Once all background documents had been reviewed, the remaining documents would all be highlighted and ready for viewing (see Figure B.11). In this experiment, the file names were named after the Intelligence discipline that generated the respected document. Category displays the file extension. The Viewed column to the right indicates whether the document has been opened/viewed. In this case, all documents have been opened as indicated by the eye icon.

Intel Database

Name	Category	Viewed
TTJO-1X	pdf	
RFI-1X	pdf	
HUMINT-1X	pdf	
IMINT-1X	pdf	
OSINT-1X	pdf	
SIGINT-1X	pdf	

Total Rows: 6

Figure B.11. Closeup view of the Intel Database.

An example of a background document can be seen in Figure B.12 while an example of an Intelligence document can be seen in Figure B.13.

The screenshot shows the JADE (Joint Adversary Detection Environment) interface. On the left, a Notepad window contains placeholder text: "Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Etiam dignissim diam quis enim lobortis scelerisque fermentum. Ac tortor vitae purus faucibus ornare suspendisse sed nisi lacus. Consectetur adipiscing elit ut aliquam. Feugiat in ante metus dictum at. Varius sit amet mattis vulputate enim nulla aliquet porttitor. Et sollicitudin ac orci phasellus egestas tellus rutrum tellus pellentesque. Suspendisse interdum consectetur libero id faucibus nisi tincidunt eget. Eu mi bibendum neque egestas congue quisque. In aliquam sem fringilla ut morbi tincidunt augue interdum velit. Penatibus et magnis dis parturient montes nascetur ridiculus." On the right, the "Resource Thoughts" window is active, displaying a map titled "Regional Threats to Joint Operations" which highlights the Great Lakes region (Upper Michigan, New Wisconsin, Pure Michigan, Illinois, Indiana, Ohio). Below the map are several flags, including the Michigan state flag and the United States flag. The interface also shows a "Time Remaining" of 00:00:50:04 and a "FINAL REPORT" button.

Figure B.12. Screenshot of the first background Intelligence document for this scenario – Threat to Joint Operations (TTJO) with sample text in the Notepad area.

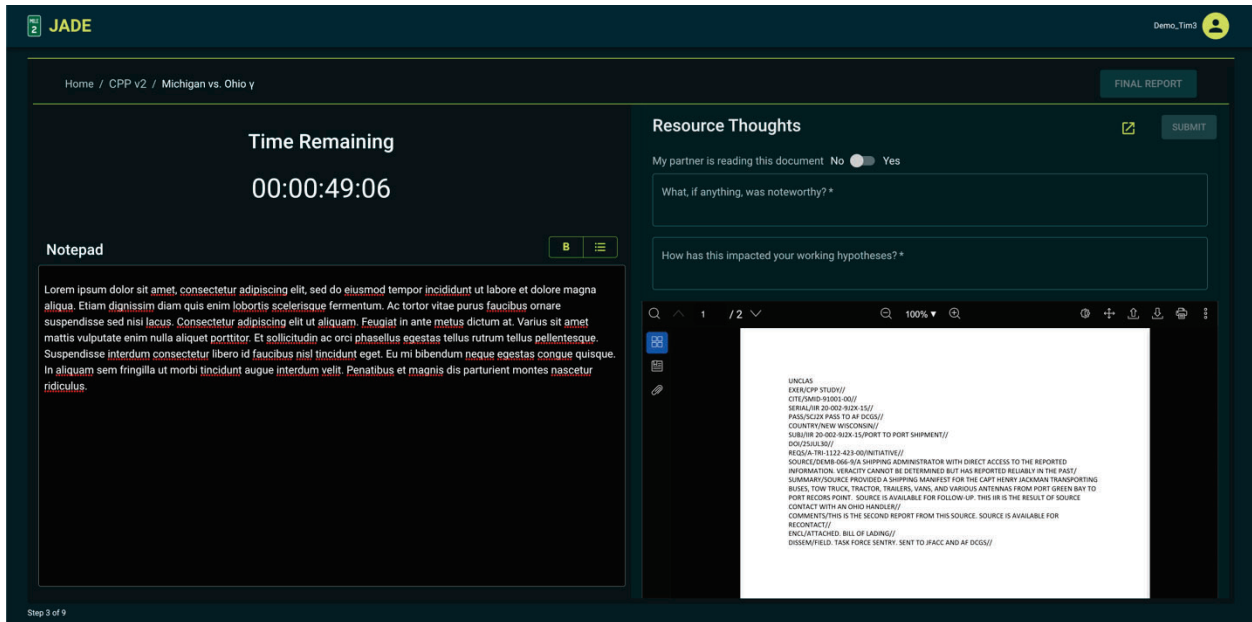


Figure 32. Screenshot of an open Intelligence document on the right with questions above requiring a response before allowing the participant to continue.

Once all documents have been viewed, participants begin the process of drafting the Report by clicking on a “Final Report” button. At that point, a small window appeared in the center of the screen and provided a set of instructions for the report (see Figure B.14).

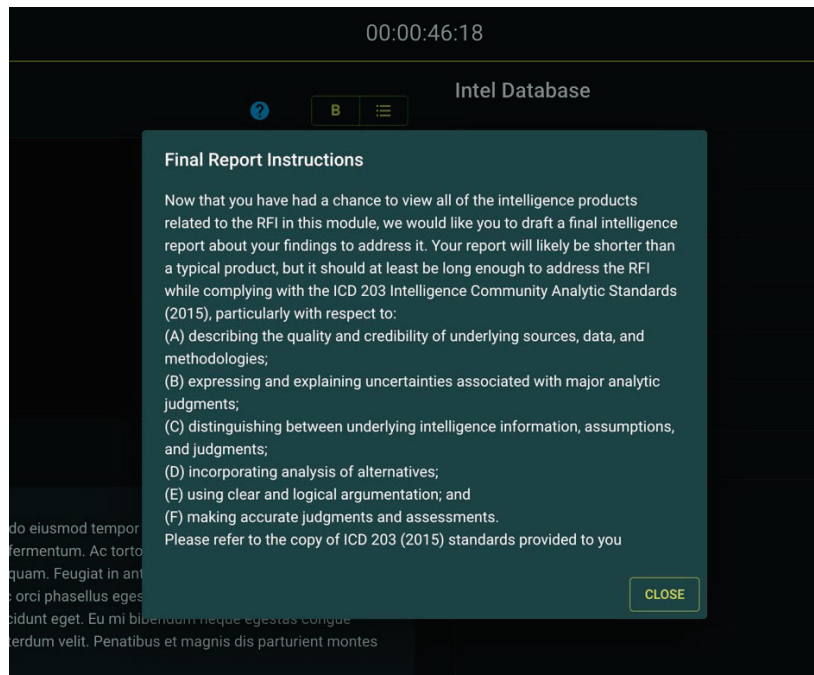


Figure 33. Closeup of the instructions given for completing the Report (i.e., Final Report).

After reading the instructions, participants began to write their report within the workspace area (see Figure B.15). Any notes in the Notepad remained available for viewing during the report writing process. Note that in Figure B.15, the Time Remaining clock has decreased in size and has moved to the top center of the screen. The area for drafting the Report is now located where the clock had been previously.

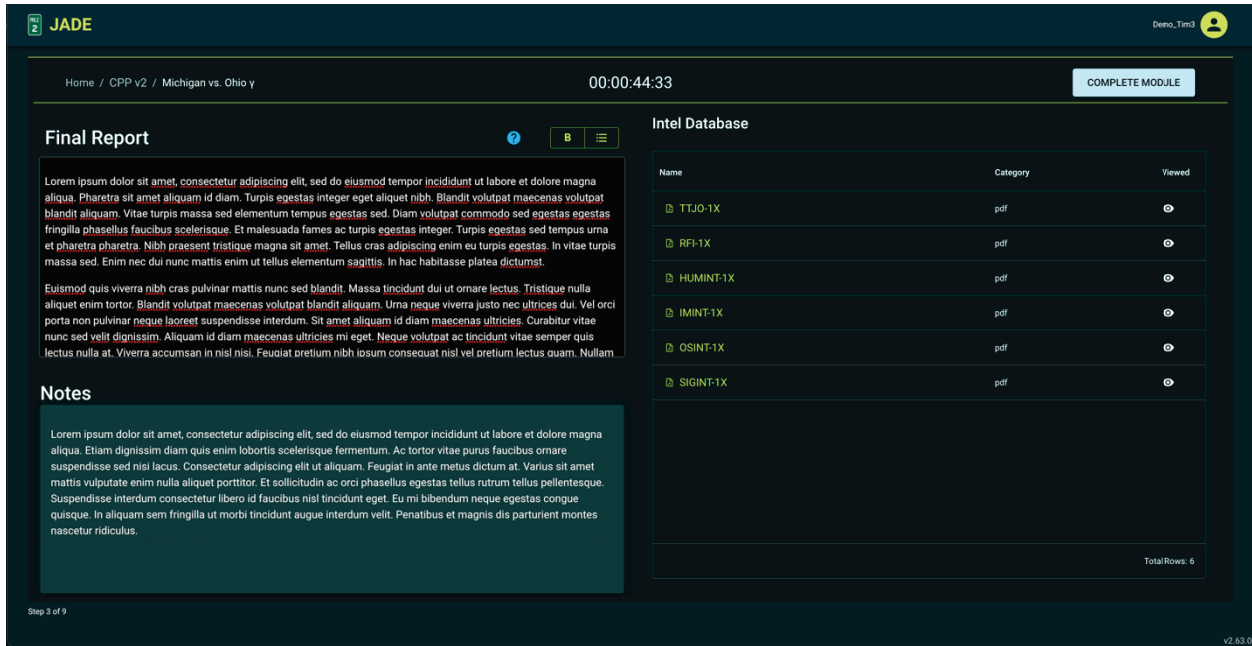


Figure 34. Screenshot of a Report in progress.

Once the Report had been submitted, participants were again instructed to wait for a facilitator to record post-task, physiological recovery measurements. Following those measurements, participants completed the post-task KSS (image not provided) and then moved on to the NASA TLX (see Figure B.16).

JADE Demo_Tim3

TLX

How mentally demanding was the task? *

Low High

How physically demanding was the task? *

Low High

How hurried or rushed was the pace of the task? *

Low High

How successful were you in accomplishing what you were asked to do? *

Good Poor

How hard did you have to work to accomplish your level of performance? *

Low High

How insecure, discouraged, irritated, stressed, and annoyed were you? *

Low High

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) *Human Mental Workload* Amsterdam: North Holland Press.

SUBMIT

Step 7 of 9

Figure 35. Screenshot of the NASA TLX as implemented within JADE.

Following the NASA TLX, participants completed the PVT once again (image not provided) and then moved on to complete the Difficulty survey (see Figure B.17). Once these final assessments had been concluded, the experimental session was considered to have been complete.

JADE Demo_Tim3

Final Thoughts

How confident were you in your assessment of this module? *

Very Confident Confident Unconfident Very Unconfident

How would you rate the difficulty of this module? *

Easy Moderately Easy Hard Extremely Hard

Please briefly describe in 1-2 sentences any feedback regarding this module *

Turpis egestas integer eget aliquet nibh. Blandit volutpat maecenas volutpat blandit aliquam. Vitae turpis massa sed elementum tempus egestas sed. Diam volutpat commodo sed egestas egestas fringilla phasellus faucibus scelerisque. |

SUBMIT

Step 9 of 9

Figure 36. Screenshot of the Difficulty survey.

APPENDIX C. TEAM SURVEY

At the end of each experimental session, participants in the Pair Condition filled out an 8-item Team Survey related to team performance. This survey was completed twice - once for each scenario. The results of questions 2 & 3 are presented in Section 0. There results for the remaining questions are presented here.

Question 1 asked, “*How well did you know your partner prior to doing the task?*” Figure shows the distribution of responses as a function of the Fatigue condition. Note that most of the Pairs knew their partner Well or Very Well. This reflects the fact that most of the Pairs signed up together. That is, partners were typically self-selected, rather than assigned.

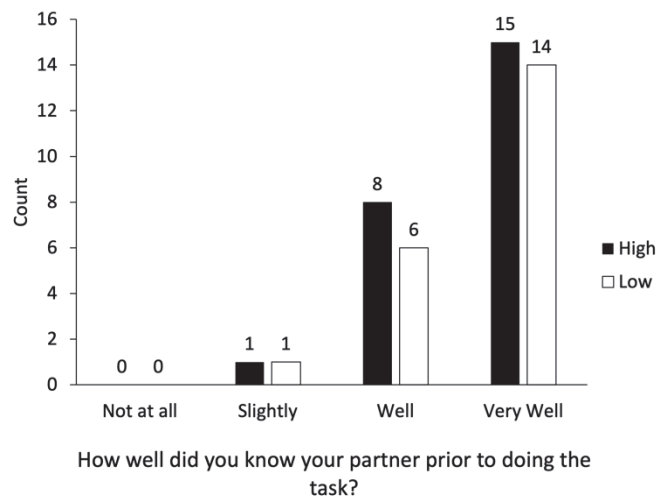


Figure C.1. The distribution of responses indicating that most of the Pairs included people who knew each other Well or Very Well prior to participating in the experiment. Bar color represents Fatigue condition.

The remaining five questions relate to team processes. A key attribute of high performing teams involves information sharing (Cooke et al., 2013; Pentland, 2015). For example, a key metric for team performance is the Push/Pull ratio which reflects the extent to which teammates anticipate the information needs of their partners and volunteer that information before being asked for it. Figure C.2 shows the distribution of response to three questions related to information sharing. These questions (top left, top right, bottom) addressed the importance of information sharing for team performance. Bar color represents the Fatigue condition from when the response was given. Overall, participants recognized the value of pushing information to their teammates and listening to the opinions of their teammates for effective team performance.

The responses indicate that most of the participants anticipated their partners information needs and volunteered information before their partner asked for it. The only team that that indicated that their partner only shared information when asked for it was the same team that was unsatisfied with the team performance. This team also had the three responses to the question about how well their partner listened to them that were less than well. The two responses

indicating that they only shared information with their partner when asked came from another team – but both responses came from the same team. Note that these questions are self-reported, but the general conclusion is that most of the participants recognized the importance of anticipating their partners information needs and pushing information to their teammate before being asked.

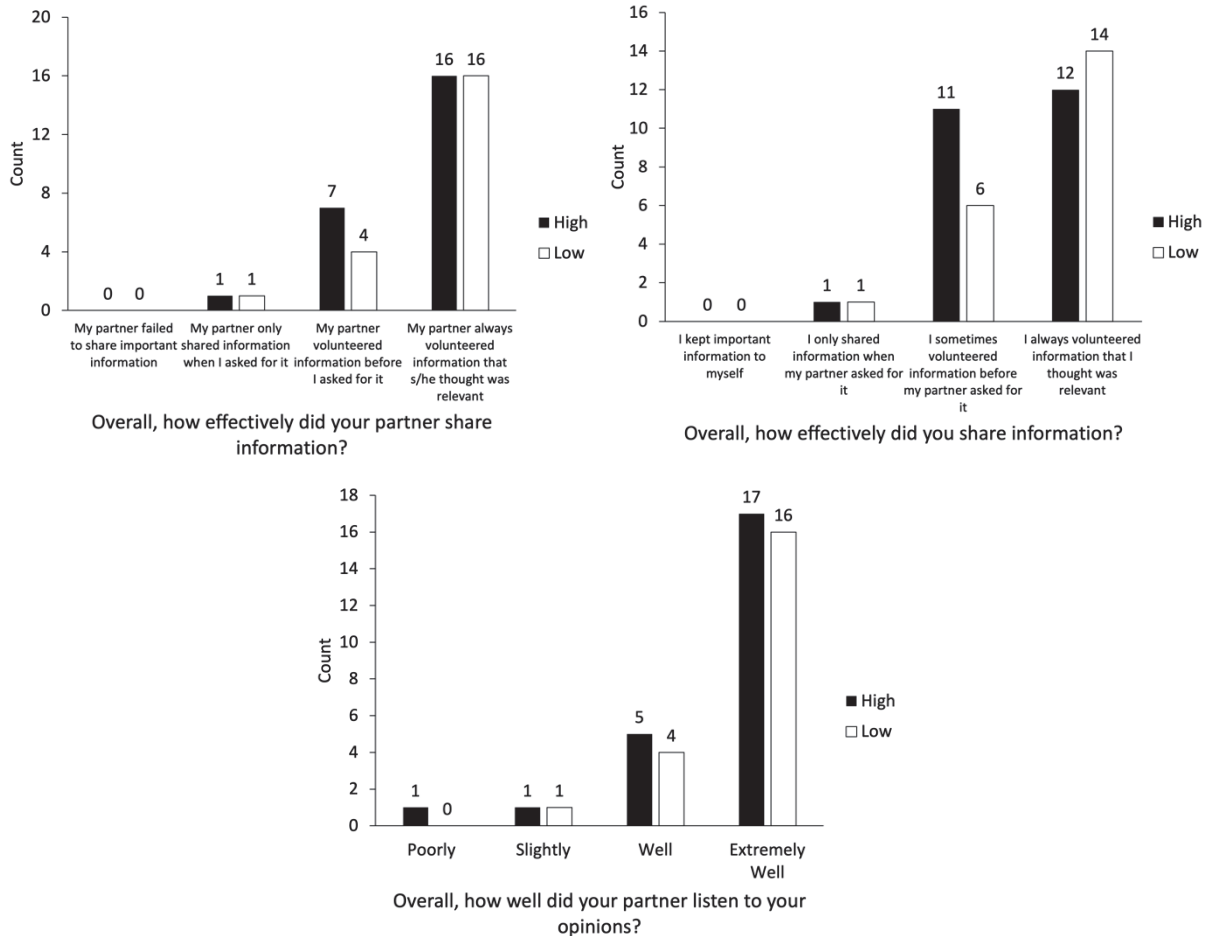


Figure C.2. Responses to Team Information Sharing Questions

A second important attribute of team performance is the work was distributed within the team. The final two questions, shown in Figure C.3, addressed how evenly the workload was distributed within the Pairs. Most of the responses indicated that the workload of both analyzing the information and creating the final Report was shared evenly. This is even though only one person was responsible for creating the joint final report and submitting it.

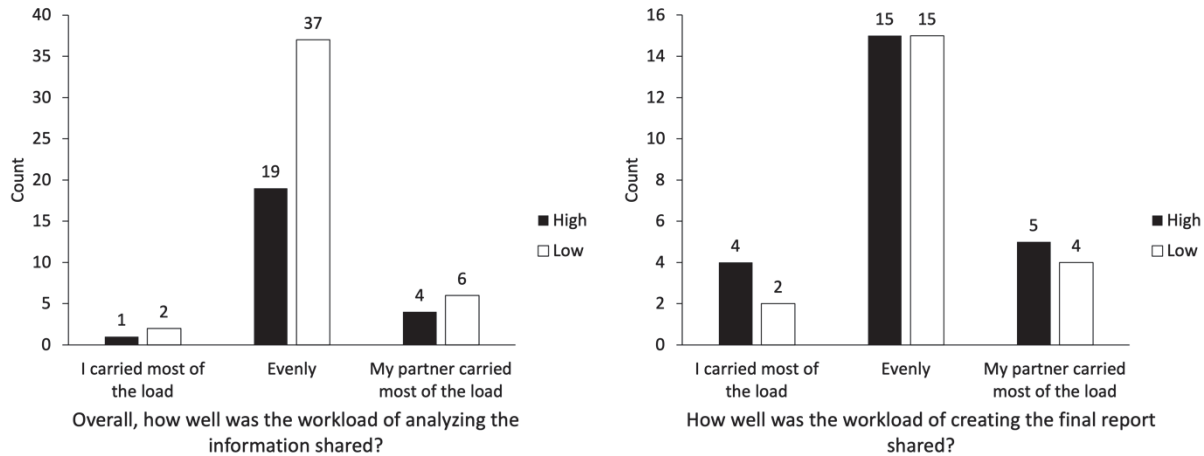


Figure 37. The distribution of responses to questions about sharing the workload. Bar colors represent Fatigue condition.

With respect to sharing the workload of analyzing the information, two different participants indicated that they carried most of the load. It is interesting to note that both of their partners responded that the workload was evenly shared. Five different people responded that their partners carried most of the load for analyzing the information (one participant indicated that for both the High and Low Fatigue conditions). Again, their partners all responded that the workload was shared evenly.

At this point, there are several hypotheses that might account for instances where the workload was not shared evenly. One possibility is that it reflects a difference in skill or experience. Seven people who indicated that the workload was not shared evenly, did so for only one of the two sessions. This might reflect the fact that the two problems required different skills (only one involved imagery analysis). So, a participant who had the right skills or experience might have contributed more to the problem solving. Also, since these are subjective reports – other factors might have led to these responses – a biased perspective might lead to over-estimating your own workload; or a false humility might lead to attributing more effort to a teammate.

With respect to sharing the workload of creating the final report, five different participants indicated that they carried most of the load (one indicated this for both High and Low Fatigue conditions). Eight different participants indicated that their partner carried most of the load (one indicated this for both Fatigue conditions). Note that only one person typed the final Report into the computer to represent a joint analysis and assessment. For two of the teams the responses were reciprocal or complementary. For both sessions for one team, one participant indicated that their partner carried most of the load and the other participant in the Pair agreed that they carried most of the load. And another team did the same for one session. For the remaining responses the partners indicated that the workload was shared evenly.

APPENDIX D. COMPARISON BETWEEN THE TOBII PRO GLASSES 2 & PUPIL CORE EYE TRACKING GLASSES

The following table compares the Tobii Pro Glasses 2 to the Pupil Labs Pupil Core eye tracking glasses. The purpose of this table is to provide readers with information about the two devices based on the experience of the Mile Two research team. The Tobii Pro Glasses 2 was used for Experiment 1. The Pupil Core glasses were used for Experiment 2.

Table 11. Comparison between the Tobii Pro Glasses 2 & Pupil Core Eye Tracking Glasses.

Feature	Tobii Pro Glasses 2	Pupil Core
Availability	Discontinued model	Available (as of June 2023)
API	No longer supported	Currently supported; enables access to controls and raw data via publicly available API
Cost	Expensive	Cheaper
Weight	45g; some participants reported discomfort and remarked on their heavy weight	23g; no participants reported any discomfort and some remarked that they forgot they had been wearing them
Mobility	Less restrictive; able to move freely within cable length distance from the computer	More restrictive; unable to move more than about 3' from the computer without needing recalibration
Metrics	Calculates many common metrics; opaque algorithms make it difficult to determine how metrics are calculated	Larger catalog of raw and calculated metrics; however, the blink rate measure produced lower rates than expected
Compatibility with Glasses	Not compatible but has a set of swappable lenses from a separately available lens kit	Not compatible
Camera Adjustability	Cameras cannot be repositioned, adjusted, or manually focused	Adjustable cameras

LIST OF ACRONYMS, ABBREVIATIONS, AND SYMBOLS

1N0 – AFSC for All source Intelligence analyst (ALL)
1N1 – AFSC for Geospatial Intelligence analyst (GEO)
1N2 – AFSC for Signals Intelligence analyst (SIG)
1N4X2 – AFSC for Cryptologic analyst & reporter (CAR)
ACC – Air Combat Command
AET – Analysis Exploitation Teams
AFRL – Air Force Research Laboratory
AFSC – Air Force Specialty Code
AFSOC – Air Force Special Operations Command
ALL – All-source Intelligence analyst
ANOVA – Analysis of Variance
ANS – Autonomic nervous system
AOD – Air Operations Directive
API – Application programming interface
CAR - Cryptologic Analyst & Reporter
CPP – Cognitive and Physiological Performance
DCGS – Distributed Common Ground Systems
DTIC – Defense Technical Information Center
ECG – Electrocardiogram
ERQ - Emotion Regulation Questionnaire
ESS – Epworth Sleepiness Scale
FMV – Full motion video
FRAGO – Fragmentary orders
GEO – Geospatial intelligence analyst
GUI – Graphical user interface
HF – High frequency
HFA_D – High frequency differences between Task Difficulty conditions
HFA_F – High frequency differences between Fatigue conditions
HR – Heart rate
HRV – Heart rate variability
HUMINT – Human Intelligence

IBI – Inter-beat-interval

ICD 203 – Intelligence Community Directive 203

IMINT – Imagery Intelligence

ISR – Intelligence, surveillance, and reconnaissance

JADE – Joint Activity Design Evaluation

KSS – Karolinska Sleepiness Scale

LF – Low frequency

LF/HF – Ratio of low- to high-frequencies

LF Δ _D – Low frequency differences between Task Difficulty conditions

LF Δ _F – Low frequency differences between Fatigue conditions

M – Mean

MATLAB – Matrix Laboratory, a proprietary multi-paradigm programming language and numeric computing environment developed by MathWorks

MBI-GS – Maslach Burnout Inventory-General Survey

mmHg – millimeters of mercury

ms - milliseconds

Multi-INT – Multi-source Intelligence

MVO – Michigan vs Ohio scenario

n – Sample size

nu – Normalized units

NASA TLX – National Aeronautics and Space Administration Task Load Index

ODNI – Office of the Director of National Intelligence

OSINT – Open-source Intelligence

p – Probability

PED – Process, Exploitation, and Dissemination

PERCLOS – Percentage of eyelid closure

PNS – Parasympathetic nervous system

PPG – Photoplethysmography

PVT – Psychomotor Vigilance Task

r – Pearson’s *r* correlation statistic

RFI – Request for Information

RR – Time between successive heart beats (based on R peaks)

RT – Reaction time

SAM – Surface-to-air-missile
SampEn – Sample Entropy
SAR – Same as Report
SD – Standard Deviation
SDK – Software Development Kit
SE – Standard Error
SIG – Signals Intelligence analyst
SIGINT – Signals Intelligence
SME – Subject matter expert
SNS – Sympathetic nervous systems
SrA – Senior Airman
SSgt – Staff Sergeant
STAI – State-Trait Anxiety Inventory
STE – Synthetic Task Environment
t – Student’s *t*-statistic
TACREP – Tactical Report
TSA – Target Systems Analysis
TSgt – Technical Sergeant
TTJO – Threat to Joint Operations
UI – User interface
USAF – United States Air Force
 Δ – Delta; symbol for ‘difference between’