USAARL-TECH-TR--2023-16

UNITED STATES ARMY AEROMEDICAL RESEARCH LABORATORY

Evaluation of Inter-subject Variability in Physiological Metrics and Workload Perception: Implications for Operator State Monitoring

Amanda Kelley, Aaron McAtee, Michelle Duffy, & Katie Feltman

Notice

Qualified Requesters

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

Change of Address

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

Disposition

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

Disclaimer

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

IRB Determination and Number

This study, USAARL log number 2022-006, was determined to be research not involving human subjects by the USAARL Exempt Determination Official on 8 February 2022.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
13-02-2023

2. REPORT TYPE
Technical Report

3. DATES COVERED *(From - To)*
2022-2023

4. TITLE AND SUBTITLE
Evaluation of inter-subject variability in physiological metrics and workload perception: Implications for operator state monitoring

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Kelley, A.[1], McAtee, A. M.[1,2], Duffy, M. J.[1,3], & Feltman, K. A.[1]

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Aeromedical Research Laboratory
P.O. Box 620577
Fort Rucker, AL 36362

8. PERFORMING ORGANIZATION REPORT NUMBER
USAARL-TECH-TR--2023-16

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

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

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

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

13. SUPPLEMENTARY NOTES
[1]U.S. Army Aeromedical Research Laboratory, [2]Goldbelt Frontier, LLC, [3]Oak Ridge Institute for Science and Education

14. ABSTRACT
Incorporation of operator state monitoring through non-invasive psychophysiological metrics would enable an objective assessment of the operator's cognitive state in real-time. Realization of such an endeavor would translate to the ability to develop adaptive automation that tailors the level of automation based on the operator's current cognitive state, as well as the ability to provide leadership with up-to-date information on the crew's cognitive state. However, while much of the work completed to-date has yielded promising results, a key obstacle remains: accounting for individual variability. This study analyzed archival data from four studies assessing physiological measures, individual differences, and cognitive workload. Minimal support for relationships between individual differences and workload levels was identified. Of those evaluated, abstract reasoning, state anxiety characteristics, and depression symptoms correlated with workload but not consistently across the four studies' datasets analyzed. With respect to physiological measures and workload, the findings show a number of physiological variables that consistently appeared as top predictors in identifying workload condition. Further research is needed to examine additional individual difference measures that may contribute to changes in physiological response from workload manipulations.

15. SUBJECT TERMS
Cognitive workload, operator state monitoring, machine learning

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Loraine St. Onge, PhD
UNCLAS	UNCLAS	UNCLAS	SAR	22	**19b. TELEPHONE NUMBER** *(Include area code)* 334-255-6906

This page is intentionally blank.

Acknowledgements

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

This page is intentionally blank.

Table of Contents

List of Tables

This page is intentionally blank.

Introduction

Ongoing work within the military (Durkee et al., 2015; Feltman et al., 2020), as well as in the civilian sector (Teo et al., 2019), continues to pursue the goal of operator state monitoring (OSM) through physiological metrics. Using physiological metrics for OSM would allow for a mostly non-invasive, objective assessment of the operator's cognitive state. While much of the work completed to-date has yielded promising results, a key obstacle remains, accounting for individual variability. Thus, this study evaluated inter-subject variability in physiological metrics and perception of workload across four independent studies of cognitive workload.

Most studies in the field have focused on identifying high versus low workload conditions by using data at the group level. This has led to the establishment of generally consistent trends in physiological responses (Borghini et al., 2013; Charles & Nixon, 2019). Table 1 summarizes some of these trends.

Table 1. Summary of Physiological Measures and Their Response to High Workload

Physiological Measure	Response During High Workload
Heart Rate	Increases (Unni et al., 2017; Hidalgo-Munoz et al., 2019)
Heart Rate Variability (HRV)	Decreases (Stuiver et al., 2014)
Pupil Diameter	Increases (Feng at al., 2018)
Eye Fixation Duration	Increases (Feng et al., 2018), Decreases (Schulz et al., 2011)
Number of Eye Fixations	Increases (Van Orden et al., 2001)
Theta Wave Activity – Electroencephalography (EEG)	Increases (Wu et al., 2017)
Alpha Wave Activity – EEG	Decreases (Van Orden et al., 2001; Wu et al., 2017)
Beta Wave Activity – EEG	Increases (Kurimori & Kakizaki, 1995), Decreases (Xiaoli et al., 2020; Hussain et al., 2021)
Oxygen Saturation (rSO_2)	Increases (Sassorroli et al., 2008)
Electrodermal Activity	Increases (Tarabay et al., 2018)
Cerebral Blood Flow Velocity (CBFV)	Increases (Warm & Parasuraman, 2007)

Although these general trends have been noted, there remains the issue of accounting for individual variability within the data. In general, individuals respond physiologically to increased workload in patterns similar to those in Table 1. However, there remains a large amount of individual variability that many models of workload have not accounted for. This limits the predictive validity of the models to identify subtle changes in workload across a range of individuals. In order to improve the predictive validity, as well as reliability, of models aimed at diagnosing operator state, it will be critical to account for individual differences.

A range of individual differences can impact physiological responses. Broadly, these can be categorized as state or trait characteristics and can apply to a range of psychological constructs. State characteristics include those that are time varying, often changing due to external factors. Trait characteristics, however, are those that tend to remain stable or time-invariant and are internally driven (Chaplin et al., 1988). Anxiety is a construct that is commonly measured at both the state and trait level and will serve as an example of the importance of identifying and accounting for individual differences. A commonly used tool for measuring anxiety is the State-Trait Anxiety Inventory (STAI) (Ercan et al., 2015), which measures a person's state anxiety (the emotional state an individual is in at a specific time point) and trait anxiety (a person's proneness to experiencing anxiety as a personality trait). Generally, an individual who is higher in trait anxiety will also exhibit greater state anxiety, but state anxiety will vary based on current circumstances, such as task demands. How individuals vary in trait and state anxiety can have implications for not only how they respond to a stressful situation, which in the context of aviation may be the presentation of an emergency, but also in how they respond physiologically. For example, Wheelock et al. (2016) found support for individual differences in self-reported stress and trait anxiety related to activity within the prefrontal cortex.

In addition to what can be considered personality-related individual difference variables, health and experience variables should also be explored. Pre-existing medical conditions (e.g., hormonal and metabolic conditions affecting brain tissue, sleep disorders, recent head injuries) can impact how physiological signals are produced and may impact how they are interpreted in a real-time OSM context. The effects of experience are well-established for how it impacts performance and can also be identified in physiological monitoring (Borghini et al., 2016). For example, Borghini et al. (2013) were able to show distinct changes in frontal theta power measured through EEG as novices began to learn a flight simulation task. In their study, changes in theta activity were noted as the participant learned the task and were also supported with ratings of lowered workload.

To-date, the literature on individualized approaches to workload monitoring is limited. For example, Teo et al. (2019) is the only peer-reviewed published study addressing this issue that we were able to find in a preliminary review of the literature. Results of their study are promising in terms of being able to accommodate individual variability in modeling physiological response to workload. However, the generalizability of their work is limited. Moreover, addressing individual variability within an Army aviation population will require the use of Army aviators as the sample to develop the model, given that aviators tend to fall within various spectrums of individual difference measures already (e.g., intelligence, personality) (Causse et al., 2011). Incorporation of state and trait characteristics that may significantly impact physiological signals during OSM could enable the development of robust algorithms that are able to predict and/or identify operator state changes in real-time. This study used archival data to evaluate the presence of individual differences and the potential causes of those differences.

Methods

De-identified datasets from four studies completed at the U.S. Army Aeromedical Research Laboratory (USAARL) were included within this study. Study one consisted of discrete computerized cognitive tasks (Feltman et al., 2020), whereas study two was completed within the UH-60 simulator (Feltman et al., 2021). The third study included was also conducted within the UH-60 simulator (Feltman et al., 2022), whereas the fourth study was completed within the unmanned aerial systems (UAS) simulator (Kelley et al., 2020). Participants from these studies included rated aviators; and cognitive-normal Soldiers under the age of 40.

All studies, except study one, included workload manipulations during simulated aviation tasks specific to the platform that was used (e.g., UH-60 or UAS). Study one used discrete cognitive tasks. The datasets included most of the same measures; however, some individual difference measures and physiological measures were not used in all studies. All studies used the NASA-Task Load Index (NASA-TLX) to assess subjective workload (Hart & Staveland, 1988). This instrument yields six subscale scores (mental demand, physical demand, temporal demand, performance, effort, frustration) and total weighted score. Table 2 summarizes the methodological details from all studies. Table 3 presents the individual differences measured across studies.

Table 2. Summary of Participants, Measures, and Methods from Included Studies

	Study 1	**Study 2**	**Study 3**	**Study 4**
Participants	16 (13 males); M_{age} = 31.33 (*SD* = 9.72)	23 males; M_{age} = 36.13 (*SD* = 4.99)	8 males; M_{age} = 37.25 (*SD* = 3.33)	22 (21 males); M_{age} = 32.14 (*SD* = 5.98)
Platform Used /Workload Manipulations	Computerized cognitive tasks across four domains (auditory, visual, physical, and cognitive)	Operationally-relevant flight scenarios in UH-60 simulator with workload manipulated across same domains as study one	Operationally-relevant flight scenarios in UH-60 simulator with workload manipulations tailored to the individual	Change/threat detection task in simulated UAS operations with stimuli presented at two rates (underload, high workload)
Number of Trials	8 total included in this study (4 high/ 4 low workload conditions)	16 total (8 high workload/8 low workload conditions)	2 total (1 high/ 1 low workload condition)	4 (2 underload and 2 high workload conditions)
Individual Difference Measures Collected	Demographics (age), Depression Scores, Sleepiness, Intelligence, Anxiety	Demographics (age), Depression Scores, Sleepiness, Intelligence, Anxiety	Demographics (age), Depression Scores, Sleepiness, Intelligence, Anxiety	Demographics (age), Depression Scores, Sleepiness, Anxiety,
Physiological Measures Collected	Electrocardiogram (ECG) (heart rate, heart rate variability); Respiration Rate; EEG (frontal alpha, theta, beta); Electrooculography (EOG) (blink rate, number of blinks, blink velocity)	ECG (heart rate, heart rate variability); Respiration Rate; EEG (frontal alpha, theta, beta); EOG (blink rate, number of blinks, blink velocity)	ECG (heart rate, heart rate variability); EEG (frontal alpha, theta, beta)	ECG (heart rate, heart rate variability); EEG (frontal alpha, theta, beta); EOG (blink rate, number of blinks, blink velocity)

Table 3. Individual Difference Measures Across Studies

Measure	Survey Used	Outcome	Studies Used In
Depression Scores	Beck's Depression Inventory-II (BDI) (Beck et al., 1996)	Total score	1, 2, 3, 4
Sleepiness Ratings	In-house developed Sleep-Wake Questionnaire;	Self-reported sleep quantity and quality	1, 2, 3
	Karolinska Sleepiness Scale (Kaida et al., 2006)	Level of sleepiness	4
Intelligence Scores	Shipley Institute of Living Scale (Shipley & Burlingame, 1941)	Verbal Score Abstraction Score	1, 2, 3
Anxiety Scores	State-Trait Anxiety Inventory (STAI) (Spielberger et al., 1983)	State score (event-dependent anxiety) Trait (persistent demonstrations of anxiety)	1, 2, 3

Statistical Approach

Data from the four studies were first examined with respect to comparability to determine whether the datasets could reasonably be combined for analyses. To do so, separate linear regression models were used to evaluate workload ratings in high and low workload conditions across the four studies. Additionally, effect sizes were compared. Correlational analyses (Pearson's *r*) were then conducted to examine the relationships between physiological measurements and individual difference variables as well as individual difference variables and perceived (reported) workload by workload condition. Finally, machine learning algorithms were computed using the classification and regression training (caret) package in R (Kuhn, 2022). These were completed individually for studies one, two, and four, given the differences identified in perceived workload across studies and unequal numbers of trials. Study three did not have a sufficient number of trials for these types of analyses and was thus excluded. The goal of these algorithms was to predict the workload level (high or low) of each participant in each trial. Models were run predicting workload condition. For each study, the datasets were split into training and test datasets, 80% and 20%, respectively.

Results

Using the archival datasets, the data from the four identified studies were compiled into a single Excel workbook. All analyses were conducted using R version 3.6.1.

Initial Steps - Evaluate Datasets to Determine whether Cases can be Grouped

Study data were first examined to determine whether the data were comparable enough to perform analyses on the aggregated data across studies. To do so, the NASA-TLX scores from the high workload conditions and low workload conditions were each compared across studies. Using separate linear regression models with repeated measures with one within-subjects factor (Study, four levels: study one, study two, study three, study four), mean weighted NASA-TLX scores were compared across the four studies for each the high workload conditions and the low workload conditions to determine the degree of comparability across studies with respect to workload conditions. The NASA-TLX scores within the high workload conditions were significantly different across the studies, χ^2 (3) = 12.3, p = 0.006. Post-hoc analysis showed that scores obtained in study three were significantly higher than those from the remaining three studies (see Table 4) suggesting that the workload manipulation in study three yielded a higher degree of workload experienced. Note that *N* refers to the total number of trials, not the total number of participants. This is due to studies one, two, and four having multiple iterations of high workload conditions.

Table 4. High Workload Summary Statistics of NASA-TLX Scores

Study Number	***N***	***Mean***	***SD***
1	58	50.2	22.8
2	183	53.9	20.4
3	8	73.7	17.7
4	39	46.5	16.8

Similar to the high workload conditions, the NASA-TLX scores in low workload conditions were significantly different across the studies, χ^2 (3) = 8.1, p = 0.04. Post-hoc analysis showed that scores obtained in study three were significantly lower than those from the remaining three studies (see Table 5).

Table 5. Low Workload Summary Statistics of NASA-TLX Scores

Study Number	***N***	***Mean***	***SD***
1	58	36.5	23.8
2	183	39.0	18.7
3	8	19.2	12.5
4	39	38.8	19.8

Next, effect sizes were examined across the studies to determine whether the workload manipulations used in each study resulted in similar effect sizes for the physiological measures of interest. To do so, the amount of change in physiological variables across the low and high workload conditions were compared, and Cohen's D was calculated for effect size. Effect sizes for each physiological variable for each study are reported in Table 6 below.

This space is intentionally blank.

Table 6. Effect Sizes for Physiological Variables Across Studies

	Study 1	**Study 2**	**Study 3**	**Study 4**
Frontal Theta	0.50*	0.07	0.77*	0.26
Frontal Alpha	0.48	0.10	1.24†	0.54*
Frontal Beta	0.16	0.09	0.94†	0.02
Heart Rate Variability (ratio of high to low frequency power (HF/LF))	0.13	0.23	0.15	0
Heart Rate	0.56*	0.27	0.55*	0.21
Respiration Rate	0.51*	0.36	N/A	0.09
Mean Blinks	0.07	0.04	N/A	0.25
Number of Blinks	0.04	0.15	N/A	0.11
Blink Rate (per minute [min.])	0.14	0.31	N/A	0.11
Blink Velocity	0.37	0.33	N/A	0.19

Note. *Interpreted as medium effect sizes (0.5 - 0.8); †Interpreted as large effect sizes (> 0.8).

Examination of the effect sizes suggests that the EEG, ECG, respiration, and blink variables are not consistent across studies. As such, each study should be examined separately, rather than aggregated. The subsequent analyses were performed on each individual study dataset.

Examine Individual Differences that Contribute to Physiological Responses to Workload Manipulations

To identify individual differences that may contribute to changes in physiological response to workload manipulations, each study was examined independently. The availability of physiological measures and individual difference measures differed across the studies. These differences are indicated below.

Study One.

Correlations were performed between the physiological variables and individual difference measures (indicated in Table 7 below) for each high and low workload manipulation. Weak to moderate correlations between the individual differences variables and physiological variables were identified. Correlations are only reported for those that reached at least 0.3 (a weak correlation).

This space is intentionally blank.

Table 7. Study One: Correlation Results Between Individual Difference Variables and Physiological Variables

	Age	**STAI State Score**	**STAI Trait Score**	**BDI Score**	**Shipley Verbal**	**Shipley Abstraction**	**Sleepiness Rating**
High Workload							
Frontal Theta	-	-0.322	-	-	-	-	-0.386
Frontal Alpha	0.399	0.322	-	-	-	-	0.379
Frontal Beta	-	0.344	-	-	-	-	0.423
Heart Rate Variability (HF/LF)	-0.394	-	0.477	-	-	-	-
Heart Rate	-	-	-0.309	-	-	-	-
Respiration Rate	-	-	-	-0.347	-	-	-
Mean Blinks	-	0.680*	0.717*	0.568*	0.303	0.359	0.436
Number of Blinks	-	-	-	-	-0.477	-0.599*	-0.350
Blink Rate (per min.)	-	-	-	-	-	0.567*	-
Blink Velocity	0.340	0.351		0.688*		0.502*	
Low Workload							
Frontal Theta	-	-0.336	-	-	-	-	-
Frontal Alpha	0.616*	0.382	-	-	-	-0.441	-
Frontal Beta	-	0.403	-	-	-	-	0.389
Heart Rate Variability (HF/LF)	-	0.375	0.566*	-	0.389	0.354	-
Heart Rate	-	-	-	-	-	-	-
Respiration Rate	-	-	-	-0.301	0.457	-	0.373
Mean Blinks	-	0.598*	0.571*	0.548*	0.394	-	-
Number of Blinks	-	-	-	-	-0.549*	-	-
Blink Rate (per min.)	-	-	-	-0.414	0.451	-	-0.365
Blink Velocity	0.451	0.364	-	0.750*	-	0.509*	-

Note. *Interpreted as moderate correlation ($0.5 < r < 0.8$).

This space is intentionally blank.

Study Two.

Correlations were performed between the physiological variables and individual difference measures (indicated in Table 8 below) for each high and low workload manipulation. Weak to moderate correlations between the individual differences variables and physiological variables were identified. Correlations are only reported for those that reached at least 0.3 (a weak correlation).

Table 8. Study Two: Correlation Results Between Individual Difference Variables and Physiological Variables

	Age	**STAI State Score**	**STAI Trait Score**	**BDI Score**	**Shipley Verbal**	**Shipley Abstraction**	**Sleepiness Rating**
High Workload							
Frontal Theta	-	-	0.325	-	-	-	-
Frontal Alpha	-	-	-	-	-	-	-
Frontal Beta	-	-	-	-	-	-	-
Heart Rate Variability (HF/LF)	-	-	-	-	-.380	-0.663*	-
Heart Rate	-	-0.385	-0.455	-	-	-0.439	-
Respiration Rate	-0.491	-0.303	-	-	0.487	-	-
Mean Blinks	-	-	-	-0.311	-0.340	-0.454	-0.523*
Number of Blinks	-	-0.381	-0.301	-	-	-	-0.474
Blink Rate (per min.)	0.387	-	-0.437	-	-	-	0.388
Blink Velocity	-	-0.401	-	-	-0.308	-0.304	-0.634*
Low Workload							
Frontal Theta	-	-	-	-	-	-	-
Frontal Alpha	-	-	-	-	-	-	-
Frontal Beta	-	-	-	-	-	-	-
Heart Rate Variability (HF/LF)	-	-	-	-	-0.454	-0.653*	-
Heart Rate	-	-	-0.437	-	-0.413	-0.559*	-
Respiration Rate	-0.426	-0.318	-	-	0.416	-	-
Mean Blinks	-	-	-	-	-	-0.423	-0.568*
Number of Blinks	-	-0.497	-0.305	-	-	-0.454	-0.463
Blink Rate (per min.)	-	0.581*	0.327	0.486	-	-	0.467
Blink Velocity	-	-0.431	-	-0.302	-0.484	-	-0.614*

Note. *Interpreted as moderate correlation ($0.5 < r < 0.8$).

Study Three.

Correlations were performed between the physiological variables and individual difference measures (indicated in Table 9 below) for each high and low workload manipulation. Note that this study did not include eye or respiration data, nor did it include BDI scores. The majority of correlations were weak to moderate between the individual differences variables and physiological variables. However, a few strong correlations were identified. Correlations are only reported for those that reached at least 0.3 (a weak correlation).

Table 9. Study Three: Correlation Results Between Individual Difference Variables and Physiological Variables

	Age	**STAI State Score**	**STAI Trait Score**	**Shipley Verbal**	**Shipley Abstraction**	**Sleepiness Rating**
High Workload						
Frontal Theta	-	-	-0.387	-0.784*	-0.375	-0.631*
Frontal Alpha	-0.467	-	-	0.562*	0.412	-
Frontal Beta	-	-	-	0.813†	0.422	0.590
Heart Rate Variability (HF/LF)	0.586*	-0.386	0.539*	0.698*	-	0.867†
Heart Rate	0.607*	-	-	-	-0.354	0.538*
Low workload						
Frontal Theta	-0.627*	-	-	-0.673*	-	-0.881†
Frontal Alpha	-	-	-	0.778*	0.358	-
Frontal Beta	0.441	-	-	0.758*	-	0.773*
Heart Rate Variability (HF/LF)	0.848†	-	0.517	0.425	-	0.896†
Heart Rate	0.684*	-	-	0.319	-	0.749*

Note. *Interpreted as moderate correlation ($0.5 < r < 0.8$); †Interpreted as strong correlation ($r > 0.8$).

This space is intentionally blank.

Study Four.

Correlations were performed between the physiological variables and individual difference measures (indicated in Table 10 below) for each high and low workload manipulations. Note that this study did not include STAI or Shipley data. Weak to moderate correlations between the individual differences variables and physiological variables were identified. Correlations are only reported for those that reached at least 0.3 (a weak correlation).

Table 10. Study Four: Correlation Results Between Individual Difference Variables and Physiological Variables

	Age	BDI Score	Sleepiness Rating
High Workload			
Frontal Theta	-0.366	-	-
Frontal Alpha	-	-	-
Frontal Beta	-	-	-
Heart Rate Variability (HF/LF)	0.497	-	-
Heart Rate	-	-	0.349
Respiration Rate	-	-	0.397
Mean Blinks	-	0.367	-
Number of Blinks	-	0.745*	-0.478
Blink Rate (per min.)	-	0.745*	-0.478
Blink Velocity	-	-	-
Low Workload			
Frontal Theta	-0.441	-	-
Frontal Alpha	0.428	-	-
Frontal Beta	-	-	-
Heart Rate Variability (HF/LF)	-	-	-
Heart Rate	-	-	-
Respiration Rate	-	-	-
Mean Blinks	0.336	-	-
Number of Blinks	-	-	-
Blink Rate (per min.)	-	-	-
Blink Velocity	-	-	-

Note. *Interpreted as moderate correlation ($0.5 < r < 0.8$).

Overall, individual differences assessed within these studies had differential impact on the physiological variables. EOG variables correlated state anxiety, trait anxiety, depression scores, abstract reasoning (Shipley A), and reported sleepiness in both high and low workload conditions. ECG variables only correlated with age, abstract reasoning (Shipley A), and reported sleepiness. EEG variables were the least impacted by individual differences and only correlated with verbal intelligence (Shipley V) and reported sleepiness.

Examine Relationships between Individual Differences and Perceived Workload

Next, correlations between the individual difference measures and perceived workload (total scores from the NASA-TLX) were performed within each study as well as on the combined data from studies one, two, and four. These were performed only for the high workload conditions given differences in the "low" workload manipulation definitions across studies (i.e., "low" was either structured as a comfortable level of workload or as an uncomfortable degree of underload). Table 11 summarizes the findings.

Table 11. Correlations with Perceived Workload

	Age	**STAI State**	**STAI Trait**	**BDI Score**	**Shipley Verbal**	**Shipley Abstraction**	**Sleepiness**
Study 1	-0.074	0.106	0.097	0.095	-0.098	-0.196	-0.202
Study 2	-0.042	-0.017	-0.008	-0.002	-0.056	-0.004	-0.062
Study 3	0.235	-0.581*	-0.256	N/A	0.160	0.230	0.297
Study 4	0.072	Not Applicable (N/A)	N/A	-0.183	N/A	N/A	0.446
Studies 1, 2, & 4	0	0.018	-0.001	0.027	0.016	-0.031	-0.064

Note. *Interpreted as moderate correlation ($0.5 < r < 0.8$).

Overall, there were minimal relationships between individual difference measures and perceived workload. Study three found a moderate correlation with STAI State scores, however, this study had the least number of participants (8 total) which is important to consider when interpreting these findings.

This space is intentionally blank.

Machine Learning Algorithms

Machine learning algorithms were computed using the classification and regression training (caret) package in R (Kuhn, 2022). These were completed individually for Studies one, two, and four, given the differences identified in perceived workload across studies and unequal numbers of trials. Study three did not have a sufficient number of trials for these types of analyses and was thus excluded. The goal of these algorithms was to predict the workload level (high or low) of each participant in each trial. Models were run predicting workload condition. For each study, the datasets were split into training and test datasets, 80% and 20%, respectively. The equal proportion of high and low workload trials was maintained in both datasets (unless otherwise noted). Missing data were imputed using the *k*-nearest-neighbor (kNN) algorithm within the caret package. All predictors were transformed to have a range from 0 to 1. Then, a recursive feature selection algorithm was used to determine the top predictor variables from the total pool of predictors (Table 12).

Table 12. Predictor Variables used in Machine Learning Algorithms

Individual Difference Measures	Physiological Metrics	NASA-TLX Scores
Age	Frontal Theta Total	Mental Demand Subscale
STAI State	Frontal Alpha Total	Frustration Subscale
STAI Trait	Frontal Beta Total	Effort Subscale
BDI Score	HF/LF Ratio (HRV)	Temporal Demand Subscale
Shipley Verbal	Heart Rate	Physical Demand Subscale
Shipley Abstraction	Respiration	Performance Subscale
Sleepiness	Blink Average	Total Weighted Score
	Num. Blinks	
	Blink Rate	
	Blink Velocity	

Study One.

For study one, there were an equal number (60 total) of high and low workload trials. These analyses assumed the Experimental Workload Level is “truth” data. EOG data (variables: blink average, number of blinks, blink rate, blink velocity) were excluded given that 53% of the data were missing. The resultant top predictor variables were: NASA-TLX total weighted workload ratings; NASA-TLX frustration subscale ratings; NASA-TLX effort subscale ratings, NASA-TLX temporal subscale ratings, and EEG frontal theta. The results from the various machine learning algorithms are summarized in Table 13 below.

This space is intentionally blank.

Table 13. Machine Learning Algorithm Results from Study One

Algorithm	**Model Accuracy**	**Top Predictors**	**Model Name in R package**
Cluster	0.45	NASA-TLX Total NASA-TLX Temporal Demand NASA-TLX Effort NASA-TLX Frustration NASA-TLX Mental Demand Respiration Rate Heart Rate NASA-TLX Performance Frontal Alpha Total	knn
Multivariate Adaptive Regression Spline	0.54	NASA-TLX Total NASA-TLX Effort Frontal Theta Total NASA-TLX Temporal Demand HF/LF Ratio	earth
Classification Tree	0.58	NASA-TLX Total NASA-TLX Frustration Respiration Rate Heart Rate Frontal Alpha Total HF/LF Ratio	adaboost
Support Vector Machine	0.58	NASA-TLX Total NASA-TLX Temporal Demand NASA-TLX Effort NASA-TLX Frustration NASA-TLX Mental Demand Respiration Rate Heart Rate NASA-TLX Performance Frontal Alpha Total	svmRadial
Extreme Gradient Boosting	0.67	NASA-TLX Total NASA-TLX Temporal Demand NASA-TLX Effort NASA-TLX Frustration NASA-TLX Mental Demand Respiration Rate Heart Rate NASA-TLX Performance Frontal Alpha Total	xgbLinear
Random Forest	0.63	NASA-TLX Total Frontal Theta Total NASA-TLX Effort NASA-TLX Temporal Demand HF/LF Ratio	rf

Study Two.

For study two, there were an equal number of high and low workload trials (184 total). These analyses assumed the experimental workload level is “truth” data. Three iterations of each machine learning algorithm was performed: The first include all predictor variables, the second only included one NASA-TLX measure and the physiological variables, and the third included only the physiological variable. The purpose of this was to evaluate whether consistent “top predictors” emerged. The resultant top predictor variables were: NASA-TLX total workload ratings; NASA-TLX frustration subscale ratings; EEG frontal beta; EEG frontal theta; and EEG frontal alpha. The results from the various machine learning algorithms are summarized in Table 14 below.

This space is intentionally blank.

Table 14. Machine Learning Algorithm Results from Study Two

Algorithm	Model Accuracy	Top Predictors	Model Name in R package
Cluster	0.69	NASA-TLX Total Score NASA-TLX Frustration Subscale Score Respiration Rate Frontal Theta Total Frontal Beta Total HF/LF Ratio (HRV)	knn
Multivariate Adaptive Regression Spline	0.63	NASA-TLX Total Score NASA-TLX Frustration Subscale Score Frontal Theta Total Frontal Alpha Total Frontal Beta Total Respiration Rate	earth
Classification Tree	0.64	NASA-TLX Total Score NASA-TLX Frustration Subscale Score Frontal Theta Total Respiration Rate Frontal Beta Total HF/LF Ratio (HRV)	adaboost
Support Vector Machine	0.69	NASA-TLX Total Score NASA-TLX Frustration Subscale Score Frontal Theta Total Respiration Rate Frontal Beta Total HF/LF Ratio (HRV)	svmRadial
Extreme Gradient Boosting	0.61	NASA-TLX Total Score NASA-TLX Frustration Subscale Score Frontal Alpha Total Frontal Theta Total Frontal Beta Total Respiration	xgbLinear
Random Forest	0.67	NASA-TLX Total Score NASA-TLX Frustration Subscale Score Frontal Beta Total Frontal Alpha Total Frontal Theta Total HF/LF Ratio (HRV)	rf

Study Four.

For study four, there were an equal number of high and low workload trials (40 total). All trials were missing STAI and Shipley data and thus excluded from the analysis. EOG data were also excluded given that 70% of the data were missing. The results from the various machine learning algorithms are summarized in Table 15 below.

Table 15. Machine Learning Algorithm Results from Study Four

Algorithm	**Model Accuracy**	**Top Predictors**	**Model Name in R package**
Cluster	0.44	NASA-TLX Effort NASA-TLX Temporal Demand NASA-TLX Total Score NASA-TLX Mental Demand Frontal Theta Total Frontal Beta Total	knn
Multivariate Adaptive Regression Spline	0.50	NASA-TLX Effort NASA-TLX Temporal Demand NASA-TLX Total Score NASA-TLX Mental Demand Frontal Theta Total Frontal Beta Total	earth
Classification Tree	0.44	NASA-TLX Effort NASA-TLX Temporal Demand NASA-TLX Total Score NASA-TLX Mental Demand Frontal Theta Total Frontal Beta Total	adaboost
Support Vector Machine	0.44	NASA-TLX Effort NASA-TLX Temporal Demand NASA-TLX Total Score NASA-TLX Mental Demand Frontal Theta Total Frontal Beta Total	svmRadial
Extreme Gradient Boosting	0.31	Frontal Beta Total NASA-TLX Total Score NASA-TLX Mental Demand Frontal Alpha Total NASA-TLX Physical Demand Frontal Theta Total	xgbLinear
Random Forest	0.31	Frontal Beta Total NASA-TLX Effort Frontal Alpha Total Frontal Theta Total NASA-TLX Mental Demand	rf

Discussion

OSM is one of the first steps toward adaptive automation. OSM will incorporate physiological data to predict performance and, subsequently, risks to mission success. By examining individual differences that contribute to physiological and behavioral responses to operational stressors during flight, this project provides input to which individual difference variables are most critical in predicting state changes. In the short-term, this will translate into providing refinements to model development that will be able to better predict how individuals will respond to stressors imposed. In the long-term, this study will contribute to the overall development of OSM algorithms in order to improve predictive validity and reliability across aviators.

Initial examination of the relationships between the individual difference measures and the physiological measures collected across these four studies found some consistent patterns. However, these consistent patterns were mainly noted within studies. That is, the individual difference measures that correlated with physiological measures demonstrated consistent patterns within studies between workload conditions. For example, in study two, HRV recorded during high workload and low workload conditions correlated with intelligence (e.g., Shipley abstract reasoning scores). Across studies, there was only one consistent pattern noted. Blink measures were found to correlate with depression symptoms scores (higher scores largely driven by fatigue/sleep related questionnaire items) in both study one and study four. These differences in relationships across studies may reflect differences in the sample populations used (e.g., studies two and three were limited to rated aviators whereas study one included military and civilian personnel and study four included military personnel).

More consistent patterns were identified within the machine learning algorithms using individual difference measures, physiological measures, and subjective workload ratings to predict workload condition. Study three was not included in this analysis given an insufficient number of trials. The initial top predictors that were identified across the analyze studies included: NASA-TLX total score; NASA-TLX frustration, temporal demand, and effort subscale scores; and EEG variables (total frontal beta, theta, and alpha values). Six machine learning algorithms were run for each study. Findings showed that heart rate variability and respiration rate as well as the NASA-TLX mental demand subscale score were significant predictors. These findings suggest that in terms of identifying the level of workload presented to an individual, subjective ratings are a key indicator, as well as a number of physiological measures. The EEG variables consistently appeared as top predictors, as did HRV and respiration. From this finding, a key takeaway is that these variables should continue to be included as candidate variables for inclusion in operator state monitoring (e.g., Li et al., 2023).

The results of these analyses provide mixed support for individual difference measures included (age, anxiety, intelligence, and depressions symptoms). While abstract reasoning, state anxiety characteristics, and depression symptoms correlated with workload, these variables did not emerge as top predictors in any of the machine learning algorithms. Additional analyses at the individual level predicting subjective workload ratings may yield different results, however, at present it does not appear that these variables' relationship to workload outweigh or strongly influence the utility of physiological metrics in detecting mental workload states. This study is not sufficient to rule out these variables moving forward, and they should still be accounted for in future studies. In fact, the mixed findings may highlight a key limitation of this study, which is the lack of objective sleep estimates in the datasets. Measures of sleep and fatigue levels must be included in future workload studies given known relationships with the physiological variables studied as well as impacts on performance levels.

There are a number of limitations in this study that pose a challenge to interpretability and highlight some of the methodological issues in this line of research. First, the manipulation of workload across studies was not consistent in terms of subjective perception thus prohibiting combination of the datasets. Using a tailored approach to manipulating workload, as was done in study three, is likely a superior approach to understanding if and how physiological measures can be used to detect changes in workload at the individual level as it yields a more effective workload manipulation at the individual level. This also controls for influence from experience level. Another limitation is the poor performance of the models generated in this study. Model accuracies from study four are very poor and below the level expected by chance which also limits the interpretability of the findings. Study one model accuracies are only slightly better than those from study four. Study two yielded models performing better than chance but still substantially lower than needed for application. Also, the top predictors across all studies were the subjective workload ratings (NASA-TLX total and subscale scores) rather than physiological variables. Given the overall poor predictive accuracies and considering that the purpose of physiological monitoring to detect operator state is to eliminate the need for subjective feedback, the reality of building such a system is still unknown.

Future studies should consider examining the data from each participant separately. In doing so, one could quantify the variability of each physiological response under high and low workload manipulations. From there, individual participant regression or classification algorithms could be developed. This would account for individual differences in physiology, another important aspect of understanding if and how physiological variables can reliably be used to detect changes in operator state and workload level.

Conclusion

Minimal support for relationships between individual characteristics (e.g., demographics, intelligence) and workload levels was identified. Of those evaluated, abstract reasoning, state anxiety characteristics, and depression symptoms correlated with workload but not consistently across the four studies' datasets analyzed. With respect to physiological measures and workload, the findings show a number of physiological variables that consistently appeared as top predictors in identifying workload condition. These were: frontal beta and theta values, respiration rate, and heart rate variability. We recommend these variables continue to be included in ongoing research aimed at operator state monitoring. Further research is needed to examine additional individual difference measures that may contribute to changes in physiological response from workload manipulations.

References

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck depression inventory-II.* Psychological Corporation.

Borghini, G., Aricó, P., Astolfi, L., Toppi, J., Cincotti, F., Mattia, D., Cherubino, P. Vecchiato, G., Maglione, A. G., Graziani, I., & Babiloni, F. (2013). Frontal EEG theta changes assess the training improvements of novices in flight simulation task. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society's Annual Conference,* 6619–6622. https://doi.org/10.1109/EMBC.2013.6611073

Borghini, G., Aricó, P., Graziani, I., Salinari, S., Sun, Y., Taya, F., Bezerianos, A., Thakor, N. V., & Babiloni, F. (2016). Quantitative assessment of the training improvement in a motor-cognitive task by using EEG, ECG, and EOG signals. *Brain Topography, 29*, 149–161.

Causse, M., Dehais, F., & Pastor, J. (2011). Executive functions and pilot characteristics predict flight simulator performance in general aviation pilots. *The International Journal of Aviation Psychology, 21*(3), 217–234.

Chaplin, W. F., John, O. P., & Goldberg, L. R. (1988). Conceptions of states and traits: Dimensional attributes with ideals as prototypes. *Journal of Personality and Social Psychology, 54*(4), 541–557. https://doi.org/10.1037/0022-3514.54.4.541

Charles, R. L., & Nixon, J. (2019). Measuring mental workload using physiological measures: A systematic review. *Applied Ergonomics, 74*, 221–232.

Durkee, K., Pappada, S., Ortiz, A., Feeney, J., & Galster, S. (2015). Using context to optimize a functional state estimation engine in unmanned aircraft system operations. https://doi.org/10.1007/978-3-319-20816-9_3

Ercan, I., Hafizoglu, S., Ozkaya, G., Kirli, S., Yalcintas, E., & Akaya, C. (2015). Examining cut-off values for the state-trait anxiety inventory. *Revista Argentina de Clinica Psicologica, 24*, 143–148.

Feltman, K. A., Bernhardt, K., & Kelley, A. M. (2020). Measuring the domain specificity of workload using EEG: Auditory and visual domains in rotary-wing simulated flight. *Human Factors*. https://doi.org/10.1177/0018720820928626

Feltman, K. A., Kelley, A. M., Bernhardt, K., Basso, J., & Morabito, C. (2021). *Psychophysiological indicators of aviator flight performance for operator state monitoring* (USAARL-TECH-FR--2021-17). U.S. Army Aeromedical Research Laboratory.

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

Feng, C., Wanya, X., Yang, K., Zhuang, D., & Wu, X. (2018). A comprehensive prediction and evaluation method of pilot workload. *Technology and Health Care, 26,* S65–S78. dOI 10.3233/THC-174201

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183.

Hidalgo-Muñoz, A. R., Béquet, A. J., Astier-Juvenon, M., Pépin, G., Fort, A., Jallais, C., Tattegrain, H., & Gabaude C. (2019). Respiration and heart rate modulation due to competing cognitive tasks while driving. *Frontiers in Human Neuroscience, 12*(525). https://doi.org/10.3389/fnhum.2018.00525

Hussain, I., Young, S., & Park, S. (2021). Driving-induced neurological biomarkers in an advanced driver-assistance system. *Sensors, 21*(21), 6985. https://doi.org/10.3390/s21216985

Kaida M., Takahashi T., Åkerstedt A., Nakata Y., Otsuka T., Haratani K., & Fukasawa. (2006). Validation of the Karolinska sleepiness scale against performance and EEG variables. *Clinical Neurophysiology, 117*, 1574–81.

Kelley, A. M., Hayes, A. M., Bernhardt, K., & Mathews, C. (2020). *Sustaining performance and vigilance in extended UAS operations* (2020-043). U.S. Army Aeromedical Research Laboratory.

Kuhn, M. (2022). Classification and Regression Training. R Package Version 6.0-92, <https://CRAN.R-project.org/package=caret>

Kurimori, S., & Kakizaki, T. (1995). Evaluation of work stress using psychological and physiological measures of mental activity in a paced calculating task. *Industrial Health, 33*, 7–22.

Li, Q., Ng, K. K., Simon, C. M., Yiu, C. Y., & Lyu, M. (2023). Recognising situation awareness associated with different workloads using EEG and eye-tracking features in air traffic control tasks. *Knowledge-Based Systems*, *260*, 110179.

Sassaroli, A., Zheng, F., Hirshfield, L. M., Girouard, A., Solovey, E. T., Jacob, R. J. K., & Fantini, S. (2008). Discrimination of mental workload levels in human subjects with functional near-infrared spectroscopy. *Journal of Innovative Optical Health Sciences, 01*(02). https://doi.org/10.1142/S1793545808000224

Schulz, C. M., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Wasmaier, M., Kochs, E. F., & Schneider, G. (2011). Eye tracking for assessment of workload: A pilot study in an anaesthesia simulator environment. *British Journal of Anaesthesia, 106*(1), 44–50.

Shipley, W. C., & Burlingame, C. C. (1941). A convenient self-administering scale for measuring intellectual impairment in psychotics. *American Journal of Psychiatry*, *97*(6), 1313-1325.

Spielberger, C. D., Gorsuch, R. L., Lushene, R.E., Vagg, P. R., & Jacobs, G. A. (1983). Manual for the State-Trait Anxiety Inventory. Madrid: TEA Ediciones.

Stuiver, A., Brookhuis, K. A., de Waard, D., & Mulder, B. (2014). Short-term cardiovascular measures for driver support: Increasing sensitivity for detecting changes in mental workload. *International Journal of Psychophysiology, 92,* 35–41. ttp://dx.doi.org/10.1016/j.ijpsycho.2014.01.010

Tarabay, R., & Abou-Zeid, M. (2018). Assessing the effects of auditory-vocal distraction on driving performance and physiological measures using a driving simulator. *Transportation Research Part F, 58,* 351–364. https://doi.org/10.1016/j.trf.2018.06.026

Teo, G., Matthews, G., Reinerman-Jones, L., & Barber, D. (2019). Adaptive aiding with an individualized workload model based on psychophysiological measures. *Human-Intelligent Systems Integration.* https://doi.org/10.1007/s42454-019-00005-8

Unni, A., Ihme, K., Jipp, M., & Rieger, J. W. (2017). Assessing the driver's current level of working memory load with high density functional near-infrared spectroscopy: A realistic driving simulator study. *Frontiers in Human Neuroscience, 11(*167*).* https://doi.org/10.3389/fnhum.2017.00167

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

Warm, J. S., & Parasuraman, R. (2007). Cerebral hemodynamics and vigilance. In R. Parasuraman & M. Rizzo (Eds.), *Neuroergonomics: The brain at work* (pp. 146-158). Oxford University Press.

Wheelock, M. D., Harnett, N. G., Wood, K. H., Orem, T. R., Granger, D. A., Mrug, S., & Knight, D. C. (2016). Prefrontal cortex activity is associated with biobehavioral components of the stress response. *Frontiers in Human Neuroscience, 10*(583). https://doi.org/10.3389/fnhum.2016.00583

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

Xiaoli, F., Chaoyi, Z., Xin, Z., Hong, L., & Wei, Z. (2020). Assessment of mental workload based on multi-physiological signals. *Technology and Health Care, 28*(S1), 67–80.

U.S. Army Aeromedical Research Laboratory
Fort Rucker, Alabama

All of USAARL's science and technical information documents are available for download from the Defense Technical Information Center.

https://discover.dtic.mil/results/?q=USAARL

Army Futures Command
U.S. Army Medical Research and Development Command