ABSTRACT

ZHANG, WENJUAN. Effect of Interface Design on Cognitive Workload in Unmanned Aerial Vehicle Control. (Under the direction of Dr. David B. Kaber).

The application of Unmanned Aerial Vehicles (UAV) is growing rapidly in research, commercial and military use. UAV control interfaces represent a key means by which to transfer information from a vehicle to an operator. Supervisory control interfaces have utility for moderating operator cognitive workload and promoting efficient, accurate and safe UAV performance. However, research on UAV interfaces has heavily focused on increasing the level of vehicle automation in order to reduce complexity of control operations and interface designs for system users. Usable design of interfaces is often overlooked as a human factors methodology for system enhancement. There is a need for unmanned systems interface design methodologies to better support delivery of vehicle and automation information to users for improved performance.

The objectives of this study included: (1) investigating how UAV control interface design features may impact operator cognitive workload in fundamental control operations based on quantitative assessments; (2) demonstrating use of physiological measures of cognitive workload in simulated UAV operations; and (3) assessing the validity of an interface evaluation tool, the Modified GEDIS-UAV, as a means of operator cognitive workload prediction.

Three UAV control interface variations were generated as part of this research. The interfaces were classified into three categories - “baseline” interface, “enhanced” usability interface and “degraded” usability interface. These conditions were developed based the degree of user accessibility and comprehension of interface features for vehicle control. The relative usability of the prototypes was verified using heuristic analysis and the Modified

GEDIS-UAV evaluation tool. A few generic UAV control tasks were formulated for interface testing and vehicle speed was used to manipulate task pacing. A computational cognitive task performance modeling approach was used to verify the task workload manipulations.

An experiment was conducted to compare the various interface alternatives under the different task workload conditions. Forty-eight (48) participants were recruited. Each participant was randomly assigned to one of the three interface prototypes. User physiological responses and subjective ratings of cognitive workload were collected. User dynamic knowledge of the system was also tested.

Results on multiple response measures indicated that the "enhanced" interface led to the lowest operator workload and was robust to increased control task demands, as compared to the "baseline" and "degraded" interface. One of the physiological workload measures, normalized blink duration, also showed sensitivity to the interface manipulation and was correlated with user dynamic knowledge on the system. In addition, both the cognitive task performance model results and the Modified GEDIS-UAV scores revealed utility for predicting workload responses.

This study identified the influence of interface design features on operator cognitive workload responses. Findings also demonstrated the application of physiological responses for predicting cognitive workload during UAV operations. In addition, results of the study demonstrated the utility of an interface evaluation tool for predicting UAV operator cognitive workload. The approaches used in this study can be applied to interface evaluation for real UAVs, as well as other unmanned systems, as a basis for improving human-automation interaction. Finally, the new interface evaluation tool, the Modified GEDIS-UAV, can be

used to support UAV interface design, guide redesign, and facilitate selection of interface alternatives that reduce cognitive workload for operators.

Effect of Interface Design on Cognitive Workload in Unmanned Aerial Vehicle Control

by
Wenjuan Zhang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Industrial Engineering

Raleigh, North Carolina

2017

APPROVED BY:

_____	_____
David B. Kaber Committee Chair	Manida Swangetr
_____	_____
Nancy Currie	Peter Bloomfield

BIOGRAPHY

Wenjuan Zhang enrolled at Beijing University of Aeronautics and Astronautics (BUAA) in 2009 and studied engineering science in the program of École Centrale de Pékin. She graduated in 2013 with a Bachelor of Science in Mathematics and Applied Mathematics. In the same year, Wenjuan was admitted to the Master of Industrial Engineering program at North Carolina (NC) State University. During her study at NC State University, she developed interest in human factors and ergonomics in systems design. She completed the degree of Master of Science in Industrial Engineering with a thesis focused on biometric analysis of virtual reality-based simulation of motor therapy tasks. She then continued in the Ph.D. program in the Industrial & Systems Engineering Department. She was employed as a teaching assistant for related courses at the undergraduate and graduate levels. She also received graduate research assistant appointments through multiple external grants to the Department. Upon completion of her degree, she plans to continue her research in human factors and ergonomics in academia or industrial fields.

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. David Kaber for his guidance and support for my research. I appreciate Drs. Swangnetr, Currie and Bloomfield serving on my committee and for sharing their knowledge. I believe the expertise of the committee substantially contributed to the rigor and meaning of my work. Moreover, I would like to thank my fellow researcher for their words of encouragement, and, especially David Feltner and James Shirley, for their contribution to all phases of my study. I'm also grateful for my family for always being supportive.

Finally, this research would not have been possible without the support of these people as well as the National Aeronautics and Space Administration (NASA) through a grant to the NC State Ergonomics Lab (No. NNX16AB23A). I thank Dr. Terry Fong, the NASA project monitor, for his interest and commitment to our Lab's research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

A	Alarm
ANOVA	Analysis of Variance
AOI	Area of Interest
ATC	Air Traffic Control
BD	Blink Duration
CBRNE	Chemical, Biological, Radiological, Nuclear, And Explosive
CTA	Cognitive Task Analysis
C	Color
Cog	Cognitive
DEC	Data Entry Command
DV	Dependent Variable
DKQ	Dynamic Knowledge Queries
DL	Display Layout
ECG	Electrocardiogram
EEG	Electro-Encephalography
ERP	Event Related Potential
FAA	Federal Aviation Administration
fNIRS	Functional Near-Infrared Spectroscopy
GEDIS	Guia Ergonómica Para El Diseño De Interfaz De Supervision
GEI	Global Evaluation Index
GOMS	Goals, Operators, Methods, Selection Rules
GOMSL	Goms Language
GUI	Graphical User Interfaces
GSC	Ground Control Station
H	Hypothesis
HF	High Frequency
HRV	Heart Rate
HRV	Heart Rate Variability
HSD	Honestly Significant Difference
HFDS	Human Factors Design Standard
IP	Information Presentation
IV	Independent Variable
JAUS	Joint Architecture for Unmanned Systems
LOA	Level of Automation
LF	Low Frequency
LTM	Long Term Memory

MCH	Modified Cooper-Harper
MCH-UVD	Modified Cooper-Harper For Unmanned Vehicle Display
M	Motor
MN	Map And Navigation
MP	Mission Planner
P	Perceptual
NASA-TLX	NASA Task Load Index
NGOMSL	Natural GOMS Language
PC	Physical Control
PFD	Primary Flight Display
RTL	Return to Launch
SD	Status and Devices
SWAT	Subjective Workload Assessment Technique
T	Text
UAV	Unmanned Aerial Vehicle
UGV	Unmanned Ground Vehicle
ULF	Ultra-Low Frequency
VLF	Very-Low Frequency
WM	Working Memory

1. Introduction

1.1 State-of-the art of UAV interfaces

In recent years, Unmanned Aerial Vehicles (UAVs) have begun to play a key role in academic, industrial, and military applications. They have been applied to various domains, including search and rescue, reconnaissance, intelligent traffic monitoring, border patrol, disaster monitoring, safety inspection, etc. (Gupta, Ghonge, & Jawandhiya, 2013; Irizarry, Gheisari, & Walker, 2012). Although some of these applications have previously been fielded by manned aircraft, the task of piloting a UAV is very different from that of a pilot-in-command of a traditional aircraft. As Olson and Wuennenberg (2001) pointed out, a traditional pilot operates in a three-dimensional (3D) world while a UAV presents the same world to a pilot in two-dimensions (2D) via a camera image on flat display screen. A primary consequence of removing the pilot from the aircraft is that the UAV operator is deprived of a range of sensory cues that are otherwise available during manned flight. Rather than receiving sensory input directly from an environment, the UAV operator receives only the information presented through a ground control station (GCS) interface. Consequently, human interaction with GCSs has emerged as an important research direction in the recent past (Demir, Cicibas, & Arica, 2015).

The objective of a GCS interface is to transfer information between a UAV and a human operator. GCSs provide tools to perceive a remote environment, to make decisions on UAV operations, and to generate control commands (Fong & Thorpe, 2001). Williams (2007) identified four levels of control in UAV pilot interfaces. The lowest level allows for direct control of aircraft flight surfaces through the use of a joystick, whereas the highest

level relies on the use of programmed waypoints in a flight path to identify the location and altitude of the aircraft at a given point in time. Two intermediate levels of control allow for command of either bank angle/turn rate or heading (for horizontal control) and vertical speed or altitude (for vertical control). Intermediate levels of control involve human operator use of physical interfaces as part of some systems, like a joystick or knob, or a computer interface that either requires a "pull-down" menu selection or interaction with a virtual control using a computer pointing device, such as a mouse or trackball. Lower levels of control provide more immediate access to aircraft attitude but require greater integration of information by a pilot to achieve a particular flight objective, while higher levels of control make flying the aircraft easier but restrict available flight options (Williams, 2007). Williams (2007) also pointed out that there is a trend in the UAV industry to replace manual control of aircraft during takeoff and landing with automation and a perception of no need for control inputs below the automation level of joystick commanding of bank angle and vertical speed.

In general, as robotic systems become more autonomous, control interfaces are used less for direct manipulation and more for monitoring and diagnosis of system faults (Fong & Thorpe, 2001). For example, supervisory control interfaces provide tools to make robot navigation and motion command generation easier for operators (Fong & Thorpe, 2001). However, automation of various functions should not eliminate the need for human intervention (Gabriel, Ramallo, & Cervantes, 2016). Fong and Thorpe (2001) identified several design challenges for supervisory control interfaces, including display layout, tools for managing human-robot interaction (HRI), and tools for facilitating shared or traded system control. These challenges are also relevant to UAV supervisory control interfaces, as

HRI research has the same general goal of facilitating effective human-automation interaction. A well-designed interface is considered critical to the effectiveness of all operator tasks, including vehicle control, navigation, environment hazard detection and system and health monitoring. The design features of UAV control interfaces may directly influence safety, task performance and operator cognitive workload. Therefore, the importance of the operator interface should not diminish with increases in the level of vehicle autonomy (Fong & Thorpe, 2001). This observation is a primary motivation for the present research. Although advanced forms of automation may be possible in UAV development, unless systems are "fully autonomous", designers need to ensure human-in-the-loop control interfaces support optimal information processing and vehicle management for mission success and resource preservation.

Yesilbas and Cotter (2014) reviewed a number of UAV accident reports and suggested that about 60% of remotely piloted aircraft mishaps involved operation-related human casual factors. The Federal Aviation Administration (FAA) has also reported that the percentage of accidents attributable to human error range from 21% to 68% relative to those attributable to equipment failures (Williams, 2004). There are a few underlying causes for human error, including excessive workload, fatigue, or inadequate training etc. (Gabriel et al., 2016). Among 36 U.S. Army Hunter accidents analyzed by Williams (2004), 15 of them were attributed to human factors issues. These issues included limitations in alerts and alarms, display design and procedural errors. Such issues have also been attributed to high operator workload as a result of poor interface design. A breakdown of human factor issues in Shadow accidents also revealed alarm and display design to account for 80% of all issues

in control of specific systems (Williams, 2004). A lack of access to alarms as well as UAV control display usability can increase operator perceptual, cognitive and motor response demands. A comprehensive interface evaluation tool capable of precisely identifying human factors issues, might facilitate elimination of such issues early in the systems design process and, consequently, prevent damage to, or loss of, systems.

1.2 Empirical studies of UAV operator workload

In the past few decades, most empirical studies of UAV operator workload have focused on approaches to vehicle automation, including level of automation (LOA), mode of automation (i.e., adjustable vs. fixed) and reliability of automation towards reducing cognitive workload. Some studies have also looked at the number of UAVs that can be controlled by an operator at any one time (e.g., Dixon & Wickens, 2003; Galster, Knott, & Brown, 2006; Squire, Trafton, & Parasuraman, 2006). Related to this work, other studies have manipulated interface features in attempts to better support operator capabilities to manage multiple UAVs simultaneously (Miller & Parasuraman, 2007). However, many of these investigations have been limited in focus to the type of cameras, depth cues, and multi-modal displays integrated in system designs.

As examples of camera studies, Scribner and Gombash (1998) compared narrow and wide field-of-view equipment and reported greater user motion sickness with the wide field-of-view condition. Nielsen and Goodrich (2006) compared video-only, map-only and video-map displays and concluded that video-only displays yielded longer task completion times. Menda et al. (2011) compared a vehicle “chase” view vs. an on-board view and concluded that the onboard view produced higher average operator blood oxygenation levels, indicating

higher cognitive workload. Related to depth of cues, Scribner and Gombash (1998) compared monoscopic vs. stereoscopic displays and found that stereoscopic cues in a control display interface resulted in fewer errors, lower stress and were more preferred by users.

Research on multi-model displays has typically made reference to Multiple Resource Theory (MRT; Wickens, 2008) as an attention allocation framework and basis for design. This research has sought to mitigate workload by offloading visual display demands onto attentional resources that may be available through other senses (Prewett, Johnson, Saboe, Elliott, & Coovert, 2010). For instance, Donmez, Cummings, Graham, and Brzezinski (2010) compared sonification and discrete audio alerts in decision aiding. They found that sonification resulted in shorter stimulus reaction times but did not impact workload. More recently, Maza, Caballero, Molina, Peña, and Ollero (2010) emulated a UAV GCS to assess operator multi-modal interaction conditions and found that speech synthesis aiding and location-based audio/haptic assistance improved performance and lowered self-reported workload in a simulated target detection task. Recent studies have also considered natural modalities, such as speech and gestures, to create intuitive interaction with UAV interfaces and potentially reduce operator workload (Peshkova, Hitz, & Kaufmann, 2017). Prewett et al. (2010) reviewed a number of empirical studies on HRI and proposed that optimal visual displays in combination with appropriate audio and tactile feedback should produce better operator performance. However, few studies have looked into how to organize UAV display components for an optimal visual layout, let alone systematically investigate and validate any effects on operator cognitive workload. Although HCI research (e.g., Dix, 2009; Helander, 2014; Molich & Nielsen, 1990; Shneiderman & Plaisant, 1987) has extensively studied

methods by which to improve visual presentation of interfaces, any guidelines resulting from such research may not be directly applicable to UAV supervisory control interfaces due to differences in context of use and user task demands.

In general, an important observation to make on the above identified literature is that the majority of studies have only assessed user cognitive workload by subjective means. Only a few studies applied a secondary-task approach or physiological measures for workload assessment. Çakır, Şenyiğit, Akay, Ayaz and İşler (2012) explored the use of functional near-infrared spectroscopy (fNIRS) to identify differences in performance and cognitive workload induced by various UAV control interface designs (i.e., single camera, dual-camera and projector views). They did not find significant differences in participant rated workload, task difficulty and pressure. However, fNIRS results showed that participants experienced greater task difficulty with a projector view (i.e., flight video displayed by a projector), supported by the fact that higher blood oxygenation levels were observed with this condition. Similarly, Menda et al. (2011) also used fNIRS to compare cognitive workload in operating a UAV with two types of camera views (i.e., an onboard view vs. chase view). They found that the onboard view produced higher average blood oxygenation levels, indicating higher cognitive workload. However, there was no statistical difference in NASA-TLX scores (overall and mental). Afergan et al. (2014) used fNIRS to model the difficulty of a multi-UAV planning and routing task. The task was designed to be either "overwhelming" or "overly simple". fNIRS input was used to intelligently add or remove UAVs in order to provide an "ideal challenge" to the user. The study compared the fNIRS-based adaptations to a control condition where UAVs were randomly added and removed. Across conditions,

participants controlled roughly the same number of UAVs overtime. It was found that participants in the adaptive condition had significantly lower failure rates, traveled significantly less distance, and encountered fewer obstacles.

These studies validated the use of a physiological workload measurement technique and supported sensitivity in detecting cognitive workload differences due to control interface manipulations. However, few other studies have applied objective measures of cognitive workload for assessing UAV operations. In addition, the type of physiological measures presented in the related literature is quite limited. There is a need for investigation of other objective measures of workload to effectively and reliably evaluate UAV system design approaches.

1.3 Research motivation

A lack of research on effective UAV supervisory control interface design motivates the present study. As noted, the majority of prior studies have focused on investigations of vehicle automation as a potential panacea for operator workload issues. Only few studies have examined UAV interface control and configuration as a basis for mediating perceptual, cognitive and motor demands on operators. Adams (2002) pointed-out that the design of the human-robot interface can directly affect an operator's ability and desire to complete a task. Interface design also affects an operator's ability to understand the current task and system situation, make decisions, as well as provide high-level commands to a robotic system and supervise execution. These issues are critical to UAV supervisory control interfaces in which task mixture and pacing can pose high cognitive demands for operators.

Moreover, the manner in which interface design features may influence UAV operator cognitive workload responses is not clear. System-specific manipulations of camera features and forms of multi-modal interface options have been investigated but no studies have conducted comprehensive usability evaluations of UAV control interfaces as a basis for inferring cognitive workload. Any attempt to systematically evaluate a UAV interface design and to predict operator workload induced or mitigated by design features would be worthwhile at this stage of the domain research. The present project attempts to shed light on how supervisory control interfaces can be designed to moderate operator cognitive workload and support safe system operations.

Finally, UAV operator workload assessment, as presented in the existing literature, has been limited to performance and subjective measures. More physiological measures need to be evaluated for sensitivity to interface manipulations and to improve objectivity and reliability of workload assessments as bases for system design improvements.

2. Literature review

In this section, the current literature is reviewed with a focus on three issues, including interface evaluation tools, an approach for enhancing existing tools, and methods of cognitive workload assessment. The review begins with coverage of existing UAV interface evaluation tools (Section 2.1), including an introduction to their development and discussion of advantages and limitations. This section represents a constrained review of tools with demonstrated relevance to UAV systems. Section 2.2 describes an extension of an existing ergonomic guideline-based tool for application to supervisory control interfaces, as developed in a previous study. The objective of this section is to identify a tool structure that has applicability to UAV interfaces. Section 2.3 reviews a range of methods to assess cognitive workload, as well as advantages and limitations of each method.

2.1 Existing interface evaluation tools for UAV

There are currently several tools and methodologies that can be used for evaluation of interface designs for supporting human performance in UAV operations. These tools range from subjective measures of cognitive workload to interface design checklists (or interface scoring systems) to quantify the degree of design adherence to human factors standards. Some of the established tools have been adapted to various domain applications in order to account for specific interface user information processing needs as well as workload demands under different operational conditions.

2.1.1 MCH-UVD

The FAA has used the Cooper-Harper subjective workload rating scale for many years to assess aircraft pilot cognitive demands and to gauge whether a cockpit interface design is

effective from a performance perspective (Cummings, Myers, & Scott, 2006). The Cooper-Harper scale gives an ordinal rating; grading interfaces based on how much additional cognitive demand a design places on pilots. Cummings et al. (2006) adapted the Modified Cooper-Harper scale to the UAV domain (Modified Cooper-Harper for Unmanned Vehicle Display; MCH-UVD). The tool mimics the FAA standard and applies UAV domain specific constraints and issues to the same scale. The MCH-UVD addresses higher-level UAV operator cognitive demands instead of lower-level control skill requirements through a process of diagnosing display issues. The tool was designed to detect an interface's capability to support operator information processing, including information acquisition, information analysis and decision making. The ratings of pilot workload from 1 to 10 represent increasingly difficult human-vehicle display interaction. The goal of developing and applying the MCH-UVD was to identify possible problems with an interface before deployment, such that the quality improvement cycle could be shortened, thus reducing training time and upgrade costs. The structure of the MCH-UVD is presented in Figure 2.1.

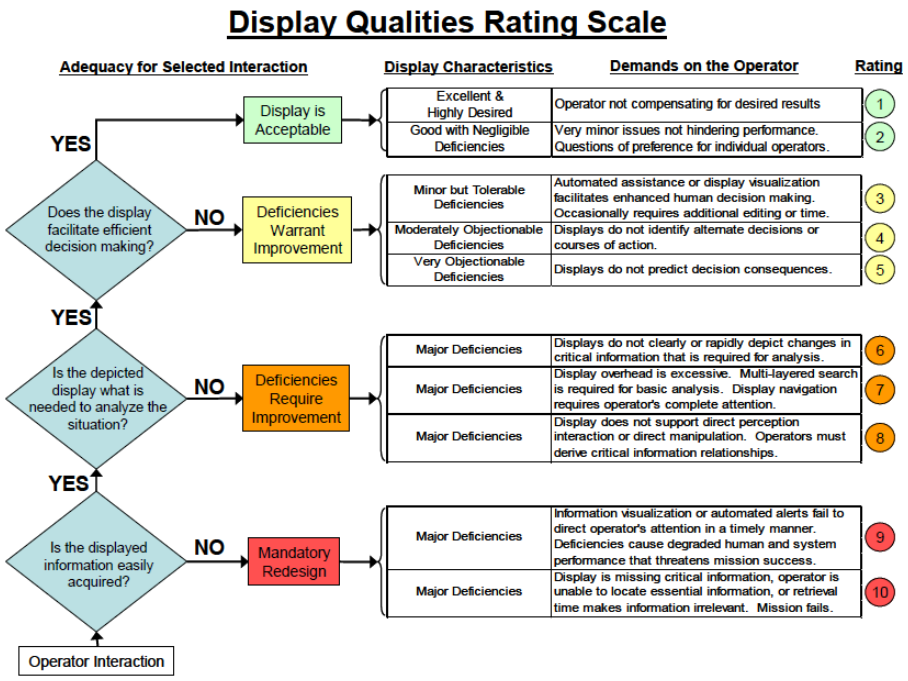

Figure 2.1 MCH-UVD Scale

Cummings et al. (2006) conducted two case studies to determine the effectiveness of the MCH-UVD. In the first study, they compared participant ratings of the interface with verbal comments for three types of tasks (preflight, enemy territory surveillance and a malfunction task) and found responses to be largely consistent. However, there was some variation among the subjective ratings of workload and comments for the malfunction task. In the second study, the MCH-UVD was used to assess a decision support tool designed to assist mission management and in-mission replanning during multiple UAV operations. The decision support tool resulted in better performance. However, even "top-performers" produced ratings of 6 or higher for the system interface, indicating major deficiencies in the design. Despite superior performance, the scores indicated the need to improve the tool, which was consistent with participant comments. In another related study, Donmez et al.

(2010) considered a UAV control interface display with ratings lower than 2 to represent an acceptable display for mission support; any display with a rating of 9 or 10 was considered to require redesign. In the Donmez et al. study, participants supervised one UAV or four UGVs in several operations. The effect of display misuse on mission success was analyzed. Results showed a correlation between fewer display misuses and better MCH-UVD scores. These results were considered as preliminary evidence of the validity of the MCH-UVD tool. Regarding feedback on MCH-UVD, 86% of participants found the tool helpful in identifying UAV display improvements and 32% said they would not have recognized certain interface deficiencies without use of the tool.

These two studies indicated that the MCH-UVD helped identify deficiencies in certain UAV control interfaces by articulating human information processing requirements. Donmez et al. (2010) demonstrated the tool to be a reliable predictor of operator performance. However, it should be noted here that superior task performance is not necessarily a comprehensive and completely reliable indicator of usable interface design. Operator performance can be influenced by other factors, such as experience level, task complexity and luck! In both studies described above, there was no comparison of performance made across different interfaces. Therefore, the validity of the MCH-UVD as an interface evaluation tool still needs to be further investigated. In addition, few other researchers (if any) have yet to apply this tool to evaluate control interface designs.

2.1.2 Usability testing

Usability testing is a broadly defined methodology for assessing whether an interface presents necessary functional features and whether the features are easy to use (Dix, 2009).

Some researchers (Irizarry et al., 2012; Kaber, Riley, & Tan, 2002) have applied usability analysis to UAV interfaces as well as manned aircraft interface design. There are many different methods for testing usability, but most measure learnability, efficiency, memorability, errors, and satisfaction in order to iteratively improve an interface (J. Nielsen, 1996). Popular usability techniques include questionnaires, heuristic evaluation, cognitive walkthroughs, contextual inquiry, the cognitive interviewing method, and focus groups, etc. (Cummings, Pina, & Donmez, 2008). This section presents a constrained review with a focus on methods with relevance to UAV interfaces.

Irizarry et al. (2012) conducted a usability analysis on a small-scale aerial drone for use by safety managers in construction jobsite monitoring and control. The analysis started with an expert evaluation applying Nielson's 10 Heuristics (Dix, 2009) to the drone interface. The heuristic evaluation revealed some user interface problems and generated recommendations. For example, the use of elements of construction safety in the design of icons and names was suggested to provide significant visual cues and facilitate user understanding. In the second part of the usability analysis, drone inspections controlled with an iPhone or iPad were compared with direct visual safety inspection. A usability questionnaire was applied to collect subjective perceptions of usability of the drone option. The research concluded that the iPad drone control interface resulted in comparable inspection accuracy to direct visual inspection and participants were satisfied with the system. As another example, Kaber et al., (2002) conducted a usability inspection of an aircraft flight management systems with expert pilots. Pilots were asked to assess a multifunction control display unit interface in terms of usability principles and identified

violations of learnability, flexibility, and robustness. Violations were translated to design recommendations, including use of global metaphors and maintaining consistency among interface screens in order to minimize pilot working memory requirements and workload and to increase available attentional resources.

These two studies demonstrated usability analysis to be effective for revealing violations of usability principles in unmanned and manned flight operations. However, heuristic evaluation does not directly recommend solutions to identified interface problems (Nielsen, 1994). Usability analysis usually requires inputs from usability experts and delivery of questionnaires to users. Due to this characteristic, usability analysis relies heavily on subjective opinion. Moreover, recruiting users and analyzing questionnaire responses is also time consuming. Therefore, usability analysis is not considered to be a favorable interface evaluation approach for the present research.

2.1.3 Cognitive task analysis

Cognitive task analysis (CTA) is an extension of task analysis techniques and is used to develop an understanding of human knowledge, thought processes, and goal structures that directly impact task performance (Hollnagel, 2003; Schraagen, Chipman, & Shalin, 2000). CTA results can serve as a basis for guidance on redesigning existing systems in many domains (e.g., Kaber, Segall, Green, Entzian, & Junginger, 2006). Kaber et al. (2006) paired goal-directed task analysis (GDTA) with abstraction hierarchy (AH) modeling to characterize the knowledge structure of an expert group of users in research on supervisory control interface design. Five subject-matter experts provided commentary on shortcomings and advantages of an existing control interface and then ranked interfaces from 1 (“low”) to 5

("high") in terms of ten usability characteristics (e.g., complexity, consistency, learnability, ease of use). The researchers formulated guidelines from the GDTA and AH results along with the results of the usability heuristic analysis to create a new interface design. This new interface was tested and achieved higher scores, according to experts, due to its intuitiveness and improved interactivity. However, the authors pointed-out that GDTA and AH modeling required significant time investment for analysts. In addition, the AH model was difficult to develop and might have depended on expert knowledge and understanding of the system. Within the unmanned systems domain, Humphrey and Adams (2011) modified GDTA in its application for an analysis of integration of multiple unmanned systems. They applied GDTA to a chemical, biological, radiological, nuclear, and explosive (CBRNE) event and assessed human responder functioning as system components. The goal was to understand how unmanned systems could be designed for effective integration into the overall domain or system. Enhanced vehicle control interface design requirements were identified from the analysis and used as feedback for a design process (Humphrey & Adams, 2011). Humphrey and Adams also considered the complexity of developing task analysis models to be a major challenge and limitation of this approach.

Both studies reviewed above demonstrated CTA to be a useful tool for identifying operator dynamic goal sets, factual knowledge stores, mental strategies, and critical decisions as a basis for interface redesign (Kaber et al., 2006). However, due to the complexity level of this approach, application time is considerable, especially with complex interfaces that can be used to complete various types of tasks.

2.1.4 Ergonomic design guidelines for supervisory control interface evaluation

Ponsa and Díaz (2007) developed a cognitive ergonomics guideline for evaluating supervisory control interface designs in the context of process control operations in order to improve the efficiency of human interaction with industrial automation. Ponsa and Díaz (2007) reviewed human interface design guidelines related to process control and supervision tasks (see Table 2.1) and proposed a checklist of key design features/indicators. They labelled this set of ergonomic guidelines for supervisory control interface design as "GEDIS" (an acronym for the Spanish phrase, "Guia ergonómica para el diseño de interfaz de supervision", which means "Ergonomic Guideline for Supervisory Control Interface Design").

The GEDIS guide consists of 10 interface design indicators that seek to cover the range of interface design features in the supervisory control domain, including: architecture, distribution, navigation, color, text font, status of the devices, process values, graphs and tables, data-entry commands, and alarms. Each indicator has a substructure including a diverse set of characteristics/subindicators. Any interface can be quantitatively evaluated based on the degree of design satisfaction of the identified subindicators (i.e., characteristics of interface features). A rating scale ranging from 0 = "not appropriate" to 5 = "appropriate" with an intermediate value of "acceptable/moderate" is used for this purpose. As an example, the design indicator of interface "Navigation" can be evaluated based on two subindicators: (1) relationship with architecture, and (2) navigation between screens. Each of the two subindicators can be assigned a score of 5, 3 or 0 depending upon whether each design

characteristic is considered to be “appropriate”, “acceptable”, or “not appropriate” for supporting human use. In some cases, a score is assigned based on the presence of certain features. For instance, with respect to the interface “Architecture”, a subindicator of “Map existence” can only be assigned a score of 5 or 0, indicating that a map exists or not. The score of each indicator is the weighted average of all subindicator ratings (see Equation 1).

$$Indicator\ Score = \frac{\sum_j^J w_j Subindicator_j}{\sum_j^J w_j} \text{ (Equation 1)}$$

where J = number of subindicators, w_j = weight of the j-th subindicator.

Similarly, a global evaluation index for the entire supervisory control interface can be obtained by calculating the weighted average of indicator scores (see Equation 2).

$$Global\ Evaluation\ Index\ (GEI) = \frac{\sum_i^I p_i Indicator_i}{\sum_i^I p_i} \text{ (Equation 2)}$$

where I = number of indicators, p_i = weight of the i-th indicator.

Weighting factors can be determined based on expert opinion of the direct relevance of any given design indicator to human information processing in the target interface application. The researchers applied the GEDIS to a Sugar Technology Center simulator interface (Ponsa, García & Díaz, 2010; Ponsa & Díaz, 2007) and generated some design recommendations in order to achieve a maximum GEI value, which was considered to represent optimal interface design. It was suggested that the GEI of an interface should have an initial value of 3-4 and designers should ultimately work to achieve a score of 5 with the help of the GEDIS checklist (Ponsa & Díaz, 2007).

Table 2.1 Guidelines reviewed for GEDIS and GEDIS-UAV

GEDIS	GEDIS-UAV
ISO 11064-7:2006 (Ergonomic design of control centres - Part 7: Principles for the evaluation of control centers)	ARINC 601 - Aims to normalize the definition of a Cockpit Display System
Human Factors Design Standard (HFDS)	STANAG 4586
Nureg 0700: Human Interface Design Review Guidelines	DO-178B - Software Considerations in Airborne Systems and Equipment Certification
I-002 Safety and Automation Systems NORSOK	Joint Architecture for Unmanned Systems (JAUS) - JAUS Unmanned Ground Vehicle Service Set
Man Systems Integration Standard (NASA-STD-3000)	Joint Architecture for Unmanned Systems (JAUS) - JAUS HMI Service Set
	ISO 9241-11: Ergonomic requirements for office work with visual display terminals - Part 11: Guidance on usability

2.1.5 GEDIS-UAV

Lorite, Muñoz, Tornero, Ponsa and Pastor (2013) adapted the original GEDIS, as described above, for application to evaluation of UAV graphical user interfaces (GUIs). In their effort to create a GEDIS-UAV evaluation tool, the researchers reviewed standards and guidelines related to the design of GUIs for UAVs (also see Table 2.1). The desired design features/indicators were carried forward from the original GEDIS, however, modifications were made to desired subindicators (feature characteristics) to accommodate UAV interface technology. For example, under the indicator “Architecture”, subindicators, including “number of display levels” and “Map existence” were removed, and “Screen number” was added. The GEDIS-UAV inherited the scoring system of the GEDIS, where the GEI and indicator scores remain weighted averages.

The GEDIS-UAV was subsequently used to evaluate a UAV ground control interface, which was designed based on the ARINC 661 standard, comprising a display mounted on a

control panel similar to a conventional aircraft cockpit. The initial GEI for the interface was 3.8. A set of design changes were proposed based on the evaluation results and indicator scores, including redistributing the components of the interface, improving data visualization, ameliorating alarm content and adjusting font sizes. These modifications brought the GEI to 5, which is the maximum value of the rating scale. The researchers expected the enhanced interface to minimize the possibility of human error in use (Lorite et al., 2013).

There are several issues with the development of the GEDIS-UAV tool that motivate the need for enhancement. The GEDIS-UAV provides a checklist-based tool for evaluating GCS interfaces. The quantitative outcomes of the tool allow for straightforward assessment of the degree of conformance of specific interface designs with existing guidelines, as well as quick comparisons among different designs. Moreover, in application of the tool, subindicators with the lowest ratings can be identified as targets for design improvements. However, there are a few issues related to application of the GEDIS-UAV. First, the original development did not provide solid justification for the set of desired interface design indicators/features. This drawback resulted in confounding indicators and an absence of some important UAV interface indicators, such as maps and navigability features. From a design effectiveness standpoint, it is important that an evaluation method take into account multiple aspects of an interface design, including usability features, functional capabilities, operator information processing quality, and resulting cognitive workload (Zhang, Feltner, Shirley, & Kaber, 2016). To comprehensively reveal design deficiencies, it is necessary to have an overarching framework for indicator identification, such as usability principles (e.g., Nielsen & Molich, 1990).

The second major limitation of the GEDIS-UAV is that the selection of subindicators (i.e., characteristics of design features) was not supported by detailed references to design standards. The subindicators are specific characteristics to be evaluated to determine the level of conformance of a design with each indicator. However, the set of subindicators under each indicator was not a complete representation of all characteristics related to UAV interface design components/features. It is possible certain characteristics were chosen over others because they were more important in terms of supporting operator performance. However, there was no explanation for subindicator selection provided by Lorite et al. (2013). Related to this, some the existing characteristic criteria do not represent a reasonable basis for interface evaluation. For example, the subindicator "number of screens" is not an absolute measure. Instead, it can be influenced by screen size and resolution. If one screen is large and clear enough, there is no need for additional screens. This issue also results in overlap of subindicators. For example, "graphs and tables" includes "format", visibility", "location" and "grouping" as specific characteristics for evaluation. However, these are not stand-alone features. Format of data presentation restricts its location and vice versa. Visibility of interface items heavily depends on format and location on a screen. To clarify the selection of subindicators, references from design guidelines (e.g., HFDS) and the existing empirical literature is necessary.

A third limitation of the GEDIS-UAV tool is that the manner of assessment of an interface, in terms of each subindicator (i.e., appropriate, moderate, not appropriate) is entirely subjective. Lorite et al. (2013) referred to specific interface characteristics in making ratings of subindicators. However, it is not clear how design characteristics relate to ratings.

In the case of the subindicator “absence of non-appropriate font combination”, the answer from different analysts could vary dramatically depending on eyesight, age, cultural background, reading preference or sensitivity to font characteristics. Without any reference to design guidelines, claiming that certain features are appropriate or not is not a justified or valid approach to interface design evaluation. As a result, the evaluation process can be highly vulnerable to personal preferences and rater emotional states (Zhang, Feltner, et al., 2016), instead of being an objective indicator of how the interface design supports performance.

As a fourth limitation, the language used in presenting many of the subindicators as part of the GEDIS-UAV tool is unclear and could lead analysts to make wrong judgments about specific designs. For example, the GEDIS-UAV inherited the indicator of “navigation” from the GEDIS tool, which referred to “the logical linkage of interface items for supervisory control of industrial automation.” However, when it comes to the UAV domain, interface navigation has a completely different meaning; that is, planning and controlling a UAV route vs. monitoring process flow channels and pump settings. Another example of confusing language among GEDIS-UAV subindicators can be found in the subindicator, “relationship with text.” It is unclear from the word “relationship” that this guideline is referring to how color influences the presentation of text and what colors are appropriate for interface design. The use of double negation is another critical language issue in GEDIS-UAV evaluation criteria. For example, “absence of small fonts (smaller than 8 pt.)”. Such criterion can increase analyst interpretation effort. It would be much more straightforward to require a sufficient font size (greater than 8 pt.).

2.1.6 Modified GEDIS-UAV

In order to address the above identified issues with the GEDIS-UAV, both a "top-down" and "bottom-up" approach were taken to enhancing the identification and organization of interface design indicators as well as specification of design characteristics for interface evaluation. The top-down approach provided an overarching structure for modifying and classifying GEDIS-UAV indicators, including two types of indicators: (1) usability-related indicators identified based on usability heuristics (Nielsen & Molich, 1990; Norman, 2013; Shneiderman & Plaisant, 1987); and (2) functionality-related indicators, based on functional interface components observed from various UAV studies (e.g., Afergan et al., 2014; Chen, Barnes, & Harper-Sciarini, 2011; Cummings et al., 2006; Fahlstrom & Gleason, 2012; Fuchs, Borst, de Croon, van Paassen, & Mulder, 2014; Lorite et al., 2013; Lu, Horng, & Chao, 2013; Menda et al., 2011). The modified set of desired interface design indicators are summarized in Table 2.2. The usability-related indicators are relevant and may be directly applicable across various platforms (e.g., unmmaned ground vehicles). The functionality-related indicators, however, are specific to UAVs and need to be modified in order apply to a different domain.

Table 2.2 Indicators in Modified GEDIS-UAV

Usability Related	Functionality Related
Display Layout (DL)	Map and Navigation (MN)
Information Presentation (IP)	Status and Devices (SD)
Color (C)	Data Entry Command (DEC)
Text (T)	Alarm (A)
Physical Control (PC)	

The bottom-up approach was used to identify an elaborate list of potential design subindicators from the existing standards literature. A number of industrial guidelines were reviewed (see Table 2.3) as part of this approach. A subset of guidelines was selected for each indicator (resulting from the top-down approach) by considering a set of selection rules, including: (1) the guideline is applicable to the UAV domain; (2) the guideline addresses design issues relevant to UAV supervisory control interfaces (and not robotic systems, in general); (3) the guideline can be applied to contemporary technologies used in UAV supervisory control interfaces (e.g., guidelines for Cathode ray tube displays were not included); and (4) the guidelines are sufficiently descriptive to allow for assessment of the degree to which a particular interface design is conforming or non-conforming. If multiple guidelines identified through the bottom-up approach described the same interface content, they were combined as one criterion. If there were conflicts among guidelines, the guideline referring to an interface use scenario closest to UAV operation was adopted.

Among all the guidelines selected for assessment of each interface design indicator/feature/component, those addressing the same (or similar) interface feature characteristic were grouped under a subindicator. The subindicators were named based on the design characteristics commonly referenced in the guidelines. The guidelines under each sub-indicator serve as criteria for evaluation of interface conformance with each subindicator (i.e., design characteristic). As an example, “Menu structure” is a subindicator under the indicator “Display Layout”. The criteria for assessing “Menu structure” are listed in Table 2.3 along with identification of the original design guidelines from which the specific criteria were derived.

Table 2.3 "Menu Structure" Subindicator under "Display Layout" Indicator

Indicator: Display Layout		
Sub-indicator	**Criteria for sub-indicator**	**Reference Guidelines**
Menu Structure	The interface provides an appropriate maximum number of options for different types of graphical controls: (a) Radio buttons: 1-6 options; (b) Static Menus: 3-10 options; (c) Menu Bars: < 10 options; and (d) Scrolling Menus: >10 options.	HFDS 8.7.5.1; HFDS 8.7.5.3.4; HFDS 8.7.5.7.4;
	The number of selections required to reach the desired option in complex menus is no more than 4 steps.	UAS_GCS_HMI 3.2.2.4
	When a user selects a menu option and no computer response is immediately observable, the software provides some other acknowledgment of the selection.	HFDS 8.7.5.3.9
	Menu options are presented in a single vertical column, aligned and left justified (exception: menu bars).	HFDS 8.7.5.6
	Destructive commands (e.g., delete, exit) are placed at the bottom of menus.	HFDS 8.7.5.6
	Options for opposing actions (e.g., save and delete) are not placed adjacent to each other.	HFDS 8.7.5.6
	Primary windows' menu bars extend the full width of the primary window.	HFDS 8.7.5.7.2
	System menus include the following options: end a session, review system status, define user preferences, manage alerts, and change a password.	HFDS 8.7.5.8

After all relevant criteria were identified from the existing UAV and human factors design standards, the phrasing of each design characteristic and the meanings were edited for clarity. This process was intended to ensure no conflicts among criteria and that all criteria referred to one (and only one) design characteristic. Phrasing was also edited to support analyst ease of interpretation and determination of whether a specific design was conforming or non-conforming. Currently the Modified GEDIS-UAV contains 290 criteria (with all

being supported by references), which provide a comprehensive list for evaluating UAV supervisory interface design characteristics.

2.1.6.1 Scoring mechanism of Modified GEDIS-UAV

The Modified GEDIS-UAV has a checklist structure. A set of design criteria is presented in the form of a checklist for each subindicator and provides a basis for objective assessment of interface designs. Given the above identified limitations of the GEDIS-UAV approach, in order to evaluate a design subindicator with the Modified GEDIS-UAV tool, an analyst compares the actual interface design characteristics with related criteria and determines whether each criteria is met or not. That is, the analyst assigns a binary score ("Yes (Y)" if met, "No (N)" if not) for each criterion. If the characteristic described by a criterion is not related to the interface under evaluation, the analyst assigns a response of "NA" and the criteria is not considered in scoring of the interface design. In this way, the Modified GEDIS-UAV eliminates the need for an analyst to provide a subjective rating of each interface design characteristic with multiple levels. The approach also eliminates the needs for justification or rationale for different ratings on characteristics. The Modified GEDIS-UAV requires calculation of the percentage of criteria satisfied by an interface design. The revised equation for calculation of a sub-indicator score is shown below.

$$Sub - indicator\ Score = \frac{\#\ of\ "Y"}{\#\ of\ "Y" + \#\ of\ "N"} \quad \text{(Equation 3)}$$

On this basis, an indicator score becomes the average of all relevant subindicator scores. That

(a) Digital display

(b) Data wall display

Figure 2.2 Interfaces with different vehicle status displays

is, for each desired design feature a percentage of interface conformance with guidelines for specific characteristics of the feature/component is calculated. Subsequently, the GEI is calculated as the average percent satisfaction of desired design features for an interface under evaluation. The GEI is simply obtained by averaging all the indicator “scores”.

The Modified GEDIS-UAV was subsequently applied to ArduPilot Mission Planner interface to test its applicability and sensitivity (Zhang, Feltner, Shirley, Kaber, &

Swangnetr, Submitted). The Mission Planner is a full-featured GCS application that presents a graphical user interface for UAV supervisory control. Three human factor experts and novice analysts reviewed videos of interface use in a few typical UAV operation tasks (e.g., setting a flightpath waypoint, monitoring vehicle status, altering vehicle flightpath). Each of the analysts evaluated the interface, independently based on the features presented and used in the videos. They were allowed to review the videos multiple times. It took 2-2.5 hours for an expert analyst to complete the evaluation. The analyst evaluation results were used as a basis for an inter-rater reliability assessment. An intra-class correlation coefficient (ICC) was calculated for the GEDIS indicator scores. This testing revealed inter-rater reliability in application of Modified GEDIS-UAV to be moderate (ICC = 0.429) with human factors experts and low (ICC = 0.204) for novice analysts. De Vet et al. (2011) considered ICC >= 0.7 to be acceptable for clinical studies. The relatively low ICC in Modified GEDIS-UAV evaluation was found to be mostly attributable to disagreement among analysts in identifying those guidelines applicable to the target UAV interface. That is, while one analyst assigned a score for some criteria ("Y" or "N"), another analyst might have considered the same design criteria to be irrelevant to the system interface. To address this issue, the three expert analysts were required to collaboratively identify relevant criteria based on their knowledge of the MP platform and perform the evaluation again independently. Following this approach, the ICC was found to increase to 0.834. It was also recommended that the tool be applied for interface evaluation by human factors experts having some familiarity with usability issues in interface design and design standards, in general, as well as some training on use the Modified GEDIS-UAV.

This study also revealed sensitivity of the Modified GEDIS-UAV to interface feature manipulations. Two interfaces with different displays of vehicle status were compared. One interface presented key vehicle status indicators (e.g., ground speed, altitude) using a concise digital format, whereas the other interface presented all vehicle status via a data "wall" (see Figure 2.2). Results revealed the concise digital display to produce significantly and consistently higher analyst scores for the design indicators of "Text", "Color" and "Status and Displays".

2.1.6.2 Advantages of the Modified GEDIS-UAV

The Modified GEDIS-UAV tool addresses both usability and functional features of a UAV supervisory control interface in evaluating interface conformance with existing design standards. In addition, the tool has the advantage of providing a quantitative assessment of an interface configuration, which allows for straightforward comparison among different interface designs. Moreover, the modified tool limits subjectivity in the design evaluation process by identifying detailed guidelines as evaluation criteria and requiring binary evaluation of interface conformation with those criteria. The evaluation results can be used to target various interface design weaknesses and guide feature redesign. The evaluation technique is also relatively easy to apply and less time consuming than usability inspection or cognitive task analysis. In addition, the tool can be applied to the range of UAV control interfaces, from smart mobile device-based control to a cockpit-like control station, by adjusting relevant criteria according to control platform characteristics. All these advantages make the Modified GEDIS-UAV an effective tool for UAV supervisory control interface evaluation.

2.2 Cognitive task complexity and workload assessment techniques

There are many different tools presented in the human factors literature that have been developed for assessing human task demands and cognitive workload in order to provide some basis for assessing and validating complex system design approaches. This section provides a review of the various types of workload analysis methods with potential relevance to use of supervisory control interfaces in UAV operations.

2.2.1 Primary and secondary task performance measure

Primary task performance measures can be used to identify the degree or level of human task difficulty. It is expected that a task posing higher workload will be more difficult, resulting in degraded performance, as compared with a low workload task. Primary task performance is effective as an index when a task: imposes long periods of workload, can lead to overload, and is sensitive to individual differences in resource use (Young & Stanton, 2004). However, primary task performance has been found to be insensitive for revealing workload differences at low and medium difficulty levels. That is, if an increase in workload is within the attentional capacity of an operator, performance may not be degraded. Therefore, primary task performance by itself is often not a reliable measure of workload in isolation.

In order to address this issue, secondary-task techniques have been developed and are typically used in conjunction with primary performance, in order to infer the level of mental workload in a task. In the use of primary and secondary task measures, participants are instructed to maintain consistent performance of the primary task while attempting to perform the secondary task when primary task demands allow. Secondary tasks must be carefully designed in order to ensure a measure of spare attentional capacity. Regarding

secondary task design, researchers have concluded that a secondary task demonstrates greatest sensitivity when there is considerable overlap in the processing demands of the primary and secondary task (Wierwille & Eggemeier, 1993). Brown (1978) suggested that researchers adopt discrete stimuli that occupy the same attentional resource pools as the primary task in order to ensure the technique is measuring spare capacity and not an alternative resource pool. As an example of the use of secondary task measures of workload, Grant, Carswell, Lio and Seales (2013) proposed a method for measuring surgeon mental load using a simple secondary task in a surgical context. The primary task was a surgical training task in which participants manipulated small objects with surgical graspers. The secondary task required participants to say the word "time" each time they thought that 21 seconds had elapsed. Participants were instructed to complete the surgical task as quickly and accurately as possible and to attempt to make the intervals in the time-keeping task (secondary task) as accurate and consistent as possible. It was emphasized for participants that performance of the surgical task was more important than the time-keeping task. Grant et al. (2013) found this simple approach to workload assessment, using the secondary task, to be sensitive to primary surgical training task manipulations.

In summary, secondary tasks have the advantage of being able to discriminate between tasks when no differences are observed in primary task performance. Such measures are also useful for quantifying short periods of workload, limited spare attentional capacity and even automaticity (Young & Stanton, 2004). However, the primary limitation of such measures is that they can be intrusive to primary task performance, particularly when the workload of the primary task is low. This intrusion can cause an operator to modify

allocation of processing resources devoted to the primary task under some circumstances (Wierwille & Eggemeier, 1993). Therefore, the secondary-task approach is not an ideal option for analysis of operator workload in UAV supervisory control.

2.2.2 Physiological responses

2.2.2.1 Review on physiological workload measures

Physiological measures of mental workload make use of various bodily response measures, including heart rate, heart rate variability (Jorna, 1992; Roscoe, 1992), eye movements and pupillary responses (Backs & Walrath, 1992), and event-related potentials (Kramer, Trejo, & Humphrey, 1996) as indices of mental effort. These measures offer promising prediction of task performance and workload and they may be useful for identifying and diagnosing influential factors in unmanned systems control performance (Abich, Matthews, & Reinerman-Jones, 2015).

Electro-encephalography (EEG) is a measure of brain activity. It has been found to vary in a highly predictable manner in response to sustained changes in task load and associated changes in mental effort required for task performance. EEG features can be analyzed in both time and frequency domains. Event related potentials (ERPs), such as the P300 (a positive wave form with a 300 ms lag relative to a perceptual stimulus) is a commonly used time domain feature to predict cognitive workload (Putze & Schultz, 2014). In a flight simulation experiment, Song, Zhuang, Song and Wanyan (2011) recorded primary and secondary task workload measures along with pilot EEG during takeoff, cruise and landing activities. They found that the main effect of task workload was significant in peak EEG amplitude and peak latency of a P300 ERP. Under higher mental workload, the ERP

peak amplitude decreased and peak latency extended as compared to responses under a lower task workload condition.

Although EEG analysis is a highly favored workload measurement technique for many researchers, the methodology has several limitations. EEG signals are easily contaminated by body movements, including eyeball movement and blinking. In UAV operations, a considerable amount of operator eye and hand movement is often required for vehicle status monitoring and operator manipulations of controls for achieving specific vehicle states. Operators are also often required to communicate with control team members and report mission status, which involves facial muscle use and can also corrupt EEG signals. All such movements will introduce artifacts in EEG response data. Artifacts require filtering through elaborate signal processing methods. This process also typically results in the loss of some meaningful signal information. Finally, collection of EEG responses also requires a participant to don an electrode cap with an attached wiring harness, which can be uncomfortable and obtrusive to task performance. Based on these facts, EEG-based workload assessment is not considered to be a select methodology for assessment of human workload in UAV supervisory control operations.

Functional near infrared Spectroscopy (fNIRS) is another measurement technique that can be used to assess human cognitive workload in terms of blood oxygenation levels. fNIRS systems integrate light-emitting diodes placed on the surface of the skull that transmit near-infrared wavelengths of light approximately 5 mm into frontal cortical tissue (Gratton & Fabiani, 2006). Measurement of the returning light allows for monitoring of changes in blood oxygenation and blood volume related to human brain function (Menda et al., 2011).

Previous research has shown fNIRS to be a sensitive measure of physiological stress (De Joux, Russell, & Helton, 2013). Ayaz et al. (2012) validated fNIRS in assessing mental workload of humans performing standardized (n-back) and complex cognitive tasks (air traffic control). Typically, performance of cognitively demanding tasks leads to elevation of regional cerebral oxygen saturation (rSO2), which may be tentatively attributed to mental effort (Matthews, Reinerman-Jones, Barber, & Abich, 2015).

fNIRS has advantages over EEG in that it is more affordable, portable and less sensitive to muscle tension (Izzetoglu, Bunce, Onaral, Pourrezaei, & Chance, 2004). For this reason, it has been applied in several studies to assess workload in UAV operation (e.g., Afergan et al., 2014; Çakır et al., 2012; Menda et al., 2011). However, this technique also suffers from noise contamination and is particularly vulnerable to artifacts caused by head motion. Displacement of sensors caused by head movement can cause capture of ambient light or light that has passed directly from a source to a detector without passing through tissue (Izzetoglu et al., 2004) leading to anomalous response readings. Therefore, elaborate signal processing is also required in order to achieve "clean" response measures. Consequently, this methodology is also not considered to be a select workload assessment technique for UAV supervisory control operations.

As another physiological response measure, heart rate has been frequently used as an indicator of human workload (Wierwille & Eggemeier, 1993). Researchers have previously used normalized heart rate (vs. absolute values) to assess cognitive workload. Normalized HR can be calculated by determining the mean difference between a participant's raw heart

rate during a test trial and a baseline measurement. Such normalization is typically necessary to account for individual differences in physiology or fitness level.

Heart-rate variability (HRV) is another physiological response measure that has also been used as an indicator of overall demands placed on human cognitive resources during task performance. HRV is the beat-to-beat fluctuation of heart rhythm and has been found to be a good indicator of cognitive workload. Such physiological measures have been used to study changes in energetical state as a function of stressors (e.g., noise, sleep deprivation, drug use) or to index workload as a function of task parameters as well as subject involvement (Jorna, 1992). Spectral analysis is a preferred method to obtain HRV measures because it reveals the total frequency content of cardiac signals (Jorna, 1992). Related to classifications for HRV frequency ranges, the Task Force of the European Society of Cardiology (1996) recommends the categories of ultra-low frequency (ULF; < 0.003Hz), very-low frequency (VLF; 0.003 Hz ~0.04 Hz), low frequency (LF; 0.04 Hz ~ 0.15 Hz) and high frequency (HF; 0.15-0.4 Hz). The powers of these frequency components are useful in making inference on human autonomic activities. The autonomic nervous system has two branches: in general, the sympathetic system is important in situations of stress; the parasympathetic systems is primarily involved in maintaining state (Brodal, 2004). Parasympathetic activity is considered to be a major contributor to the HF component of the HRV response (Task Force of the European Society of Cardiology, 1996). Therefore, a decrease in the power of HF components would indicate an increase in stress or cognitive workload. However, the interpretation of the LF component remains controversial: some studies suggest that LF, when expressed in normalized units, is a quantitative marker for

sympathetic modulations, while other studies view LF as reflecting both sympathetic and parasympathetic activity (Task Force of the European Society of Cardiology, 1996). Some researchers (e.g., Harriott, Buford, Zhang, & Adams, 2015; Swangnetr & Kaber, 2013) have used the mean ratio of low-frequency HRV over the high frequency HRV (LF/HF ratio) to represent the level of sympathetic activity in relation to parasympathetic activity and found the ratio to increase with an increase in the level of mental workload (Hwang et al., 2008). However, the physiological interpretation of this indicator remains unclear.

As an example application, Brookings, Wilson and Swain (1996) measured heart rate and HRV for assessing cognitive workload in air traffic control (ATC) tasks. They used a computer-based ATC simulation and controlled the levels of scenario difficulty by varying the number of aircraft to be handled, the arriving to departing flight ratios, and time limit for handling a task. The HRV in the high-frequency band showed sensitivity to the task difficulty manipulation. In addition to ATC research, HRV has also been considered as a reliable measure for assessing pilot workload (Roscoe, 1992). Similarity in the contexts of manned and unmanned vehicle piloting make HRV a good candidate for assessing workload in UAV operations, where no strenuous body movement is involved. In addition, the technique is affordable and relatively simple to use.

There are, however, a few essential requirements in the application of HRV. First, spectral analysis is necessary to ensure the stationarity of the signal. A sudden disruption of the HRV signal (e.g., standing up) can produce undesired frequency shifts. Spectral analysis can be used to effectively reveal such frequency shifts. Related to this issue, experimental conditions should be designed to be as stable as possible with regard to task demands and

complexity. It is also important to carefully control the measurement period. Long task exposure periods can result in fluctuations in signals that can affect stationarity. However, selecting too short of an exposure period may degrade signal reliability and analysis, as the resolution of low-frequency component capture may not be sufficient (Jorna, 1992). Typically, 5 min recordings are recommended as a practical solution (Jorna, 1992; Task Force of the European Society of Cardiology, 1996).

Other research has explored eye-tracking as a tool for human cognitive workload assessment in various application domains. Specific measures of workload are often based on pupillography, such as pupil diameter. Pupil size has been shown to reflect processing load or mental effort (Beatty, 1982). Mental activities, such as recollection from memory, attention and thinking, can result in pupil dilation and pupil size returns to baseline within a few seconds of completing mental work (Beatty & Lucero-Wagoner, 2000). More specifically, increase in pupil size has been shown to strongly correlate with increases in mental workload for individuals (Hess & Polt, 1964; Reiner & Gelfeld, 2014). Related to this, Schulz et al. (2011) assessed anesthetist workload in a full-scale operation simulation with multiple test sessions. The sessions differed in critical incident severity, which was used to manipulate the mental workload response. It was found that pupil diameter increased with the severity of the simulated critical incident. In another study, Palinko, Kun, Shyrokov, and Heeman (2010) found that the percent change in pupil size had a higher value for mentally complex tasks than easier tasks and might be more sensitive to task demand changes than the pupil size measure.

It is important to note here that a number of researchers have found that pupil diameter can be affected by ambient illumination levels (Marshall, 2002). Moreover, pupil diameter changes related to mental workload are of a relatively small magnitude, and can be overwhelmed by larges changes due to the pupillary light reflex (Ahlstrom & Friedman-Berg, 2006). Beatty (1982) suggested that influences of light reflex on pupillary response must be considered whenever subjects are permitted to shift their gaze across a non-uniform visual field, such as a photograph.

Blink rate is another pupillography measure that has been evaluated as an indicator of cognitive workload demands in task performance. Blink rate is usually quantified as the number of blinks per minute. Decreased blink rate has been associated with the requirement to extract information from a visual environment (Kramer, 1990). Sirevaag et al. (as cited in Kramer, 1990) found decreases in blink rate when subjects transitioned from a single to dual-task scenario. However, Wierwille, Rahimi and Casali (as cited in Kramer, 1990) found increases in blink rate when navigational demands of a simulated flight task increased. Based on these discrepancies in results, Kramer (1990) suggested that more theoretical and empirical research needed to be conducted before blink rate could be recommended as a measure of mental workload.

In addition to the blink rate response, blink duration has been found to decrease with increases in task demands (Kramer, 1990; Wierwille & Eggemeier, 1993). In an ATC study, Ahlstrom & Friedman-Berg (2006) used blink duration to predict operator workload in controlling different numbers of aircraft with different settings of a weather data display. They reported that blink duration decreased (i.e., the eyes remained open longer) when the

operators controlled an increased number of aircrafts. Furthermore, the blink duration was found to be lower when the weather data display was absent, indicating a higher workload level for the ATC task. Benedetto et al. (2011) also explored the application of eye-tracking responses in a study related to in-vehicle information systems usage. Participants performed a lane change task in a driving simulator under single- and dual-task conditions. The secondary task was a visual search task in which level of difficulty ("easy" vs. "difficult") was manipulated in terms of the number of distractors. Results revealed shorter blink duration for the dual task condition and the "difficult" visual search task. Researchers concluded that blink duration was a more sensitive and reliable indicator of driver visual workload, as compared with blink rate.

Due to the design of remote contemporary eye-tracking technologies, this methodology has the advantage of being non-obtrusive to user task performance and visual behavior, as compared to other workload measurement techniques (review above). The approach is also insensitive to artifacts from other body movements. The literature has provided some support for use of different pupillography measures as indicators of cognitive workload. Although disagreements exist in research results, eye-tracking-based measures of workload appear to be a promising approach for application in studies of UAV supervisory control.

2.2.2.2 Applications of workload assessment in related domains

Physiological measures have been adapted to various human-automation system domains as indicators for predicting performance and workload in cognitively demanding operations. Abich et al. (2015) investigated physiological indices and subjective stress measures as

predictors of performance and workload within two simulated unmanned ground vehicle (UGV) tasks. They applied a host of physiological measurement equipment to study participants, including EEG, electrocardiogram (ECG), fNIRS, and eye tracking sensors. In regard to subjective measures, they used the NASA Task Load indeX (NASA-TLX) and Dundee Stress State Questionnaire to assess perceived workload and stress responses to task conditions. Results showed high internal consistency and reliability among all response measures. They also found a general trend of higher workload responses under task conditions (i.e., threat detection, change detection, or dual-tasking) relative to the baseline condition (i.e., no task). This study demonstrated sensitivity of the applied workload measures; however, it was suggested that a multi-variate battery of physiological and subjective workload measures is needed for performance prediction (Abich et al., 2015).

Harriott et al., (2015) investigated workload differences in human-human vs. human-robot teaming in a reconnaissance task by using a combination of workload metrics. The applied physiological metrics included HR, HRV and respiration rate. They also used secondary task questions as a workload indicator of primary task performance. In addition, NASA-TLX and In Situ Mental Workload Ratings were applied to assess perceived workload. The authors did not find significant differences in task performance, but the physiological responses showed sensitivity for detecting workload differences. In general, both of these applications of workload measurement techniques indicate use of physiological and subjective response ratings may be useful for studying workload implications of supervisory control interface designs for UAV operations.

2.2.2.3 Summary

Physiological measures provide for continuous monitoring of human workload states and offer increased sensitivity in measurement as compared to other performance-based measures as well as subjective ratings. Several physiological measures also provide the advantage of not interfering with primary task performance. Nevertheless, some physiological indicators of cognitive load can be easily confounded by extraneous interference (e.g., physical workload; Young & Stanton, 2004) and influenced by individual differences. Beyond this, some equipment used to collect responses may be physically obtrusive to subject behavior. Due to these limitations, prior research has recommended that a combination of subjective and objective measures be used to provide the most complete and reliable cognitive workload assessment possible, given existing measurement technologies (Abich et al., 2015; Harriott et al., 2015).

Given the pros and cons of the workload measurement techniques reviewed above, cardiac and eye-tracking responses appear to have the most promise for sensitivity and reliability for the present research on UAV interface design and use. Additional subjective measures of cognitive workload are reviewed in the following section.

2.2.3 Subjective ratings of workload

Subjective ratings are another commonly used method for workload assessment (Hart & Staveland, 1988). Ratings have been found to be sensitive to perceived difficulty, concurrent activities, and demand for multiple resources (Hockey, Briner, Tattersall, & Wiethoff, 1989). Subjective measures also have the advantage of sensitivity to changes in effort when such effort maintains primary task performance at stable levels (Hockey et al., 1989).

Methodologically speaking, subjective ratings can be unidimensional and multidimensional. Unidimensional measures tend to be simpler to apply and analyze, but they offer only a general workload score. Multidimensional measures provide some diagnostics for identifying sources of mental workload (e.g., cognitive vs. physical demands), however the procedures for administration are typically more complex (Young & Stanton, 2004). Although such subjective ratings are generally reliable and simple, they also have limitations. They typically can only be administered after task performance, which can influence the reliability of ratings for evaluation of workload during long-duration tasks. Furthermore, metacognitive processes, such as self-appraisal of one's own performance, can cloud accurate reporting. It can also be difficult to make absolute comparisons between individual ratings due to internal differences in perceptions of workload and scaling of responses (Young & Stanton, 2004).

An example of a popular subjective rating method is the Subjective Workload Assessment Technique (SWAT). The SWAT was developed by Reid and Nygren (1988) and is a commonly applied method. The SWAT measures workload in three dimensions – time pressure, effort and task stress. Interval integer scales (ranging from 1 to 3) are used. Subjects rank in ascending order 27 index cards with different combinations of the three dimensions, which represent workload. However, the scales do not allow for fine discrimination of different workloads. Output values of the technique are discrete and not conducive to simple parametric statistical analysis.

The NASA-TLX is another workload subjective rating system that was developed by Hart and Staveland (1988). It also involves ranking of workload demand components (Effort,

Performance, Frustration, Temporal demand, and Mental and Physical demand) in terms of perceived importance to a task. A total of 15 pair-wise comparisons of demands are used. Each demand component is then rated subsequent to task performance by using a 5-in bi-polar scale with anchors of "low" and "high". Rankings and ratings are combined to compute an overall workload score. Currently, computerized versions of this technique are available.

The Modified Cooper-Harper (MCH) Scale was previously identified in the review of system usability evaluation methods with reference to the MCH-UVD, as developed by Cummings et al. (2006). The MCH is yet another subjective workload analysis technique that is particularly applicable to dual-task performance (e.g., pilot evaluation of new aircraft in terms of flight handling and stability). The MCH uses a 10-point unidimensional rating scale that yields a global rating of workload. A decision tree can be used to assist a rater in determining the most appropriate rating to assign to a system. The MCH scale has the advantage of requiring less time for workload ratings than some other techniques. According to some research, it is also a measure that has great utility for field applications (Hill et al., 1992).

By way of comparison of subjective rating methods, Hill et al. (1992) compared four workload scales in a series of studies, including SWAT, MCH, NASA-TLX and overall workload. The overall workload was literally a subject's rating of overall workload on a unidimensional scale from 0 to 100, with 0 representing very low workload and 100 representing very high workload. In five studies with three U.S. Army systems, the scales were compared with one another in terms of four dimensions: sensitivity (as measured by factor validity), operator acceptance, resource requirements, and special procedures. The

authors found the NASA-TLX to have the greatest correlation with the operator workload factor and identified the technique to have the highest factor validity. Operator acceptance data showed that the NASA-TLX was preferred and was rated as being superior in capability to represent workload. By analyzing resource requirements, the authors also found the NASA-TLX and SWAT to require more time for data reduction and analysis because of their multidimensional nature. However, these two multidimensional measures may also be used diagnostically to address the question of what is causing reported workload, and results may point toward ways to alleviate excessive workload.

Although the above reviewed methods all have limitations, it is well accepted that the NASA-TLX has advantages over other subjective MWL measures, for a number of reasons. First, the TLX provides a multidimensional evaluation of workload, with the potential to identify components of mental demand most affected by an experimental task paradigm. Second, the TLX is more sensitive to mental workload differences than the SWAT, particularly at low workload levels (Nygren, 1991). Hart and Staveland (1988) stated the procedure of the TLX is practically and statistically superior to the SWAT because the independent components of the TLX provide additional diagnostic information unavailable in the SWAT. Finally, and possibly most important, the TLX is much easier to administer than other subjective measurement techniques. Hill et al. (1992) stated the method is more acceptable to participants and increases the likelihood of genuine responses.

The NASA-TLX has also become a tool commonly applied in UAV operation studies. Crescenzio, Miranda, Persiani and Bombardi (2009) compared human use of a "mission replanning" interface under different levels of automation (manual, semi-

autonomous, fully autonomous) in a UAV supervision task. The NASA-TLX was applied as the major workload measure. Results revealed workload under the semi-autonomous mode to have a mean level of the other two cases. In another study, Lu et al. (2013) used the NASA-TLX to compare workload responses in using a situation-augmented display for UAV monitoring task performance. They concluded that the new interface improved performance without increasing operator cognitive workload. Ruff, Narayanan and Draper, (2002) also used the NASA-TLX to identify any changes in workload in a study of human interaction with levels of automation in UAV control. Results revealed workload increases with more UAVs under operator control. In addition, a decision aiding tool was found to reduce operator workload under high demand task scenarios. Given these applications of the NASA TLX, the present study made use of the subjective rating method in combination with physiological measures in order to make assessment of UAV operator cognitive workload in supervisory control interface use.

2.2.4 Model-based workload assessment

This section presents a constrained review of a model-based workload assessment approach. Among methods presented in the literature, the Goals, Operators, Methods and Selection Rules (GOMS) modeling technique has previously been applied for research on HRI (Swangnetr, Kaber, Vorberga, Fleischera, & Thurowa, 2014) and unmanned vehicle interfaces (Kaber, Kim, & Wang, 2011). Due to its potential relevance to UAV interfaces, background on GOMS models and applications are reviewed in this section.

2.2.4.1 Background on GOMS models

The GOMS modeling approach was developed by Card, Newell and Moran (1983). GOMS is a popular, formalized cognitive task performance modeling method that has been commonly used in human-computer interaction (HCI) studies (Drury, Scholtz, & Kieras, 2007) and aviation domain research (Foyle et al., 2005). GOMS facilitates quantification of procedural knowledge necessary to operate a system and allows for quantitative predictions of user performance (Kaber et al., 2011).

GOMS has a few basic components, described as follows: "Goals" represent what a user intends to accomplish with a system and can be hierarchically broken-down into subgoals; "Operators" (hereafter referred to as operations) describing actions performed in service of a goal, which can be perceptual, cognitive or motor in nature; and "Methods" or sequences of operations and subgoal invocations that accomplish a goal. When more than one method is applicable to task performance, "selection rules" are needed to represent a user's knowledge of which method should be applied, given a particular set of circumstances (John & Kieras, 1996).

There are variations on GOMS model coding. NGOMSL (Natural GOMS language; (Kieras, 1994) is a structured natural-language notation in which task methods are represented in a program format or list of steps that contain both external keystroke-level operators and internal cognitive operators (i.e., adding and removing working memory information or setting up subgoals; John & Kieras, 1996). GOMSL (GOMS Language) is an executable form of NGOMSL, structured language notation for developing cognitive task performance models (Kieras, 1999).

2.2.4.2 Application of GOMSL modeling technique

GOMS modeling has been used as a method for representing cognitive task processes (e.g., visual operations, long term memory (LTM) and working memory (WM) use) and for predicting user performance (e.g., task learning and execution times). Kaber et al. (2011) applied computational GOMS models with the objective of understanding human cognitive strategy and behavior in remote rover control. They mounted a laptop computer on a robot for programmatic control of the unit (e.g., local control of sensors and manipulators) by an executable GOMSL model as well as remote wireless control by a human operator. Human rover control performance was compared with cognitive model-based control in a path-following task. For human rover control, a remote workstation was set up in a separate room that allowed for input of direction and speed commands. The operator could view the target rover motion path (captured by an on-board camera) and control the rover's movement through the remote computer control interface. For GOMSL model control, a stepwise control model, representing an exact path-following strategy, and a ballistic-motion control model, representing a minimum task-time strategy, were implemented and tested. The authors compared task completion time and accuracy and found the ballistic model to better approximate human control. The model control performance did not perfectly mimic human performance due to some inherent limitations of the modeling technique (e.g., assumption of serial processing of operations). However, this study identified some utility of developing cognitive models to represent human behavior in HRI tasks. It was suggested that cognitive models could provide flexibility and robustness in modeling human behavior along multiple

performance dimensions and allow for critical insights into operator internal strategies and decision criteria.

Swangnetr et al. (2014) also used GOMSL models as a means for identifying specific types of human perceptual, motor and cognitive operations in HRI for laboratory task performance as well as the flow of human operator information processing. They developed GOMSL models of life science lab technician manual performance of organic sample content analysis. The goal of the modeling effort was to identify high workload methods as part of the process, as a basis for recommending implementation of automation and/or robotics. The GOMSL model output (i.e., motor, cognitive and combinations of both operation counts) was found to significantly correlate with operator subjective ratings of workload measured with the NASA-TLX. It was noted that use of the cognitive task modeling methodology combined with the cognitive workload rating technique provided for a higher resolution analysis of operator demands. The researchers were able to obtain an accurate representation of the cognitive load for technicians and to make precise specification of where automation should be applied in life science lab processes.

The above studies have demonstrated some utility of using GOMSL model outputs to predict cognitive workload in the HRI domain. For example, the peak number of tags present in working memory at any given time can provide a measure of how much information must be recalled by a human operator. Related to this, Kieras (1999) noted that GOMSL model predictions of the need for an operator to maintain more than five tags in WM at once is likely to be indicative of problematic task performance. Beyond this outcome measure, the mere structure of a model (i.e., the stack of goals and subgoals) as well as the number

operations can be used as indicators of level of task complexity and for predictions of operator workload states.

2.3 Summary of literature review

The literature review supports the research interest of the present study in selection of interface and workload evaluation tools. The Modified GEDIS-UAV has advantages over other interface evaluation tools because of its comprehensive nature. The new method takes into consideration both usability and functionality features of UAV interfaces. It also allows for diagnosis of interface design problems and provides guidance for redesign. Beyond this, the revised scoring technique as part of the method provides for greater objectivity in interface evaluation, as compared to the GEDIS and GEDIS-UAV tools. Consequently, the Modified GEDIS-UAV is considered to be a good option for analyzing and classifying interface designs in terms of degree of conformance with existing UAV interface design guidelines as well as general guidelines for human factors consideration in systems design. The method allows for objective quantification of the nature of an interface design, which may ultimately be related to interface user workload and performance outcomes.

The review of literature also provided a basis for selecting workload measurement techniques and metrics for evaluation of UAV control interfaces. A few measures, including cardiac and eye-tracking responses, are considered to be valid indicators of cognitive demands and to have applicability to a typical UAV operation setting, as compared to other physiological responses. Subjective ratings, such as the NASA-TLX, are also considered to have substantial utility for UAV operator workload assessment, based on prior research. Beyond these measures, GOMSL model outcomes are also considered to be useful for

objectively assessing UAV control task complexity and to offer a valid basis for predicting workload.

2.4 Problem statement

The three primary research issues to be addressed by this study include: (1) characterizing any relationship between UAV interface design features and operator cognitive workload responses through use of the new Modified GEDIS-UAV quantitative interface evaluation tool; (2) identifying any complementarity of physiological responses and subjective rating techniques for assessing UAV operator cognitive workload in supervisory control operations; and (3) determining whether the Modified GEDIS-UAV interface evaluation tool has utility for predicting UAV operator workload responses, as represented using a cognitive task performance modeling approach. A lab experiment was conducted to address all of these issues with the use of high-fidelity simulation of instances of a real UAV supervisory control interface. Details of the experiment and results are described in the following sections.

3. Methodology

This section describes the methodology for the present study. As stated above, the primary research interest is to describe the relation between supervisory control interface design features and user cognitive workload in UAV control operations. The study also seeks to determine whether the Modified GEDIS-UAV has utility for predicting UAV operator workload demands as quantified using cognitive task modeling. For this investigation, several variations on an existing UAV supervisory control interface were generated. The interfaces were tested in different tasks scenarios involving typical UAV operations. The operations were used to assess user interactions with interface design features. The study also involved measuring operator cognitive workload in using the different interfaces by applying physiological and subjective response measures.

3.1 Original MP interface

The ArduPilot MP interface was used as a basis for prototyping the various UAV control interface conditions to be tested in this study. The MP interface is a commercially available UAV supervisory control interface that can be integrated with a range of vehicle hardware. The MP interface contains a few functional components, as presented in Figure 3.1 and Figure 3.2. The upper-left section of Figure 3.1 shows a primary flight display (PFD) that presents a simulated horizon line and critical flight parameters. The upper-right section of Figure 3.1 shows a navigation display that presents the map of the operation area, programed waypoints, and projected vehicle route. The lower-left section of Figure 3.1 is a configurable auxiliary box that allows the operator to access flight parameters in a concise digital format. The lower-right section of Figure 3.1 presents a mission-planning tool that allows various

operations on waypoints and overall planning of a mission. Command input in the mission-planning tool needs to be verified by using the waypoint command buttons on the right-hand side of the interface (Figure 3.1).

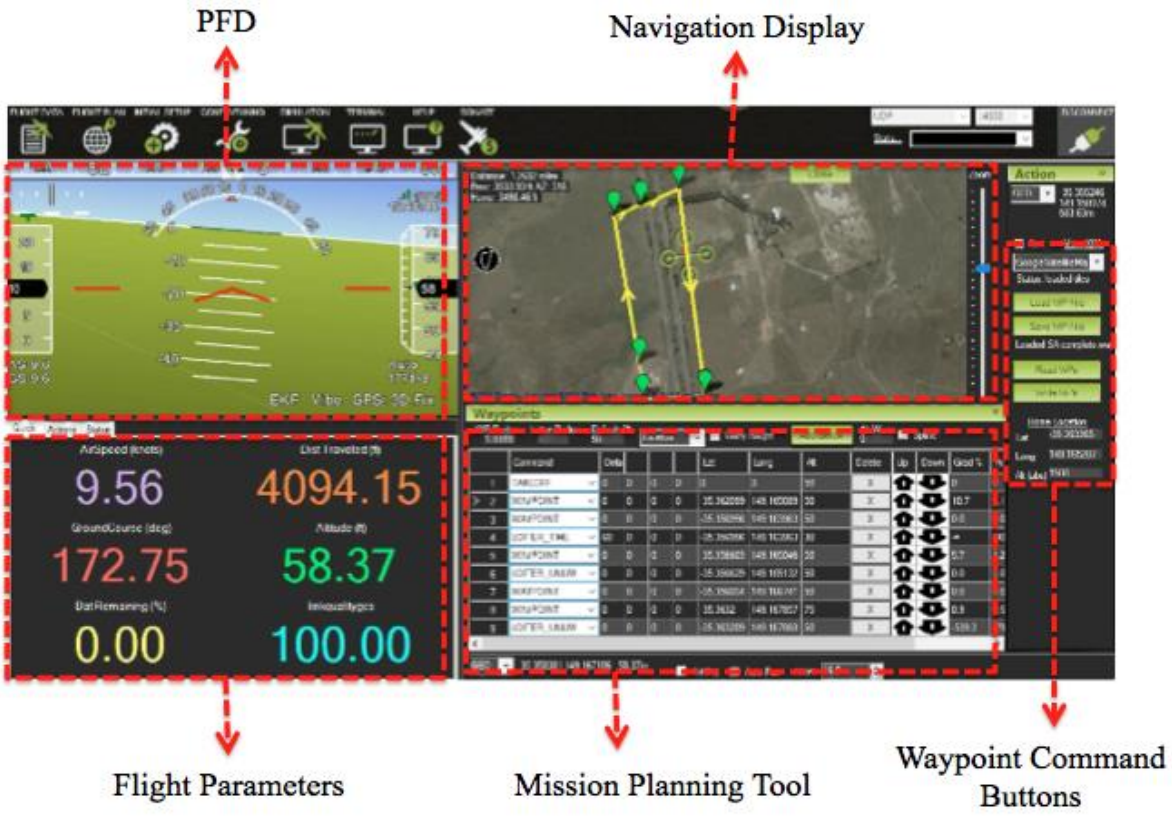

Figure 3.1 Functional Components of MP (Part 1)

There is a configurable auxiliary box in the lower-left of the interface, which also provides the operator with a range of immediate action commands for the UAV, including "Launch UAV", "Return to Launch Point (RTL)", "Loiter", etc. (see Figure 3.2). The auxiliary box enables a user to switch between the parameter display and the immediate action display by toggling option tabs at the top of the box. The interface also has an alarm component. If a user error or abnormal system state is detected during a UAV mission, the

interface will provide warning messages via the PFD or pop windows with error messages (see Figure 3.2).

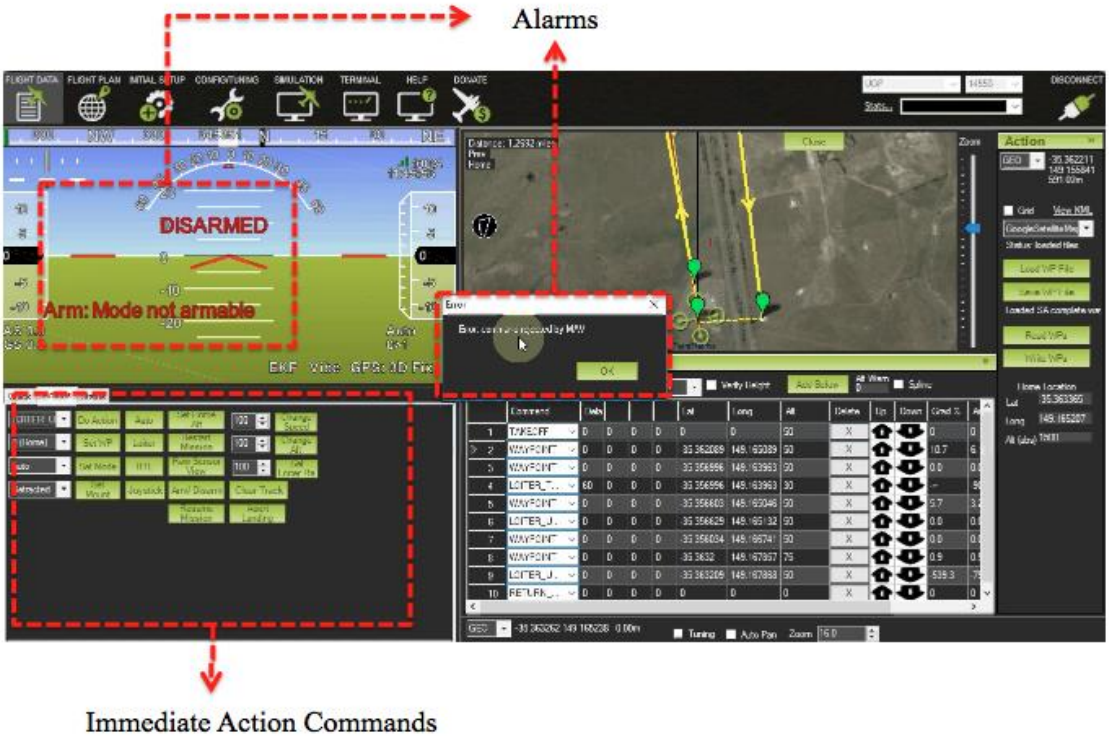

Figure 3.2 Functional Components of MP (Part 2)

3.2 Variations of MP interface

To investigate the relation between supervisory control interface design features and user cognitive workload, a prototyping tool (Just-in-Mind, San Francisco, USA) was used to generate three variations of the MP interface, including an "enhanced" usability/functionality interface, a "baseline" interface and a "degraded" usability/functionality interface. The interface prototypes were highly interactive but did not present the full MP functionality. All the interface variations supported interactions (e.g., mouse clicking, toggling between tabs in the auxiliary box) and provided sufficient visual information presentation (e.g., flight

parameter updating, moving UAV icon) for users to complete vehicle control tasks as part of test scenarios (described in Section 3.3). However, interface features that were not relevant to testing were non-functional.

A Modified GEDIS-UAV evaluation on the original MP interface was used as a basis for proposing interface feature variations for the enhanced and degraded conditions. The "baseline" interface was prototyped to maintain most of the original features of the MP interface. However, some features were modified to allow for discrimination of usability/functionality from the degraded and enhanced interfaces. In addition, some features of the original MP were simplified for all three variations in order to facilitate training for participants. For example, the immediate action commands available in the auxiliary box were reduced and some commands were reworded so that users could easily interpret their meaning without extensive training on the platform. There were a number of vehicle control actions presented in a pop-up menu, which was available by right clicking anywhere on the navigation display. Due to limitations of the prototyping tool, these actions were grouped in a menu that could be accessed via a "Map Action" button on top of the navigation display for all interface variations (see Figures 3.3, Figure 3.4 and 3.5). The enhanced interface attempted to satisfy as many Modified GEDIS-UAV guidelines as possible and, therefore, would produce a higher Modified GEDIS-UAV GEI than the baseline MP interface. The degraded interface was intentionally made less usable.

As an example of one of the interface manipulations, user selection of the Launch UAV command before arming the vehicle triggered an error message in the original MP. Similar to the original MP, the baseline interface prototype indicated through the PFD if the

vehicle was disarmed. If the user attempted to launch a disarmed UAV, an error message informed them that the UAV was not able to launch because it was not armed (see Figure 3.3). The design of this error message was found to pose a few violations of the Modified GEDIS-UAV design characteristic guidelines. In order to address these violations, the error message in the enhanced interface not only informs the user of the cause of the error (i.e., "UAV not armed. Cannot launch."), but also advises appropriate actions to solve the problem (i.e., the message suggests "click Arm/Disarm button" and highlights the "Arm/Disarm" button). The enhanced interface was also designed to prevent user error by greying out the "Launch UAV" button and presenting a reminder message with mouse hover, if the UAV is disarmed (see Figure 3.4). To make the interface less usable in the degraded interface, this error message was removed (see Figure 3.5). Due to the lack of feedback, users needed to rely on their knowledge of the system to figure out the reason for no response/action of the vehicle. Although these "degraded" feature manipulations may not seem reasonable from the standpoint of usable system design, they are representative of some existing commercially available interfaces.

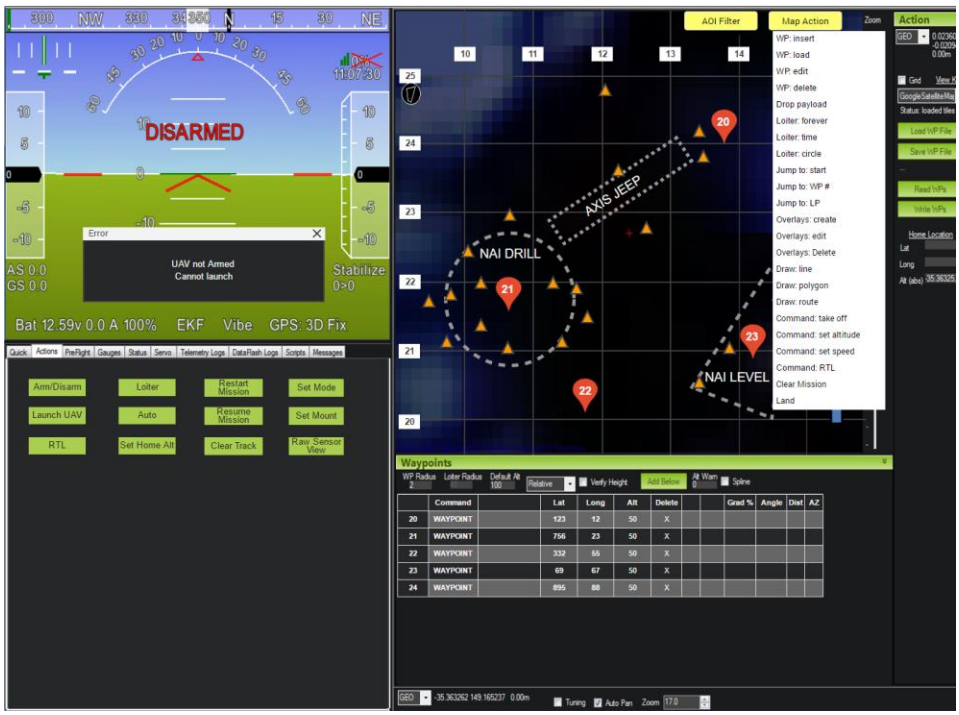

Figure 3.3 Baseline Interface

Figure 3.4 Enhanced Interface

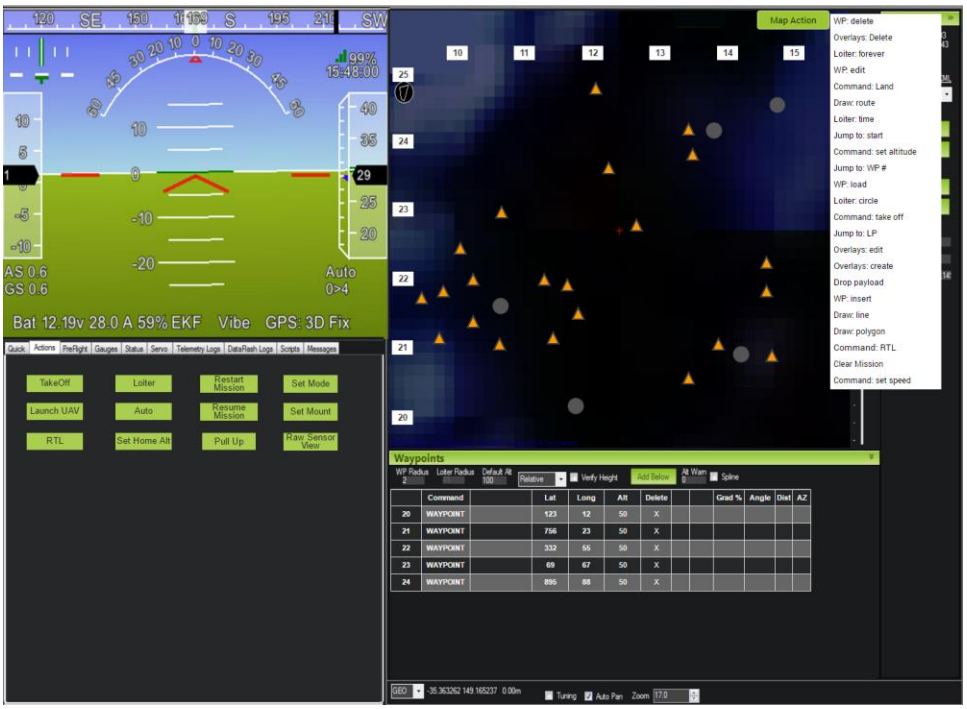

Figure 3.5 Degraded Interface

Following a similar approach, a set of modifications were made for each of the interface variations (see Table 3.1). In general, the goal was to impact as many aspects of the interface design (usability and functional features) as possible in order to cause substantial differences in Modified GEDIS-UAV scores among the prototype interfaces.

Table 3.1 Modifications on Interface Variations

Variation ID	Interface Components	Baseline	Enhanced	Degraded
1	Menu	Menu items are grouped logically but no menu hierarchy.	Menu items presented in hierarchical organization based on logical groups.	All menu items are presented in a long list in random order.
2	Flight parameter display	Acceptable range of value presented with each parameter	Acceptable range of value presented with each parameter; visual reminders (i.e., a small warning icon) when a parameter is out of range.	Only flight parameter displayed. No additional information.
3	Waypoints	Traditional waypoint icons are used. Waypoints are numbered according to flight plan.	Traditional waypoint icons are used. Waypoints are numbered according to flight plan. Waypoint color changes once the UAV flies through.	Waypoint icons are not easy to distinguish from background map features. Waypoints are not numbered.
4	Area of interest (AOI)	A filter presents line coding for AOIs (i.e., dotted margin) and is accessible from a shortcut button.	A filter presents shade coding for AOIs (i.e., shaded area) and is accessible from a shortcut button.	No filter available.
5	Coordinates	Auxiliary lines presented with coordinates.	Auxiliary lines presented with coordinates. "Coordinate" and "Distance" tools are available to facilitate determination of locations and distances.	No auxiliary lines or additional tools available.
6	Alarms	Alarm messages indicate the cause of alarms. The message also provides information on alarm priority.	Alarm messages indicate the cause of alarms and advise actions to resolve the alarms. The message provides information on alarm priority with color coding.	Lack of feedback for some user actions. Alarms only indicate the fault/failure state but not the cause. No information on alarm priority.

Table 3.1 Continued

Variation ID	Interface Components	Baseline	Enhanced	Degraded
7	Confirmation for writing waypoints	Users need to click distant button to confirm waypoint change. Confirmation windows block a portion of the navigation display. Confirmation windows disappear automatically after the process is complete.	Users only need to press "Enter" on keyboard to confirm waypoint change. Confirmation windows do not obstruct any portion of the navigation display. Confirmation windows disappear automatically after process is complete.	User needs to click a distant button to confirm waypoint change. Confirmation windows block a portion of the navigation display. User needs to manually close the window when the process is complete.

The features of the enhanced interface were expected to reduce user errors and mitigate workload by reducing operation complexity and working memory load during UAV missions. The degraded interface, however, was expected to increase task complexity and require users to recall more information from prior training for the mission. It is important to note that the interface manipulations were not intended to achieve an "optimal" design or to resolve shortcomings of the existing MP interface from a usability standpoint. Rather, the motivation was to generate interface conditions that would result in a range of Modified GEDIS-UAV scores and that might also lead to differences in operator workload in simulated UAV control operations. The interface conditions were also expected to allow for determination of whether Modified GEDIS-UAV scores were predictive of UAV operator cognitive workload outcomes.

3.3 Mission Scenario

Based on the proposed interface conditions, two mission scenarios were developed in order to observe operator usage of the various enhanced, control and degraded interface features in

fundamental vehicle control operations. Through a preliminary study of UAV control operations, fundamental control tasks were identified that made use of the range of functional components of the MP interface. The list of generic tasks and a mapping to related interface features is presented in Table 3.2.

Table 3.2 Mapping of Generic Tasks to Interface Features

Generic Tasks	Interface Components
Launch UAV	Immediate Action Commands
Monitor route	Navigation Display
Monitor vehicle status (e.g., ground speed)	Flight Parameter Display
Drop Payload	Navigation Display (“Map Action” Menu)
Identify hazards	Alarms
Change Altitude	Mission Planning tool
Modify Route (e.g., return to launch point)	Immediate Action Commands
Flight path tracking	Flight Parameter Display

The task list provided a basis for generating mission scenarios for the experiment. A subset of the generic control tasks was selected and included in the scenarios. The scenarios assumed the operator was to fly a simulation of a quad-rotor UAV prior to an actual mission. The operator was asked to perform a series of control tasks following a scheme of maneuver, including launching the UAV, flying through a few pre-programmed waypoints in numerical order, determining a few environmental features (i.e., number of targets) and eventually returning to the launch point. At certain points of the mission, the operator was asked to drop a payload, which could be completed through the “Map Action” menu or a shortcut button. The operator was also asked to change the altitude settings for two waypoints. However, the mission scenarios did not include loiter (i.e., dwell time on a target). Vehicle speed was used to control event rates in test trials. During a mission, a pre-recorded audio message was used

to provide verbal cues to alert operators of the need for control actions (e.g., “Fix the current alarm”; also see Table 3.6 for two complete lists of verbal cues). In addition, operators were required to closely monitor flight parameters and the UAV navigation display. The operator needed to report flight parameter anomalies (i.e., values beyond acceptable ranges) or targets of interest during the mission. This requirement was intended to engage participants in the test scenarios. Their responses were not used as additional bases for performance evaluation. All test scenarios included the same control tasks but followed a slightly different timelines for the scheme of maneuver. Two mission maps were generated to allow replication of test conditions. The geographical features (i.e., locations of area of interest (AOIs)) presented in the mission map (see Figure 3.6) were made slightly different in order to prevent carryover learning effects among replications. The AOIs were used as a basis for identifying targets of interest. The targets were not presented in the mission maps that were provided to operators. Instead, they were presented at arbitrary locations in the navigation display as part of the prototype interfaces in order to simulate unpredictable environmental features.

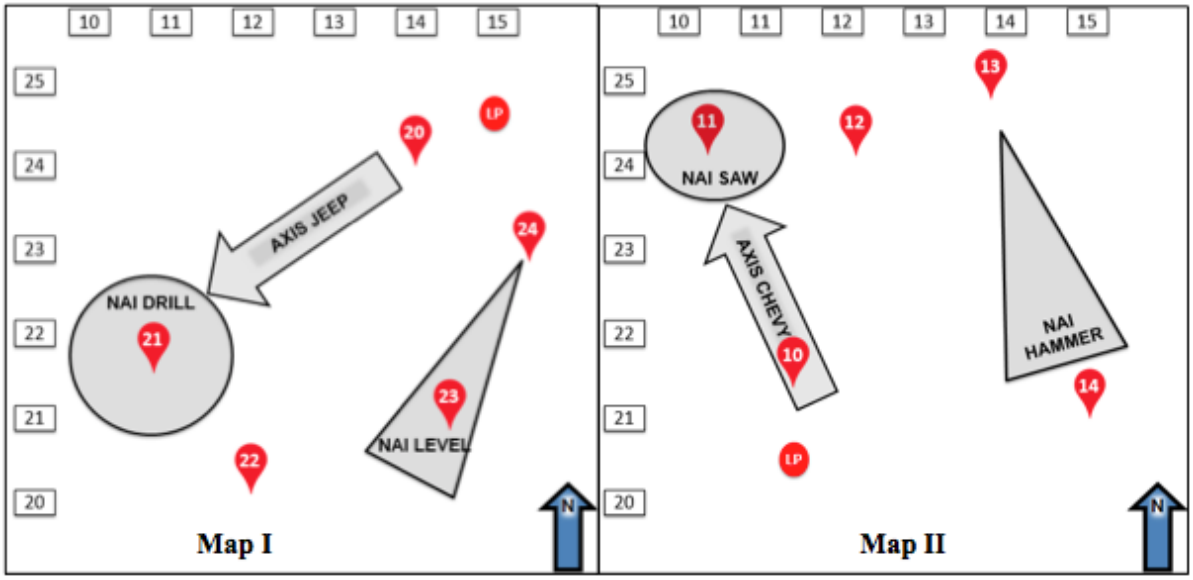

Figure 3.6 Mission Maps for Test Scenarios (Numbers in small rectangles represent eastings and northings; numbered round icons represent waypoints and the LP; grey shapes represent AOIs for a mission)

Beyond these test scenario designs, a UAV control training scenario was also developed. The training scenario was a reduced version of the test scenarios with different geographical features (see Figure 3.7). However, the operator was required to perform the same control tasks as in the test scenarios. The training scenario was used to familiarize participants with the use of each of the three interface prototypes, including control structures, while potentially avoiding carryover effects of participant scenario learning to test trials. Consequently, any differences in participant workload responses observed during the experiment results were attributed to the characteristics of the interface designs vs. familiarity with the UAV control simulation.

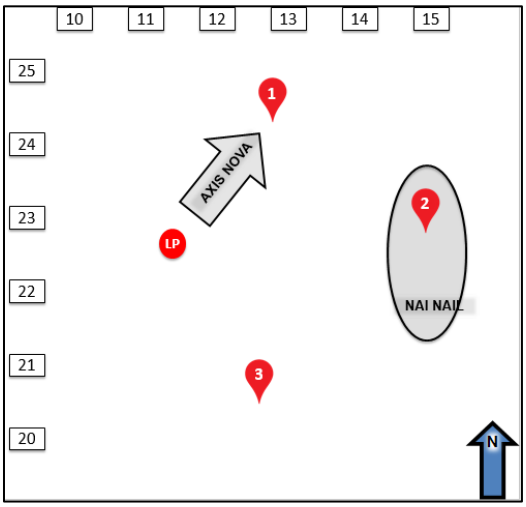

Figure 3.7 Mission Map for Training Scenario

3.4 Interface usability and task complexity verification

All interface variations were evaluated using the Modified GEDIS-UAV tool. The interface prototypes were used as a basis for the evaluation. The components of the MP interface (as well as the variations) can be mapped to the indicators in the Modified GEDIS-UAV (see Table 3.3). The features of each component were compared with the guidelines for the

related functionality indicator of the Modified GEDIS-UAV as well as all the usability indicators.

Table 3.3 Mapping of MP Interface Components to Modified GEDIS-UAV Indicators

Modified GEDIS-UAV Indicators	Interface Components
Data Entry and Commands	Immediate Action Commands Mission Planning tool
Map and Navigation	Navigation Display
Status and Devices	Flight Parameter Display
Alarm	Alarms

To improve the reliability of application of the interface evaluation tool, three human factors experts identified relevant design criteria through a group effort. Subsequently, the three analysts evaluated the interface designs, independently. The averages of their evaluation scores were calculated as assessment results for the interfaces. As expected, the enhanced interface had the highest average Modified GEDIS-UAV score (GEI=0.86), followed by the baseline interface (GEI=0.78) and then the degraded interface (GEI = 0.67). The GEI values reflected the capability of each variation to support effective and efficient UAV operations.

In addition to use of the Modified GEDIS-UAV tool, heuristic analysis was used as an additional form of interface usability assessment in order to confirm/validate potential differences in the Modified GEDIS-UAV scores for the interface variations. Each of the interface variations were compared with the usability principles for improving HCI (Molich & Nielsen, 1990). The interface variations were judged by a small sample of human factors experts as to whether features satisfied, partially satisfied or violated the usability principles

(see Table 3.4). As expected, the enhanced interface satisfied all the principles and the degraded interface posed the most violations.

Table 3.4 Usability Heuristic Analysis on Interface Variations (with "Satisfied"=100%, "Partially Satisfied"=50% and "Not Satisfied"=0% for each heuristic)

Usability Principles	Enhanced interface	Baseline interface	Degraded interface
Simple and Natural Dialogue	Satisfied	Satisfied	Not satisfied
Speak the User's Language	Satisfied	Satisfied	Partially satisfied
Minimize the User's Memory Load	Satisfied	Partially satisfied	Not satisfied
Be Consistent	Satisfied	Partially satisfied	Not satisfied
Provide Feedback	Satisfied	Satisfied	Partially satisfied
Provide Clearly Marked Exits	Satisfied	Satisfied	Satisfied
Provide Shortcuts	Satisfied	Not satisfied	Not satisfied
Provide Good Error Messages	Satisfied	Partially satisfied	Not satisfied
Error Prevention	Satisfied	Partially satisfied	Not satisfied
Total Percent Satisfaction (across heuristics)	100%	~67%	~22%

Beyond this, GOMSL models were developed for each task as part of the identified task scenarios that were completed with the three different MP interface variations. For each model, the number of perceptual (P), motor (M), cognitive (Cog) operations and the number of working memory (WM) chunks required for each individual task were counted (Table 3.5). These model outcomes allowed for task demand and operator workload quantification as well as task performance projections. In general, the enhanced interface required fewer steps and lower WM load for performing the UAV tasks, as compared to the baseline and degraded interface, indicating less task complexity and lower cognitive workload. From a task perspective, "monitoring vehicle status" was found to be the least cognitively

demanding task with reporting coordinates or distances and “prioritizing alarms” posing the highest WM and cognitive demands.

Table 3.5 GOMSL Models Output

	Enhanced				Baseline				Degraded			
Tasks	**P**	**M**	**Cog**	**WM**	**P**	**M**	**Cog**	**WM**	**P**	**M**	**Cog**	**WM**
Launch UAV	4	5	17	3	4	5	17	3	3	5	18	3
Report coordinate	5	6	34	5	7	4	42	6	9	2	43	7
Report distance	5	7	30	6	4	1	30	7	5	1	31	7
Change altitude	3	12	16	5	7	8	24	5	7	9	26	5
Drop payload	2	2	14	2	4	6	22	3	4	6	22	3
Fix alarm	3	2	13	3	4	2	16	4	5	2	23	6
Prioritize alarm	8	11	28	3	11	11	40	3	17	11	46	6
Return to launch	3	4	17	3	3	4	17	3	3	4	17	3
Monitor vehicle status	1	1	5	1	2	1	10	2	2	1	13	3

3.5 Experiment

3.5.1 Participants

Forty-nine (49) participants (26 male, 23 females; age: 24.8± 3.9) were recruited from NC State University campus population and surrounding community. After the initial data collection with 37 participants, a power calculation was performed with the “Sample Size and Power” function in JMP. Observed differences among condition means and variances were used as input for the calculation. An Alpha level of 0.5 and desired power of 0.8 was submitted to JMP in order to identify the number of observations required to reach statistical significance based on the initial test data. This process resulted in the need for 12 additional participants. Participant age was restricted to 18-40 years due to the fact that the lens of the eye thickens after 40 years of age and significantly affects eye activity, specifically a slowing

of focus in shifting attention (Bruce, Atchison, & Bhoola, 1995). In order to ensure the validity of data collection, a few inclusion criteria were applied, including: (1) participants being required to have 20/20 or corrected vision and no color-impairment due to the study requirement for visual perception of colored items/text on a display screen; (2) participants being required to have general familiarity with use of computers, as the study involved using an interface simulation presented on a desktop/laptop computer; and (3) participants being required to have no experience with UAV supervisory control operations so as not to bias performance in the simulated control tasks. Participants were not allowed to wear contact lenses as lenses have been found to directly influence a person's blink (Eckstein, Guerra-Carrillo, Miller Singley, & Bunge, 2016), which was considered to be an important response indicator of workload in this experiment. However, participants were permitted to wear glasses for vision correction. Participants were also asked to complete a demographic survey regarding their age, gender, daily computer usage, experience with gaming, and flight (or flight simulations). This information was used as a basis for covariate analyses in assessing any effects of the interface manipulations on operator cognitive workload responses.

3.5.2 Apparatus

The experiment setup (see Figure 3.8) included a desktop computer for presenting the interface simulations as well as two sheets of paper for presenting mission information. A keyboard and mouse were provided for participant interaction with the UAV control simulation. Participants sat in front of the computer monitor during the training and test trials. An eye tracking system (Seeing Machines, Austrlia) was mounted under the computer monitor and in front of the participant in order to capture pupillography measures (also see

Figure 3.8). The Facelab system consists of two cameras and an infrared light emitting pod. When reflected on the eyes, the light emitted from the pod is captured by the cameras along with the outline of the pupil. Eye movements were captured at a frequency of 60 Hz with an accuracy of 0.5° to 1° of rotational error. In gaze tracking, the two points of reflection on the surface of the eye (the (infrared light) glint and edge of the pupil) allow for determination of gaze direction. The Facelab system registered blinks in the output file, which allowed calculation for blink duration during the recording, as a basis for identifying fluctuations in cognitive load.

Beyond this equipment, participants were required to don a Polar HR monitor (V800 model and H7 monitor, Polar Electro Inc., Kempele, Finland). The system was used to capture HR and HRV response measures. The HR monitor requires participants to wear a chest band and a watch during experiment trials. Neither component was obstructive of participant interaction with the simulation workstation.

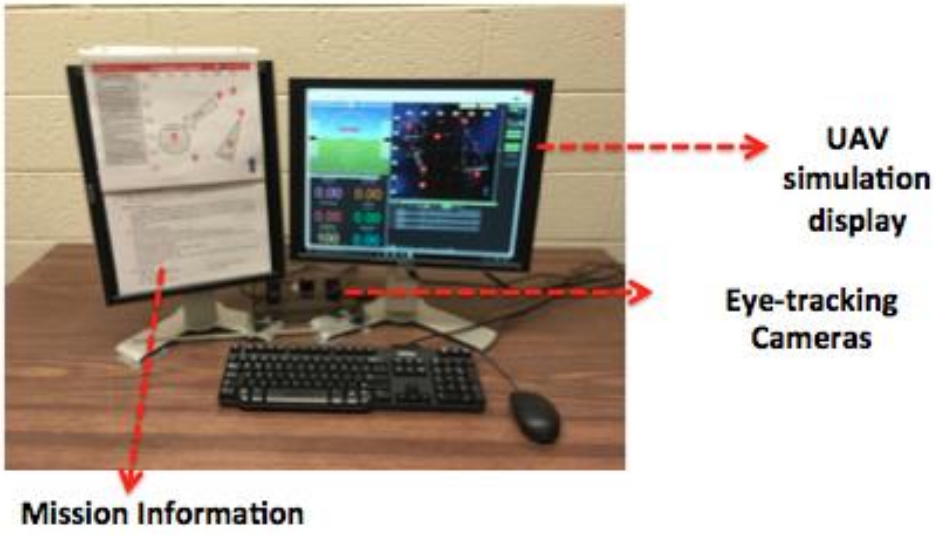

Figure 3.8 Experiment Setup

3.5.3 Independent variables

This study manipulated two independent variables (IVs), including the UAV control interface variation (V) and the simulated vehicle ground speed (S). The interface variation (V) had three levels, namely the enhanced, control and degraded interface conditions. In addition, there were two levels of vehicle ground speed (S), slow and fast, applied in presenting the test scenarios. These vehicle speeds can be translated to two levels of task event rate, also including slow and fast. As shown in Table 3.6, the number of events presented in the mission remained fixed between the test conditions and the change in event rates was simply a reflection of the vehicle speed manipulation. As a result, the trial lengths associated with the high speed and low speed conditions were 5.5 and 8.5 minutes, respectively. These trial durations were elected based on operator performance and perception of workload during a pilot study as well as the need for sufficient sampling of HRV for data analysis (at least 5 min as discussed in Section 2.2.2.1). The vehicle speeds were held constant during each test scenario with the intent of delivering a pure task demand manipulation. The fast speed or event rate was expected to introduce time pressure to participants and, therefore, make a scenario more demanding. This manipulation was also expected to allow for assessment of the extent to which degraded interface design might magnify high task demands and the extent to which enhanced interface design might serve to mitigate workload in a high demand scenario. Although the focus of this investigation was on how UAV supervisory control interface design features may translate to operator workload, there was the possibility of some evidence of specific feature designs allowing operators to better manage task demand increases.

Table 3.6 Example Event Timelines for Different Vehicle Speed Conditions

Fast		Slow	
Time (m:ss)	**Verbal Cue**	**Time (m:ss)**	**Verbal Cue**
0:00	Launch when ready.	0:00	Launch when ready.
0:13	Report the coordinate of the Western most target outside of NAI Drill.	0:21	Report the coordinate of the Southern most target in NAI Level
0:26	What Northing is Waypoint 20 closest to?	0:41	What Easting is Waypoint 22 closest to?
0:39	Change WP 21 altitude to be 35 feet	1:03	Change WP 21 altitude to be 35 feet
1:05	What is your current Battery Remaining?	1:23	In knots, what is your current speed?
1:18	What is the name of the Axis?	2:04	What is your current completion percentage for this mission?
1:40	Report the distance from WP 24 to the Launch Point.	2:39	Report the distance from Waypoint 22 to Waypoint 23.
1:57	For what Map Action command is "Land" an option?	3:06	In Map Action, what are 2 of the 3 Overlay actions available?
2:10	What map action is Speed found under?	3:26	Prioritize the alarms per your training.
2:29	Drop payload	3:58	Drop payload
2:31	Change WP 23 altitude to be 60 feet	4:00	Change WP 23 altitude to be 60 feet
2:49	In map actions how many loiter actions are available?	4:29	What is the normal range for UAV altitude?
3:07	How many targets are in NAI Level?	4:49	What is your total distance traveled for this mission so far?
3:28	Fix the alarm shown.	5:10	Fix the alarm shown.
3:41	What is the target distance for the UAV?	5:51	In map actions, what are the waypoint options available?
3:54	What is your current altitude?	6:11	In Map Action, where can you jump to?
4:27	Prioritize alarms per your training.	6:53	How many targets are inside NAI Drill?
4:46	What is the normal range for ground speed?	7:14	What is your current ground course?
4:53	Return to launch.	7:44	Return to launch
5:05	What is the minimum acceptable value for Battery Remaining?	8:05	In meters/second, what is your current speed?
5:12	End of mission (UAV reached launch point)	8:15	End of mission (UAV reached launch point)

3.5.4 Experiment design

The experiment followed a mixed factor design. The three interface variations served as a between-subject factor. The two levels of vehicle speed (each with one replication) served as a within-subject factor. In order to determine if the interface variations resulted in different cognitive workload responses, a within-subject design would have been effective for minimizing the effect of individual differences on experiment results. However, pilot testing revealed that delivering multiple interfaces to one participant could introduce confusion among interface features and learning effects. Therefore, the mixed factor design was selected. Each participant was randomly assigned to an interface variation and completed two mission scenarios under the two vehicle speed/event rate settings. This implies a Split Plot Design (SPD) as a basis for data analysis, in which the whole plot has a Completely Randomized Design (CRD) and the split plot has a Randomized Complete Block Design (RCBD). In this design, the interface variations and the vehicle speed were considered as whole-plot and split-plot factors, respectively. The mission maps were considered as replications. Geographical features presented in the maps (or navigation display) were different in order to prevent potential participant learning effects. However, the vehicle control tasks to be executed remained the same. Example treatments for participants are summarized in Table 3.6.

Table 3.7 Experimental Treatments

Participant #1	**Participant #2**	**Participant #3**	**…**
Enhanced Interface	Baseline Interface	Degraded Interface	…
Map I - Slow	Map I - Slow	Map I - Slow	…
Map II - Slow	Map II - Slow	Map II - Slow	
Map I - Fast	Map I - Fast	Map I - Fast	
Map II - Fast	Map II - Fast	Map II - Fast	

3.5.5 Procedure

Prior to the experiment, potential participants were screened according to the inclusion criteria. Only eligible participants were scheduled to visit the NC State Ergonomics Lab for testing. Upon arrival at the lab, participants were given a brief introduction to the study and were provided with an informed consent form and a demographic questionnaire. If a participant consented to the experiment procedures and potential risks (minimal), an experimenter calibrated the eye-tracking system for the participant.

Subsequently, participants were provided with familiarization on the assigned interface. Participants were introduced to the specific interface features and functionalities. The goal of this familiarization session was to ensure that participants understood the capability of the interface prototypes and critical interactions/commands to complete the control tasks in a mission scenario. Subsequently, participants were familiarized with the mission scenario. A list of events as part of the "scheme of maneuver" and a mission map sheet were provided to participants (as described in Section 3.3). An experimenter then trained the participant to recognize map features (e.g., AOIs, targets, coordinates and map orientation) and provided a brief introduction to the mission tasks. The familiarization session lasted approximately 30 min in total. Once a participant was comfortable with the

interface and the mission, (s)he was presented with an interactive prototype for the training scenario. Any vehicle control errors were noted and identified for the participant to help them develop proficiency in the operation. Participants were required to repeat the training scenario until they could complete all tasks without any errors. At the close of the training session, participants were required to complete the NASA-TLX demand component ranking form. They were required to make the 15 pairwise comparisons of the six different demand components for each interface type.

Prior to the test trials, an experimenter assisted the participant in donning the HR monitor. The various combinations of scenario and vehicle speed/event rates were delivered through the test trials in randomized order. The test trials lasted 8.5 min under the slow vehicle speed and about 5.5 min with the fast vehicle speed. These test durations were selected to ensure stationarity and reliability of the HRV signals (Jorna, 1992). After each test trial, a participant was asked to rate their perceived cognitive workload for each demand component by using the NASA-TLX survey. The participant was also asked to rate the interface using the MCH-UVD, which is a simpler usability evaluation tool (reviewed in Section 2.1.1). MCH-UVD ratings were collected for comparison with the Modified GEDIS-UAV scores and comparison of the sensitivity of the two tools for UAV interface evaluation. To prevent eyestrain and mental fatigue from continuous task performance, 2-min rest periods were provided between each of two experiment trials. Upon completion of the experimental trials, participants were asked to complete a usability questionnaire regarding their general impression of the interface. Subsequently, participants were instructed to stay relaxed and look at any location within the range of the eye-tracking cameras for about 5

min. Meanwhile, baselines of participant cardiac and pupillometry responses were recorded as bases for subsequent test trial data normalization. Table 3.7 presents an example experiment procedure for a participant.

Table 3.8 Experiment Procedure

Example Procedure			
Pre-experiment screening			
Informed consent form			
Orientation			
Equipment calibration			
Familiarization and training			
NASA-TLX demand component ranking			
Trial 1: Scenario I - Slow	NASA-TLX rating	MCH-UVD	Rest
Trial 2: Scenario I - Fast	NASA-TLX rating	MCH-UVD	Rest
Trial 3: Scenario II - Slow	NASA-TLX rating	MCH-UVD	Rest
Trial 4: Scenario II - Fast	NASA-TLX rating	MCH-UVD	Rest
Usability questionnaire			
Baseline measurement for eye-tracking and cardiac responses			
Debriefing and payment			

3.5.6 Dependent variables

Cognitive workload was measured in this study through a number of responses. The first DV for indicating participant cognitive workload was blink duration (BD). Previous research (Kramer, 1990) identified blink duration as being sensitive to task demand changes with longer blink durations indicating lower cognitive workload. The FaceLab system registered participant blink activity during test trials at a rate of 60 Hz. The system output was used as a basis for determining participant average blink duration during each trial. As another common eye-tracking measure identified in the literature, change in pupil size has been found to be sensitive to task demand fluctuations (Palinko et al., 2010). However, this

measure was not suitable for the present study due to inconsistent illuminance levels within the visual field. The mission materials were presented using white paper (see Figure 3.8) and, therefore, had a high illuminance level (~78 Lux). However, the interface presented via the monitor had a dark background and, consequently, a lower illuminance level (~30-50 Lux). When participants shifted their gaze between the interface and mission materials, any change in pupil size due to cognitive demand fluctuations was confounded with the pupil's adaption to the different illuminance levels.

The second and third DVs collected during the experiment were the cardiac responses, including HR and HRV (i.e., HF power). As previously discussed, both measures have been found to indicate cognitive task demands (Jorna, 1992; Wierwille & Eggemeier, 1993) with higher HR indicating higher cognitive workload. Spectral analysis was suggested for short time recordings of HRV (Jorna, 1992) and a decrease in HF power is indicative of higher workload (as discussed in Section 2.2.2). Polar HR monitor captured HR and beat-to-beat intervals of heart rhythm (also known as RR interval) of the participants. The RR interval data was processed with the Kubios HRV software (Tarvainen, Niskanen, Lipponen, Ranta-aho, & Karjalainen, 2014). This package supports spectrum estimatation with Welch's periodogram and divides estimates into VLF, LF and HF bands according to the recommendations of the Task Force of the European Society of Cardiology (1996). It also computes commonly used time-domain and frequency-domain HRV parameters, including mean HR and HF power. The ambient temperature of the test room was well controlled during all experiments (~70-73 ℉) and was not found to pose any influence on participant cardiac responses in the pilot study.

Normalization of physiological responses is necessary to account for individual differences. Physiological responses during rest (i.e., relaxed states) were recorded for each participant as a basis for data normalization. However, some participants showed lower cognitive load (i.e., longer BD, lower HR, and higher HF HRV) during test trials. Therefore, the lowest observations of cognitive demand state, either from rest or the test trials, was used to normalize the physiological responses with the formulas presented in Table 3.8.

Table 3.9 Normalization of Physiological Responses

DVs	Normalization Formulas	Indication of Higher Workload
DV1: Blink duration (BD)	BD decrease ratio = $\frac{Max\ BD - Trial\ BD}{Max\ BD}$	Higher ratio value
DV2: Heart Rate (HR)	HR increase ratio = $\frac{Trial\ HR - Min\ HR}{Max\ HR}$	Higher ratio value
DV3: HF power	HF decrease ratio = $\frac{Max\ HF - Trial\ HF}{Max\ HF}$	Higher ratio value

The fourth DV for assessing cognitive workload was the NASA-TLX subjective demand rating. The overall TLX value was computed for each test trial and use of a specific interface design, with higher values indicating higher workload.

In addition to the workload response measures, participants were also tested on their dynamic knowledge of UAV state and operation status during the test trials. These queries (Table 3.10) were delivered in between verbal cues of vehicle control tasks (see Table 3.6) using the same audio playback device. The accuracy of participant responses to the dynamic knowledge queries (DKQ) served as a fifth DV to assess the utility of the various interface designs for supporting UAV control operations.

Table 3.10 Example Queries for Testing Dynamic Knowledge

Example DKQs
What is the name of the [flight] Axis?
What is the normal range for UAV altitude?
In Map Action, how many "Delete" options are available?
What Northing is Waypoint 12 closest to?
What is your current ground speed?
What is your current completion percentage for this mission?

3.5.7 Research Hypotheses

The specific hypotheses tested as part of this study pertained to the expected workload effects of the various independent variables. In general, the interface design variations were expected to influence participant cognitive workload responses during the simulated UAV operations. More specifically, the enhanced supervisory control interface design was expected to generate the lowest cognitive workload responses (physiological and subjective) for both vehicle speeds/task event rates, followed by the baseline interface, and then the degraded interface. In terms of the specific response measures to be collected, it was hypothesized (H) that the enhanced interface would result in the lowest BD decrease ratio (H1), lowest HR increase ratio (H2), lowest HF power decrease ratio (H3), lowest NASA-TLX scores (H4), and highest DKQ accuracy (H5), followed by the baseline interface and then the degraded interface.

Beyond these hypotheses, the enhanced interface was also expected to be more robust in terms of moderating user workload and supporting performance under highly demanding task circumstances. More specifically, under the fast vehicle speed/event rate condition, the enhanced interface was expected to moderate operator workload, as compared to the baseline

and degraded interfaces, greater than under the slow vehicle speed/event rate condition. The interaction of the "Interface Variation * Vehicle Speed" was expected be significant in effect on the workload response measures with any difference in BD decrease (H6), HR increase (H7), HF decrease (H8), TLX ratings (H9) and DKQ accuracy (H10) being greater among interfaces used at the higher vehicle speed.

3.6 Data analysis

3.6.1 Data cleaning

The data for one participant was excluded from analysis due to unexpected interruptions during the test period. In addition, cardiac data for two other participants and two test trials by a third participant were excluded from analysis due to signal corruption, as identified through visual inspection of the data. (Variations in statistical test degrees of freedom (DOFs) occurred due to these data reduction steps.) Related to the BD response, previous study (Zhang, White, et al., 2016) suggested only trials with more than 80% valid observations should be used for further analysis. For this reason, three data points were removed from the present study dataset due to excessive loss of observations during test trials.

3.6.2 Variance analytics

Analysis of Variance (ANOVA) was applied to all response measures collected during the experiment. Prior to application of the parametric procedure, diagnostics on response measure data were performed in order to determine if the assumptions of the parametric procedure were met, including constant variance of responses among settings of the IVs and normality of response residuals across all experimental conditions. Barlett's test (Snedecor &

Cochran, 1989) and Shapiro-Wilk's test (Shapiro & Wilk, 1965) were used to assess data conformance with the homoscedasticity and normality assumptions, respectively. The equal variance and residual normality assumptions were violated for all DVs. Transformations (e.g., log or square root transformations) were ineffective for ensuring parametric assumptions. Consequently, ranked observations were submitted to the parametric test procedure (ANOVA). However, in cases where the nonparametric results were similar to results of ANOVAs on the raw response measures, analyses performed on the untransformed responses were considered valid (Montgomery, 2017) and reported.

In order to assess the workload and performance effects of the UAV control interface (I_i) and vehicle speeds (S_k; event rates) as well as their interaction (IS_{ik}), a statistical model was formulated and subjected to the ANOVA procedure. As shown in Equation 4, y_{ijklm} represents the dependent variable for analysis; $P(I)_{j(i)}$ represents the whole plot error (i.e., the participant nested within the interface) and was used to test the significance of interface variation; $SP(I)_{kj(i)}$ represents the split plot error (i.e., the speed condition crossed with participants as nested within the interface condition) and was used to test the significance of S_k and VS_{ik}. As shown in Table 3.11, the DOFs of the error terms varied from one DV to another due to data for certain participants being removed from the analysis of some response measures.

Table 3.11 DOFs of Error Terms

	Number of Participants	Number of Observations	DOF of Whole Plot Error	DOF of Split Plot Error
BD decrease	48	188	3*(16-1) = 45	1* 45 = 45
HR increase	45	177	3*(15-1) = 42	1 * 42 = 42
HF decrease	45	177	3*(15-1) = 42	1 * 42 = 42
TLX ratings	48	191	3*(16-1) = 45	1* 45 = 45
DKQ accuracy	48	189	3*(16-1) = 45	1* 45 = 45

Trial number (T_l) was included in the statistical model to represent potential learning during experimental trials. The trial number effect was also tested using the split plot error since it is a covariate at the split plot level.

$$y_{ijklm} = \mu + V_i + P(V)_{j(i)} + S_k + VS_{ik} + SP(V)_{kj(i)} + T_l + \varepsilon_{ijklm}$$ (Equation 4)

This statistical model was applied to each dependent variable in order to identify significant main effects and whether the interaction was influential in responses. The ratings of MCH-UVD were also analyzed with this model. The goal of this additional analysis was to identify the sensitivity of Modified GEDIS-UAV scores to interface manipulations, as compared to the MCH-UVD ratings. In addition to ANOVAs, post-hoc analyses were conducted on all response measures using Tukey's Honestly Significant Difference (HSD) procedure to identify specific differences among the interface conditions as well as the settings of the two-way interaction.

3.6.3 Regression analysis

This additional analysis focused on assessing the utility of particular cognitive modeling parameters and the application of the Modified GEDIS-UAV for predicting subjective workload outcomes. Due to the continuous nature of the GOMSL model parameters and the

Modified GEDIS-UAV scores, regression modeling was applied. With respect to the model predictors, GOMSL parameters described the complexity of task performance in using the interfaces. However, task complexity was dependent on the interface design, which was characterized by the Modified GEDIS-UAV scores. For this reason, the model predictors were found to be highly correlated with each other. Due to multi-collinearity among predictors, multi-parameter models were abandoned, and a set of single parameter regression models were developed and compared, as alternatives.

3.6.4 Correlation analysis

If any physiological workload response measures revealed sensitivity to the influence of the interface design variations, correlation analyses were conducted with the subjective workload ratings and performance measures. These analyses were considered as a form of cross-validation of the physiological responses for indicating cognitive load during the UAV control operations. Spearman's rank correlation coefficients were used to assess monotonic relationships between the responses (Higgins, 2003) due to violation of the normality assumption for multiple response measures.

4. Results

A significance level of α=0.05 was used as a criterion for establishing statistical significance of all ANOVA and post-hoc test results. All plots of condition response measures include untransformed mean values appearing at the top of bars. All physiological responses were normalized and, therefore, could span a range from 0 to 1. However, based on the actual range of observed responses, and to allow for ease of visual comparison of conditions, the anchors of the vertical axes in response plots was adjusted to 0 to 0.5. In addition, all error bars represent +/- 1 standard deviations from the mean. Tukey's HSD significance groupings were represented by letters (e.g., "A", "B") on the graphs. Conditions connected by the same letter are not significantly different from each other. Covariates of gender, age, computer usage, gaming experience and flight simulator experience did not have significant effects on any of the response measures and were removed from all models.

4.1 BD decrease ratio

An ANOVA on the BD decrease ratio revealed a significant effect of interface variation ($F(2,45) = 3.84$, $p = 0.0289$). Tukey's HSD post-hoc tests revealed the BD decrease ratio with the enhanced interface to be significantly lower than with both the baseline and degraded interfaces, indicating lower cognitive workload. In general, an increasing trend of BD decrease ratio was observed as the interface become less usable, as shown in Figure 4.1. The results revealed no significant effect of vehicle speed ($F(1, 45) = 2.97$, $p = 0.0916$) However, descriptive statistics showed a higher BD decrease ratio under the high vehicle speed condition (0.127) as compared to the low speed condition (0.101). The trial effect

(F(1,45) = 3.70, p = 0.0607) or interaction effect of the interface variation and vehicle speed (F(2,45) = 0.10, p = 0.9031) were not significant.

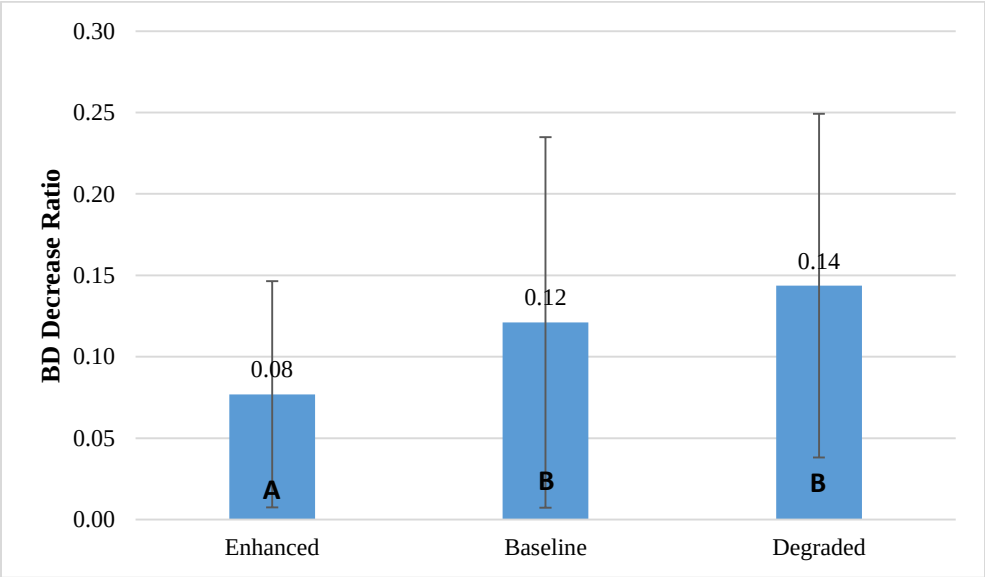

Figure 4.1 Interface Variation Effect on BD Decrease Ratio

4.2 HR increase ratio

Although the enhanced interface achieved a slightly lower HR increase ratio (Table 4.1), the ANOVA results on HR increase ratio did not reveal a significant effect of interface design variation (F(2, 42) = 2.35, p = 0.1195), vehicle speed (F(1, 42) = 3.52, p = 0.0676) or their interaction (F(2, 42) = 0.35, p = 0.7028). The effect of trial number (F(1,42) = 30.65, p < .0001) was found to be significant with a decreasing trend towards later trials (Figure 4.2).

Table 4.1 HR Increase Ratio by Interface Variations

Interface Variations	Enhanced	Baseline	Degraded
HR Increase Ratio Mean (SD)	0.037 (0.030)	0.060 (0.049)	0.061 (0.052)

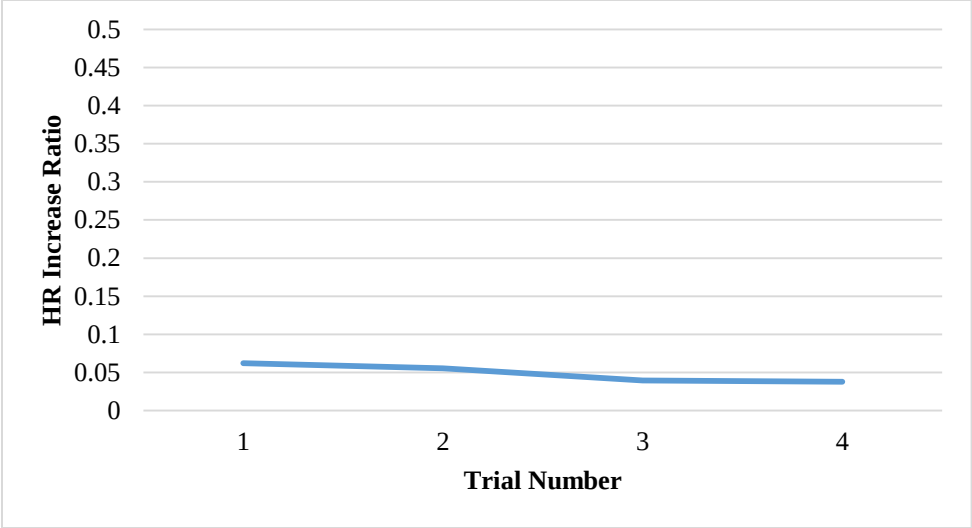

Figure 4.2 Trial Number Effect on HR Increase Ratio

4.3 HF decrease ratio

The ANOVA on the HF decrease ratio revealed no significant effect of the interface design variation ($F(2,42) = 0.46$, $p = 0.6373$), vehicle speed ($F(1, 42) = 0.04$, $p = 0.8474$), or their interaction ($F(2,42) = 3.111.13$, $p = 0.0550$). The trial effect was found to be significant ($F(1,42) = 5.10$, $p = 0.0292$), with a decreasing trend towards later trials (Figure 4.3).

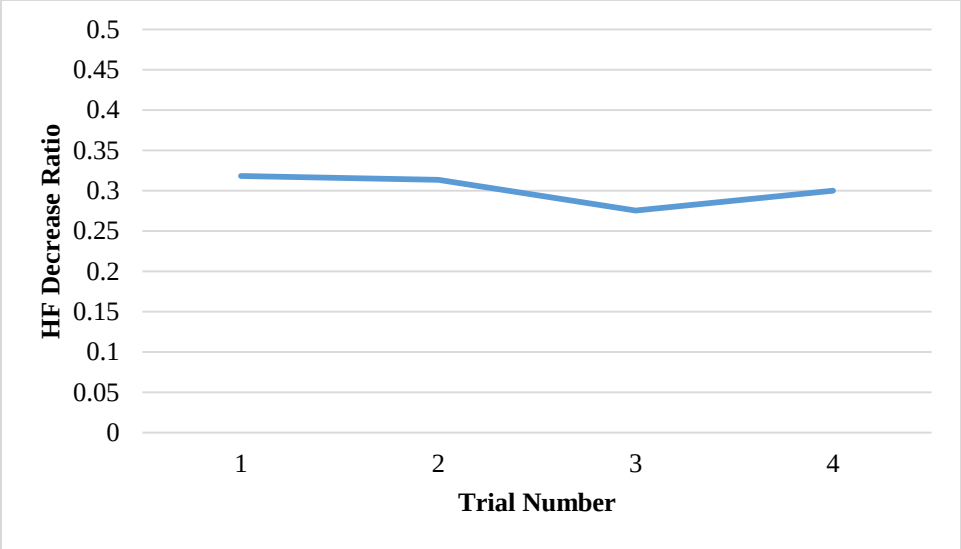

Figure 4.3 Trial Number Effect on HF Decrease Ratio

4.4 NASA-TLX ratings

The ANOVA on ranked NASA-TLX ratings revealed a significant effect of the interface design variation ($F(2,45) = 3.24$, $p = 0.0486$), vehicle speed ($F(1, 45) = 16.73$, $p = 0.0002$) and trial number ($F(1,45) = 26.81$, $p < .0001$). Tukey's HSD post-hoc tests revealed TLX ratings for the enhanced interface to be significantly lower than both the baseline and degraded interface conditions (Figure 4.4). Figure 4.5 shows that the high vehicle speed condition produced significantly higher TLX ratings than the low speed condition. Similar to the trend on other workload response measures, ratings were found to decrease towards later test trials (Figure 4.6).

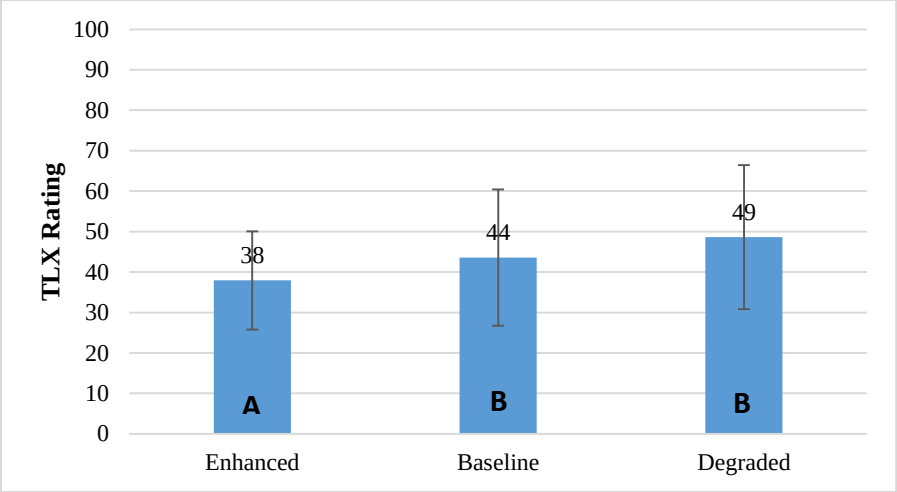

Figure 4.4 Interface Variation Effect on TLX Rating

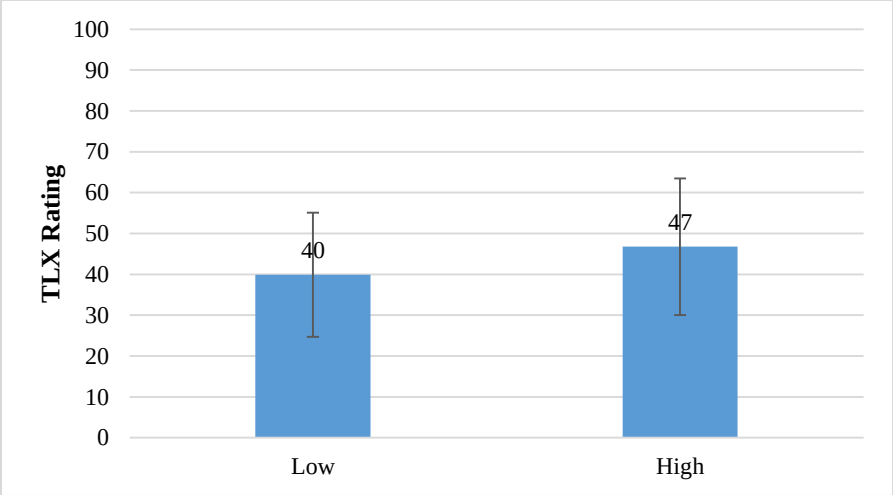

Figure 4.5 Vehicle Speed Effect on TLX Rating

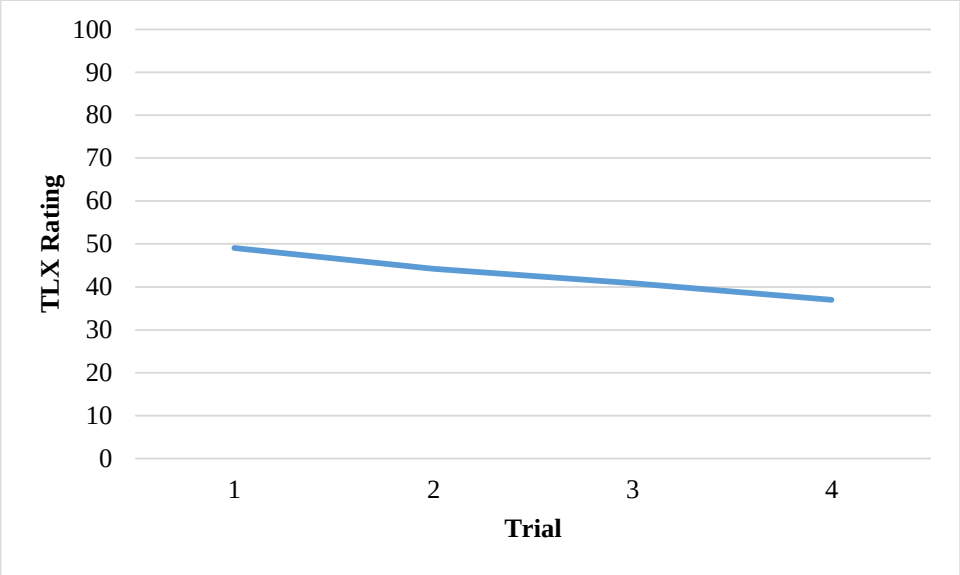

Figure 4.6 Trial Number Effect on TLX Rating

ANOVA results did not reveal significance of the "Interface Variation * Vehicle Speed" interaction effect ($F(2,45) = 2.88$, $p = 0.0665$) on the TLX scores. However, the untransformed data indicated that the enhanced interface lead to the smallest increase in subjective workload under the high vehicle speed (i.e., task event rate) condition, followed by the baseline interface and then the degraded interface (Figure 4.7).

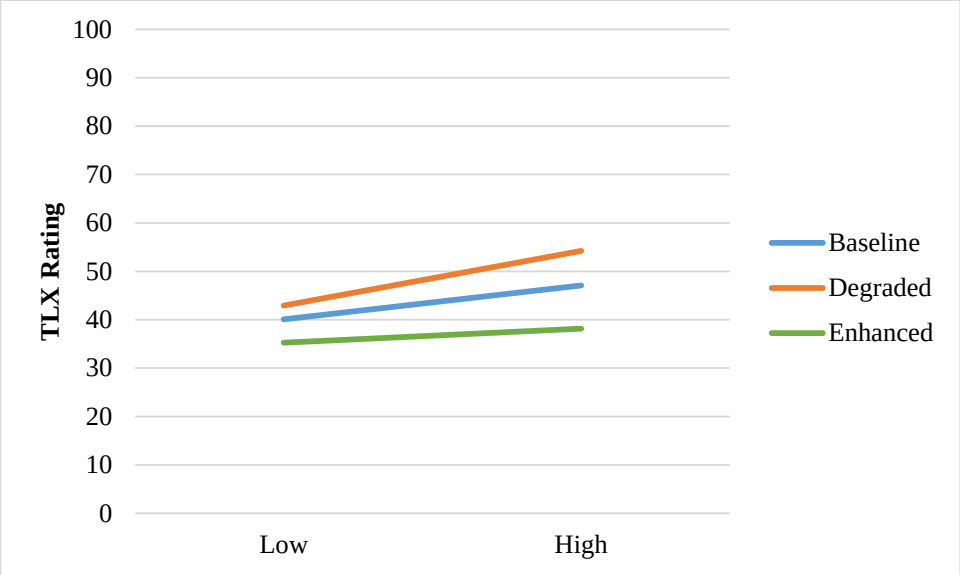

Figure 4.7 Interaction Effect of "Interface Variation * Vehicle Speed" on TLX Rating

In order to gain further insight into the potential causes of operator workload in the UAV control operations, ratings on each dimension of the TLX (physical, mental, effort, performance, frustration and temporal) were also analyzed separately. Results revealed significance of the interface design variation on "effort" and "frustration" ratings, but not for the other sub-dimensions of the overall index (Table 4.2). Tukey HSD test results revealed the enhanced interface to require significantly less effort than both the baseline and degraded interfaces. The enhanced interface also produced the least amount of frustration for participants, followed by the baseline condition and then the degraded interface.

Table 4.2 Interface Variation Effect on TLX Sub-component Ratings (Note: * indicates statistical significance.)

	ANOVA Results	Tukey HSD Results
Mental Demand	$F(2,45) = 0.52$, $p = 0.5973$	-
Physical Demand	$F(2,45) = 0.96$, $p = 0.3897$	-
Temporal Demand	$F(2,45) = 0.84$, $p = 0.4365$	-
Performance	$F(2,45) = 2.63$, $p = 0.0830$	-
Effort	$F(2,45) = 3.43$, $p = 0.0412$*	Enhanced < Baseline ≈ Degraded
Frustration	$F(2,45) = 3.51$, $p = 0.0384$*	Enhanced < Baseline < Degraded

The results also revealed a significant effect of vehicle speed (Table 4.3) on the ratings for “mental demand”, “physical demand”, “temporal demand” and “performance”, with lower ratings under the low speed (event rate) condition. (Note: As a result of the orientation of the anchors as part of the TLX rating scales, a lower performance rating indicates better performance and lower workload induced by task performance.) Although the vehicle speed effect was not significant for “effort” and “frustration” ratings, it was observed that the low speed condition produced, on average, lower ratings for these two sub-dimensions.

Table 4.3 Vehicle Speed Effect on TLX Sub-component Ratings

	ANOVA Results	Trends
Mental Demand	F(1, 45) = 7.81, p = 0.0076*	Low < High
Physical Demand	F(1, 45) = 6.42, p = 0.0149*	Low < High
Temporal Demand	F(1, 45) = 20.53, p < .0001*	Low < High
Performance	F(1, 45) = 13.08, p = 0.0007*	Low < High
Effort	F(1, 45) = 2.67, p = 0.1089	Low < High
Frustration	F(1, 45) = 3.47, p = 0.0690	Low < High

The interaction effect of interface design variation and vehicle speed was only found to be significant for the "effort" ratings (Table 4.4). As shown in Figure 4.8, the enhanced interface produced the smallest increase in ratings under the high vehicle speed (i.e., high event rate) condition, as compared with the low speed. The usable interface was followed by the baseline interface and then the degraded interface in terms of increasing workload demands.

Table 4.4 Interaction Effect of Interface Variation and Vehicle Speed on TLX Sub-component Ratings

	ANOVA Results
Mental Demand	F(2,45) = 1.36, p = 0.2683
Physical Demand	F(2,45) = 0.85, p = 0.4340
Temporal Demand	F(2,45) = 2.34, p = 0.1083
Performance	F(2,45) = 1.03, p = 0.3649
Effort	F(2,45) = 3.72, p = 0.0320*
Frustration	F(2,45) = 0.14, p = 0.8721

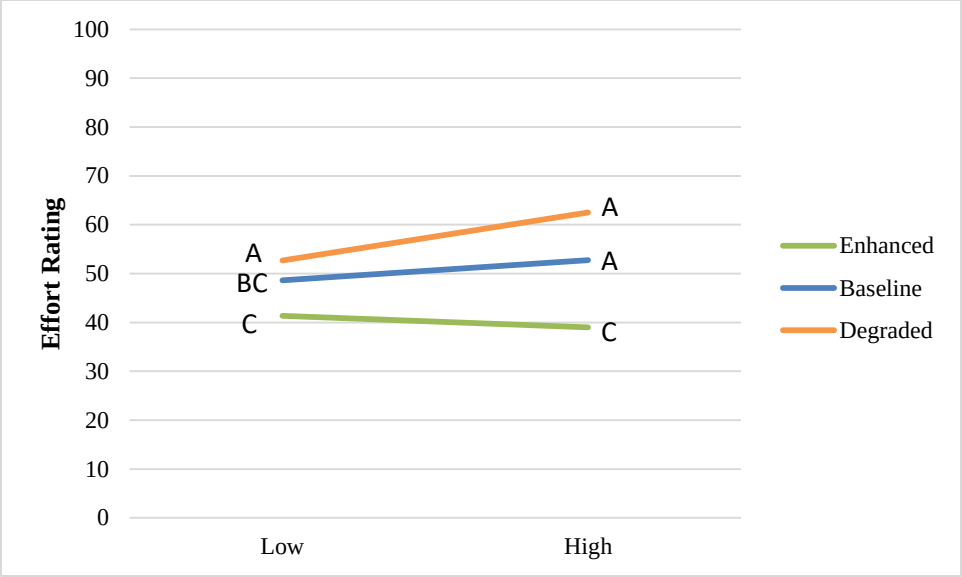

Figure 4.8 Interaction Effect of "Interface Variation * Vehicle Speed" on Effort Ratings (letters in the plot indicate Tukey groupings of the condition mean values)

For most TLX sub-dimension ratings, the trial number effect was significant (Table 4.5), with a decreasing trend towards later trials.

Table 4.5 Trial Number Effect on TLX Sub-component Ratings

	ANOVA Results
Mental Demand	$F(1,45) = 28.15$, $p < .0001$*
Physical Demand	$F(1,45) = 0.59$, $p = 0.4459$
Temporal Demand	$F(1,45) = 15.28$, $p = 0.0003$*
Performance	$F(1,45) = 12.19$, $p = 0.0011$*
Effort	$F(1,45) = 15.00$, $p = 0.0003$*
Frustration	$F(1,45) = 14.46$, $p = 0.0004$*

4.5 DKQ accuracy

An ANOVA on DKQ accuracy revealed a significant effect of the interface design variation ($F(2,45) = 16.43$, $p < .0001$), but not for vehicle speed ($F(1, 45) = 1.04$, $p = 0.3133$) or their interaction ($F(2,45) = 2.03$, 0.1431). Tukey's HSD tests revealed that participants using the enhanced interface had the highest DKQ accuracy, followed by participants using the baseline interface and then the degraded interface (Figure 4.9). The trial effect was also significant ($F(1,45) = 6.93$, $p = 0.0116$), with an increasing trend towards later trials (Figure 4.10).

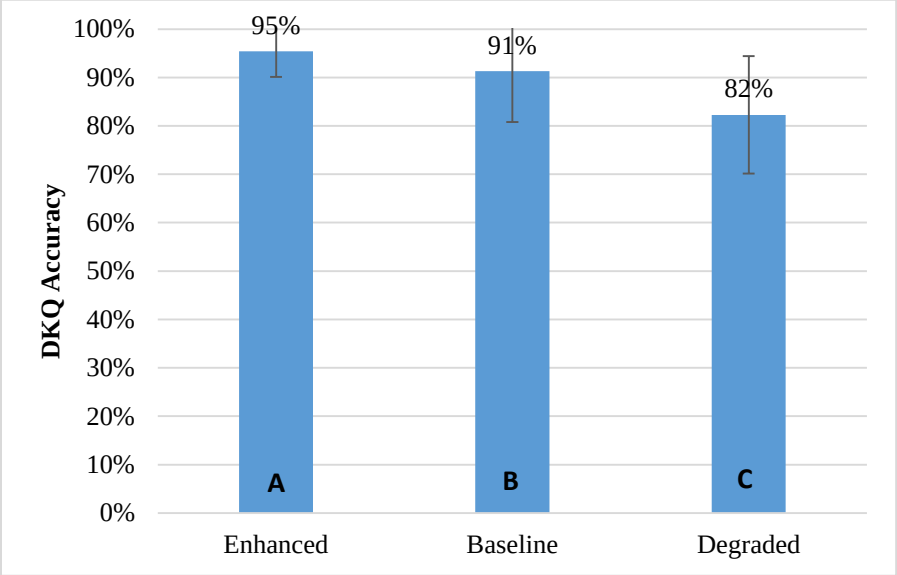

Figure 4.9 Interface Variation Effect on DKQ Accuracy

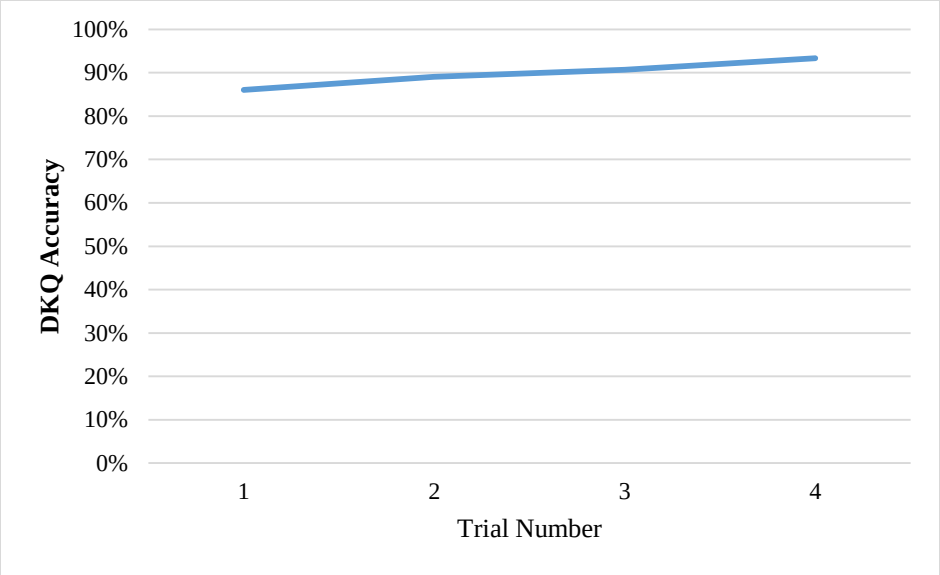

Figure 4.10 Trial Number Effect on DKQ Accuracy

4.6 MCH-UVD

Due to the timing of the collection of MCH-UVD ratings as part of the experiment, observations were only collected on eight participants for each interface. Results revealed no significant effect of the interface design variation ($F(2,21) = 2.45$, $p = 0.1107$) or vehicle speed ($F(1, 21) = 1.11$, $p = 0.3042$). The mean ratings did not show the expected trend of increase with degraded interface usability (Table 4.6); that is, there was a lack of sensitivity of the measure. However, the "interface variation * vehicle speed" interaction appeared to be significant ($F(2,21) = 5.09$, $p = 0.0158$; Figure 4.11). Ratings for the degraded interface were higher (indicating "unsatisfactory" design) under the high speed (event rate) condition, as compared to the low speed condition.

Table 4.6 MCH-UVD Ratings for Interface Variations

Interface Variations	Enhanced	Baseline	Degraded
MCH-UVD Ratings Mean (SD)	3.1 (1.9)	2.5 (1.1)	4.4 (2.5)

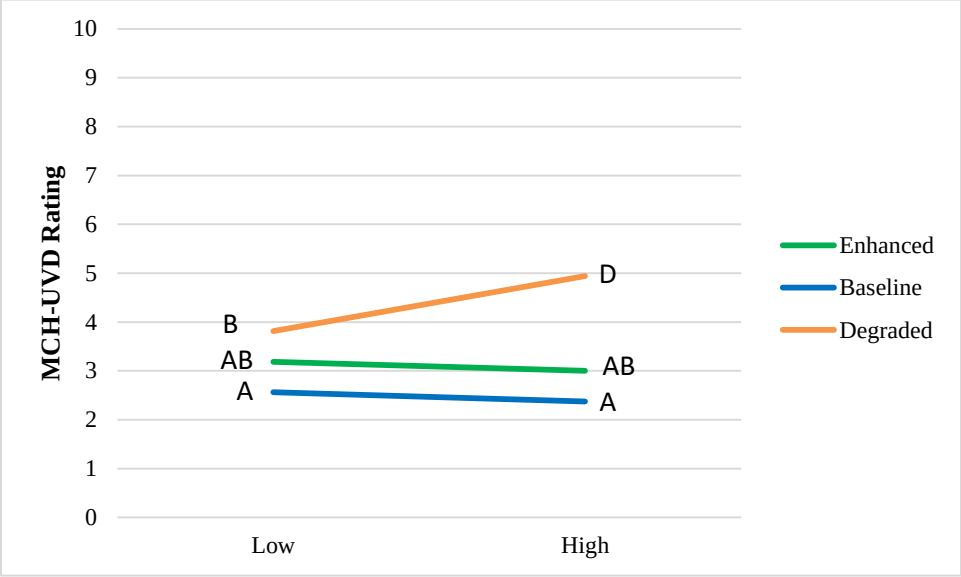

Figure 4.11 Interaction Effect of "Interface Variation * Vehicle Speed" on MCH-UVD Rating (letters in the plot indicate Tukey groupings of the condition mean values)

4.7 Regression analysis

4.7.1 Selection of predictors

Many of the GOMSL model outputs can be considered as potential predictors of operator workload in UAV control. However, in this study, output selection was restricted by the mission scenario design. In addition to the series of UAV tasks listed in the vehicle "scheme of maneuver" (see Section 3.3), the UAV missions also required operators to monitor vehicle status parameters. In every mission (i.e., test trial), operators performed one time each of the tasks listed in the "scheme of maneuver". Consequently, the GOMSL model parameters related to these tasks could be easily determined. However, the amount of time that each participant allocated to the monitoring task (or the number of times the monitoring task was performed during each trial) was unknown. Without this information, some GOMSL model

parameters, such as the total number of cognitive operators during an entire mission, could not be calculated. For this reason, only the maximum number of cognitive operators for all UAV tasks (denoted as MaxCog) and the average number of peak WM chunks (denoted as WM) across all tasks were used as predictors of workload responses. Here, it is important to note that the peak WM chunk count for an entire operation has previously been used as a measure of cognitive workload (e.g., Lerch, Mantei, & Olson, 1989). However, using the average of multiple peak values represents an alternate approach that has been an established in physical task workload analysis using electromyography measurement (e.g., Zhang, Ma, et al., 2016). The mean and the (standard deviations) of values for all workload predictors, including MaxCog, WM chunk count and the Modified GEDIS-UAV scores (denoted as GEDIS), are summarized in Table 4.7. In order to avoid implicit weighting caused by different ranges and units of measures, all predictors were subsequently normalized based on maximum response values.

Table 4.7 Workload Predictors

Predictor Information: Actual value (Normalized value)			
	Enhanced	**Baseline**	**Degraded**
MaxCog	39 (0.8298)	46 (0.9787)	47 (1)
WM	3.11 (0.666)	4.11 (0.8801)	4.67 (1)
GEDIS	0.86 (1)	0.78 (0.907)	0.67 (0.7791)

4.7.2 Regression outputs

In order to isolate the effect of the interface condition on workload responses, the dataset for the regression analysis was parsed into two parts - one for the low speed test condition and one for high speed condition. Regression analysis was applied to each half of the dataset. The

outcomes for a set of simple (single parameter) regression models are summarized in Table 4.8. Although the intercept parameter was not significant ($p > 0.05$) for any model, it was still included forcing a no-intercept model might have compromised the degree of model fit to the data for other values different from zero.

Table 4.8 Regression Analysis Results

	Response	**Predictor**	**Parameter Estimates**	**RSquare**	**P value**
High Speed Condition	TLX Score	Normalized WM	6.26 + 47.40*WM	0.1501	<.0001*
	TLX Score	Normalized MaxCog	-30.09 + 81.80*MaxCog	0.1349	0.0002*
	TLX Score	Normalized GEDIS	110.98 - 72.05*GEDIS	0.1478	<.0001*
	BD Decrease Ratio	Normalized WM	-0.05 + 0.21*WM	0.0656	0.0127*
	BD Decrease Ratio	Normalized MaxCog	-0.20 + 0.35*MaxCog	0.0557	0.0221*
	BD Decrease Ratio	Normalized GEDIS	0.41 - 0.32*GEDIS	0.0677	0.0113*
Low Speed Condition	TLX Score	Normalized WM	20.06 + 22.80*WM	0.0415	0.0454*
	TLX Score	Normalized MaxCog	1.95 + 40.00 *MaxCog	0.0388	0.0532
	TLX Score	Normalized GEDIS	69.90 - 34.06*GEDIS	0.0393	0.0516
	BD Decrease Ratio	Normalized WM	-0.06 +0.19*WM	0.0870	0.0039*
	BD Decrease Ratio	Normalized MaxCog	-0.24 + 0.36*MaxCog	0.0923	0.0029*
	BD Decrease Ratio	Normalized GEDIS	0.34 - 0.27*GEDIS	0.0717	0.0090*

In general, results revealed normalized WM chunk count to account for the greatest fraction of variability in the TLX scores under both the low and high vehicle speed test conditions. There did not appear to be a particular predictor with consistently greater utility in accounting for variability in the BD decrease ratio. Overall, the regression results indicated that there are likely other factors, beyond task complexity and interface usability and functionality, that drive operator workload in UAV control operations.

4.8 Correlation analysis

Among the physiological response measures collected during the experiment, only the BD decrease ratio revealed sensitivity to the effect of the interface design variations. Therefore, correlation analyses were conducted among the BD decrease ratio, composite TLX ratings and the DKQ accuracy. Both the BD decrease ratio and TLX ratings were found to significantly and negatively correlated with the DKQ accuracy (Table 4.9). That is, as workload demands in the UAV control simulation decreased, participant were more accurate in responding to the dynamic knowledge queries on vehicle state and operation status.

Table 4.9 Correlation Analysis Results

Variable	By Variable	Spearman ρ	P value
TLX	BD Decrease	0.1365	0.0618
DKQ	BD Decrease	-0.1895	0.0098*
DKQ	TLX	-0.2279	0.0017*

However, results did not reveal significant correlations between the BD decrease ratio and composite TLX ratings. In order to further understand the workload sources (or demands) that might be influential for the eye-tracking responses, correlation analysis between BD decrease ratio and TLX sub-component ratings was conducted. The results revealed a significant and positive correlation of BD decrease ratio with the Physical Demand and Performance ratings (Table 4.10). Note that a higher performance rating indicates unsatisfactory or unsuccessful performance. Therefore, the positive correlation suggested the decrease in performance with increased BD decrease ratio (i.e., increased cognitive workload).

Table 4.10 Correlation between BD Decrease and TLX Sub-component ratings

Variable	By Variable	Spearman ρ	P value
BD Decrease	Mental Demand	0.1401	0.0551
BD Decrease	Physical Demand	0.2089	0.0040*
BD Decrease	Temporal Demand	0.0717	0.3285
BD Decrease	Performance	0.1516	0.0378*
BD Decrease	Effort	0.0569	0.4376
BD Decrease	Frustration	0.0754	0.3039

5. Discussion

5.1 Effect of interface variation

In general, expectation was for the enhanced interface to produce the greatest utility for moderating UAV operator cognitive workload. The first hypothesis (H1) stated that the enhanced interface would result in the lowest BD decrease ratio, followed by the baseline interface and then the degraded interface. This hypothesis was partially supported by the results. The enhanced interface resulted in a lower BD decrease ratio than both the baseline and degraded interfaces. However, there was no significant difference between the baseline and degraded interfaces.

This result may be attributable to the type of interface feature manipulations among the three interfaces, which can be generally classified as pure usability feature improvements and automation (see Tables 3.1 and 5.1). The enhanced interface provided more functionality by implementing automation. For example, when the operator was required to obtain coordinates on a specific target in using the degraded interface, the operator had to visually align the target location with horizontal and vertical axis on the map display and make an estimation at the required level of accuracy. The baseline interface provided auxiliary gridlines to assist the visual alignment and target location identification. However, with the enhanced interface, all the operator had to do was to click the "Coordinates" tool button and then the desired target. This automation significantly reduced operator workload in terms of information acquisition and analysis (Parasuraman, Sheridan, & Wickens, 2000). With respect to usability features, improvements in the enhanced interface design were also substantial. For example, the baseline interface presented text messages to operators

regarding alarm solutions and priority levels in order to facilitate alarm handling tasks, while the enhanced interface highlighted the messages with colors and different warning icons. This feature further reduced operator effort in reading and interpreting the text messages on alarms. As the most challenging tasks during the UAV mission, the alarm handling and coordinates/distance estimation tasks were significantly facilitated with the usability and functionality features provided by the enhanced interface. However, the differences between the baseline and degraded interfaces were not as notable.

This situation may also be explained in terms of the GOMSL model outcomes. In general, the enhanced interface posed the lowest degree of task complexity and required the fewest steps, followed by the baseline condition and then the degraded interface. However, a close inspection of the number of cognitive operations and WM requirements provides more insight into the experiment results. While the maximum number of cognitive operations (MaxCog) in UAV control tasks was only 39, when using the enhanced interface, the numbers for the baseline and degraded interface were 46 and 47, respectively. That is to say, the difference in cognitive resources required by the baseline and degraded interfaces was not as significant. Similar observations can be made on the WM requirements posed by each interface design. As Kieras (1999) suggested, the need for an operator to maintain more than 5 tags in WM at once is problematic. When using the enhanced interface, none of the UAV control tasks required use of more than 5 WM chunks. In contrast, 2 tasks with the baseline interface and 3 tasks with the degraded interface likely overloaded user WM. Moreover, based on the WM usage requirement, “Report distance” appeared to be the most challenging among all the UAV control tasks. While the enhanced interface required users to maintain

five WM tags at once in performing this task, the number increased to 7 chunks for both the baseline condition and degraded interface. These GOMSL outcomes also justify the lack of significant difference between the baseline and degraded interfaces.

Table 5.1 Classification of Interface Feature Manipulations

Variation ID	Interface Components	Baseline vs. Degraded	Enhanced vs. Baseline
1	Menu	Usability features	Usability features
2	Flight parameter display	Usability features	Usability features and automation
3	Waypoints	Usability Features	Usability Features
4	Area of interest (AOI)	Usability Features	Usability Features
5	Coordinates	Usability Features	Usability features and automation
6	Alarms	Usability features and automation	Usability features and automation
7	Confirmation for writing waypoints	Usability features and automation	Usability features and automation

Nevertheless, the BD decrease ratio was the only physiological measure that revealed workload differences among interface designs. It is likely that this measure was sensitive because of the highly visual nature of the UAV control tasks as well as the simulator displays being the only recourse to address cognitive workload. Previous literature (e.g., Kramer, 1990; Wierwille & Eggemeier, 1993) also suggested blink duration to be a sensitive measure of cognitive workload. The above findings further demonstrated the utility of eye-tracking technology for workload prediction in UAV operations.

The research also posited that the enhanced interface would result in the lowest HR increase ratio (H2) and HF decrease ratio (H3), followed by the baseline interface and then the degraded interface. Results did not support this hypothesis. Neither the HR increase ratio nor the HF decrease ratio revealed different workload levels among the participant groups using the three different interfaces. This finding is not consistent with some previous research. Abich et al. (2015) found HRV to be a good predictor of map icon change (i.e.,

move, disappear, or appear) detection task performance and to correlate with EEG responses. However, mixed results appear in the literature. Brookings et al. (1996) applied HR and HRV in assessing ATC workload under different levels of scenario difficulty. They reported that HR did not reveal significant differences among difficulty or traffic manipulations and that there were only marginal significant differences for HF HRV responses. The authors speculated that HR and HRV were not sensitive to differences in the levels of mental demand for the ATC task. It was concluded that one level of HR activity was sufficient to address the metabolic demands of the brain at all levels of task demand. Harriott et al. (2015) also collected HR and HRV to compare workload in human-human teams vs. human-robot teams. They found no significant difference in HR and HRV among the two conditions while other workload measures (e.g., respiration rate and subjective ratings) showed sensitivity. Jorna (1992) reviewed a number of studies and summarized that the sensitivity of HRV depended on the manipulation of task characteristics. For example, when stimulus degradation, response alternatives or stimulus-response incompatibility was manipulated in the tasks, no effects were found in spectral measures of HRV. However, HRV did show sensitivity when it came to distraction in visual search or time-sharing in working memory search. Interestingly, for workload memory load tasks, increasing working memory demands did not affect HRV with college students serving as subjects, but showed significance when bus drivers were tested (Jorna, 1992). Based on these findings, Jorna (1992) concluded that rest-task effects consistently resulted in changes in HRV, but it seemed very difficult to affect HRV by simply manipulating task characteristics.

It is also important to note that the interface design variation in the present study was manipulated as a between-subject factor. The normalization approach was intended to account for individual differences in physiological state but did not appear sufficient for several response measures. In addition to the physical conditions of participants, such as fitness level and age (Jorna, 1992), other researchers have also found different task coping strategies and personality traits (e.g., neuroticism and anxiety) to result in different levels of cardiovascular arousal. With respect to the present study, although the enhanced interface produced a lower HR increase (3.7% increase) than both the baseline condition (6% increase) and degraded interface (6.1%), a large standard deviation in the response (4.5%) of the sample population prohibited identification of statistically significant differences among the test conditions.

The fourth hypothesis (H4) surmised that the enhanced interface would result in the lowest NASA-TLX ratings, followed by the baseline interface and then the degraded interface. Results partially supported this hypothesis. The enhanced interface resulted in significantly lower TLX ratings than both the baseline and degraded interface. However, no significant difference was found between the baseline and degraded interfaces. An analysis on the TLX sub-component ratings revealed that participants using the enhanced interface also reported significantly lower ratings for the "Effort" component, suggesting that the enhanced interface design required significantly less effort than both the baseline and degraded interfaces for task performance. As discussed earlier, the absence of a significant difference between the baseline and degraded interfaces, in terms of workload, may have been due to limited variations in usability and functionality features. It is also possible that

participants reached a plateau in workload and ceased additional effort with the degraded interface. More interestingly, the results on "Frustration" ratings not only revealed a significant effect of interface variation, but also significant pairwise differences among all three interfaces with the lowest ratings occurring for the enhanced interface, followed by the baseline interface and then the degraded interface. The "Frustration" component of the TLX measures how insecure, discouraged, irritated, stressed and annoyed a participant feels during task performance. The present findings suggest that the usability improvements and automation implementation as part of the enhanced interface were effective in supporting smoother and easier UAV operations. In general, the NASA-TLX has been a well-accepted cognitive workload survey for a substantial period of time (Hill et al., 1992). Numerous researchers (e.g., Crescenzio et al., 2009; Lu et al., 2013; Ruff et al., 2002) have utilized the technique and reported its sensitivity to system design manipulations and task difficulty. The present study provides further validation and supported for application of the NASA-TLX for assessing operator cognitive workload in UAV operations.

Participants using the enhanced interface prototype were also expected to produce the highest DKQ accuracy levels, followed by the baseline interface and then the degraded interface (H5). This hypothesis was supported by the experiment results. Differences in DKQ accuracy among the three interface design variations can be explained in terms of the usability and functionality features of the designs. As compared to the degraded interface, the baseline interface condition made it easier for operators to answer queries (see Table 3.9) as a result of the AOI filter, UAV parameter normal/acceptable range display, logically grouped menu options, auxiliary gridlines on the map display, and easily recognizable waypoint

icons. The enhanced interface further supported accurate responses to queries by making the AOI filter more visually elaborate (see prototype display figures), implementing a hierarchical menu structure and color-coding waypoints, as an indication of mission progress. Beyond this, the enhanced interface design made the UAV operating tasks easier to accomplish and, therefore, left participants with greater cognitive resources for addressing DKQs.

In summary, the general expectation that the enhanced interface design would be superior for moderating operator workload was supported by multiple workload measures. However, any distinction between the baseline and degraded interface conditions was not as notable and influential in operator workload.

5.2 Vehicle speed manipulation

Vehicle speed was used to manipulate task event rate during the experiment. The fast vehicle speed or event rate was expected to introduce time pressure for participants and, therefore, make task performance more cognitively (and physically) demanding. Results on the physiological responses and DKQ accuracy were not in line with expectation. Although the high speed condition resulted in a slightly higher HR increase ratio (5.7% vs. 4.7%), higher BD decrease ratio (12.7% vs. 10.1%), and lower DKQ accuracy (88.9% vs. 90.6%) than for the low speed condition, these differences were no statistically significant. However, results on the TLX ratings and several sub-components ratings (i.e., mental demand, physical demand, temporal demand and performance) were consistent with expectation. Findings suggested that the speed manipulation was enough to cause a difference in self-perceived cognitive workload. However, it was not sufficient to cause changes in physiological states.

Time pressure has been demonstrated to be a reliably influential factor in cognitive workload (e.g., Wang, He, Chen, 2016). In future studies, an even higher speed condition may be applied to produce faster event rates and generate higher task demands for participants, possibly leading to different physiological responses.

5.3 Interaction effect of interface variation and vehicle speed

The enhanced interface was also expected to be more robust in terms of moderating user workload under highly demanding task circumstances. That is, under the fast vehicle speed/event rate condition, the enhanced interface was expected to produce the smallest increase in workload (if any), as compared with the slow vehicle speed condition.

Related to this, the interaction of the "Interface Variation * Vehicle Speed" was expected be significant for the physiological workload measures, with differences in BD decrease (H6), HR increase (H7), HF decrease (H8) being greater among the interface designs when used at the higher vehicle speed. Several hypotheses (H6 – H8) were refuted by experiment results. None of the physiological responses revealed a significant interaction effect, as expected. The literature includes several studies that have identified interaction effects of "time pressure induced task demand" with other influential workload factors. Liu, Peterson, Vincenzi and Doherty (2016) investigated the effect of time pressure and target uncertainty on human operator performance and workload in control of "autonomous" unmanned aerial systems. Tasks included target acquisition, processing intruder aircraft, monitoring mission mode and handling "pop-up" threats based on automation recommendations. In addition to a significant effect of time pressure on task performance, it was found that the high time pressure condition exaggerated differences in responses caused

by target uncertainty. However, the authors did not find similar results in terms of TLX ratings due to an insignificant uncertainty effect. Zhao and Wu (2013) studied the effects of image size and time pressure on human performance and mental workload in an image search task. They found time pressure to mitigate any difference in search time caused by image size but to exaggerate differences in cognitive workload (supported by HRV responses but not NASA-TLX ratings). In the present study, the findings on the physiological workload responses were not consistent with these prior results. The present findings may be attributable to large individual differences in physiological states. In addition, the lack of significance of the vehicle speed manipulation also compromised the opportunity to assess the robustness of the enhanced interface design for moderating operator workload under high task demands.

The interaction of "Interface Variation * Vehicle Speed" was also expected to be significant on NASA-TLX ratings (H9). The overall TLX scores did not support this hypothesis. Although descriptive statistics revealed greater differences in workload ratings among the interface conditions under the high vehicle speed (see Figure 4.9 in Section 4.4), the differences were not statistically significant. Nonetheless, ratings for the "Effort" sub-component were in line with expectation. There were greater variations in effort among the interface designs when participants were posed with the high vehicle speed condition. While high vehicle speed resulted in increased levels of effort (as compared to low vehicle speed) for both the baseline and degraded interface, there was barely any increase in effort for the enhanced interface. This finding suggests that the enhanced interface was robust in moderating overall operator effort when the task became highly demanding.

The last hypothesis (H10) speculated that the interaction of "Interface Variation * Vehicle Speed" would be significant for DKQ accuracy. This hypothesis was refuted by the results. The enhanced interface did not appear to be more robust for supporting participant dynamic knowledge of the system. This is not surprising since the DKQ accuracy was not influenced by the event rate manipulation.

5.4 Trial number effect

A significant effect of trial number was observed for most of the workload responses, including HR increase ratio, HF decrease ratio and the TLX scores. The trial effect was also significant for multiple TLX sub-component ratings, including mental demand, temporal demand, performance, effort and frustration. Moreover, significant improvements in DKQ accuracy were also observed in later test trials. These findings suggest that participants experienced additional learning during experiment trials. In order to prevent potential carryover effects caused by learning, familiarization with interface features and UAV mission requirements, hands-on training was provided for participants prior to testing. Participants were required to repeat mission training until they were able to complete a zero-error operation. Due to this requirement, most participants practiced the training mission with their assigned interface with two or three repetitions. The amount of training was considerable and was expected to eliminate further learning during experiment trials. However, the results suggest that participants could have been trained to a higher degree of proficiency in the UAV control operations. In addition to further proficiency developing during test trials, participants might have been anxious for the beginning of test trials and then became more relaxed and emotionally prepared in later trials. This would also explain

the drop in cognitive workload and improvement in dynamic knowledge of the system over time.

5.5 Utility of GOMSL outputs and GEDIS scores in cognitive workload prediction

The single parameter regression models demonstrated some utility of the GOMSL model outputs and the GEDIS scores for predicting cognitive workload responses. However, due to the limited number of parameters included in the models, the RSquare values were not high. This finding means that operator workload cannot be reliably predicted on the basis of the single predictor models alone. Nevertheless, the parameter estimates suggests that an increase in WM items and cognitive operations during UAV control tasks causes an increase in workload responses. On the contrary, higher Modified GEDIS-UAV scores are associated with lower cognitive workload. Since the model predictors were normalized to their maximum values, the parameter estimates also suggest relative utility of the predictors in accounting for variability in workload responses. Across all speed conditions, WM chunk count revealed the greatest influence in subjective workload rating in UAV control, followed by the Modified GEDIS-UAV scores (i.e., GEDIS) and then the maximum number of cognitive operations (i.e., MaxCog). However, the Modified GEDIS-UAV scores and maximum number of cognitive operations accounted for greater degree of variability in the physiological response. One inference is that operator perceived workload was driven by the WM requirements from the tasks, but the most complex task and interface features (as indicated by the “MaxCog” and “GEDIS”) were more influential to operator physiological state.

5.6 Correlation among workload responses

Counter to expectation, results did not reveal significant correlations between the physiological workload response measures (i.e., BD decrease ratio) and subjective ratings. Although the BD decrease ratio successfully revealed operator workload differences in using the three interface designs, individual response values did not vary closely with subjective workload ratings. The lack of significant might also suggest the complementarity of eye-tracking measure to subjective approach in workload assessment. Related to the analysis on BD decrease ratio and TLX sub-component ratings, results revealed significant and positive correlation between the BD decrease ratio and the Physical Demand ratings. The Physical Demand was rated based on how much physical activity was required and if the task was easy vs. demanding, slow vs. brisk, slack vs. strenuous and restful vs. laborious. Therefore, higher Physical Demand ratings were likely associated with more demanding control actions during the task performance, which possibly produced greater visual attention needs and, therefore, higher BD decrease ratio for participants. The BD decrease ratio also showed significant positive correlation with the Performance Demand, indicating the decrement in performance under higher cognitive workload. With respect to DKQ accuracy, a significant negative correlation was found with both the BD decrease ratio and TLX scores. This finding also suggested that the increase in cognitive workload led to a decrement in performance and cognitive resources.

5.7 Usability questionnaire

The usability questionnaire provided additional insights on user opinions regarding the interface designs. As shown in Table 5.2, 90% of participants that used the enhanced

interface agreed that it was easy to use, as compared to 72% for the baseline interface and 50% for the degraded interface. Ninety percent (90%) of participants assigned to the enhanced interface agreed it was easy to find information, as compared to 64% for the baseline interface and 60% for the degraded interface. Moreover, 90% of enhanced interface users considered the interface to be effective for supporting task completion, versus 73% and 60% for the baseline and degraded interface, respectively. However, the enhanced interface did not show advantages over the baseline or degraded interfaces in terms of making participants feel task performance was easier. Regardless of which interface was used, most participants (70% ~ 80%) thought the UAV control tasks were easy to accomplish. The reason for this finding might be the UAV missions were not completely challenging for participants. It is also possible that participants gained more confidence for task performance by the time that they completed the questionnaire at the end of experiment.

Table 5.2 Participant Response to Usability Statements (1 = Strongly disagree; 5 = Strongly agree)

		Rating				
Statements	Interface	1	2	3	4	5
Overall, the UAV control interface was easy to use.	Enhanced	0%	0%	10%	50%	40%
	Baseline	0%	18%	9%	36%	36%
	Degraded	0%	0%	50%	40%	10%
It was easy to find information I needed.	Enhanced	0%	0%	10%	70%	20%
	Baseline	0%	9%	27%	55%	9%
	Degraded	0%	30%	10%	50%	10%
The interface was effective in helping me complete the tasks.	Enhanced	0%	0%	10%	50%	40%
	Baseline	0%	9%	18%	64%	9%
	Degraded	0%	10%	30%	40%	20%
The UAV control tasks were easy to accomplish.	Enhanced	0%	0%	30%	40%	30%
	Baseline	0%	18%	9%	64%	9%
	Degraded	0%	10%	10%	60%	20%

Related to open comments on interface usability, many enhanced interface users thought the variation was well designed and they especially appreciated the color-coding features for alarm handling. Some baseline interface users found it difficult to monitor deviations of UAV status parameters, locate targets on map and estimate distances. These complaints were similar and even more prevalent with the degraded interface design. In addition, participants also complained about the illogical layout of menu structure with the degraded interface and suggested that the interface provide recommendations for appropriate actions to address time-sensitive tasks. These usability questionnaire responses indicated that the design of the enhanced interface successfully addressed many complaints from UAV operators associated with less usable interface designs.

5.8 Comparison between Modified GEDIS-UAV and MCH-UVD

The MCH-UVD ratings did not prove to be sensitive to variations in UAV control interface designs. The lack of significant differences among interface designs may have been due to the limited number of ratings collected during the experiment. However, the trend of ratings across interface designs was concerning. MCH-UVD ratings revealed operator information processing to be most difficult with the degraded interface. However, the average rating was ~4.38, which indicates design "Deficiencies Warranting Improvement" but not major deficiencies meriting mandatory redesign. Moreover, the baseline interface resulted in a lower average rating (greater perceived usability) than the enhanced interface, which is counter to other findings of the present study (i.e., Modified GEDIS-UAV scores, usability heuristic evaluation, cognitive workload measures and usability questionnaire responses). It

is possible that the MCH-UVD is only able to detect serious deficiencies in interface design for supporting human information processing. However, as long as the interface is functionally "acceptable" for required tasks, the tool may not identify the need for further design improvements. Interestingly, results on the MCH-UVD ratings did reveal a significant interaction effect of the interface design and vehicle speed setting. While ratings only varied slightly among the interface designs under the low vehicle speed condition, the ratings for the degraded interface design under the high vehicle speed were substantially greater than for the enhanced and baseline interfaces. This outcome suggests that interface evaluations based on the MCH-UVD may be significantly influenced by task demands. The MCH-UVD ratings also appeared to be highly subjective in nature (based on personal opinions) and influenced by participant performance. (Based on anecdotal observations, participants who were superior in task performance tended to give better ratings for the interface.) There were participants who rated the enhanced interface with a MCH-UVD value of 8 (mandatory redesign) and participants who rated the degraded interface to with a value of 1 (excellent design). As concluded by Donmez et al. (2010), the MCH-UVD may be predictive of participant task performance. However, the present study indicates that it is not a reliable tool for interface evaluation.

The Modified GEDIS-UAV provides detailed guidelines/criteria for evaluating interfaces and, therefore, eliminates subjectivity in design evaluation. In addition, the Modified GEDIS-UAV considers both usability and functionality features, which makes the evaluation more comprehensive. Since the tool consists of several "indicators" (i.e., interface features), it is also diagnostic in nature. Indicators with lower scores can be targeted for

redesign. For example, the degraded and baseline interfaces produced the lowest scores for "Map and Navigation" and "Alarms" indicators, as compared to the enhanced interface (Table 5.3). Redesign of these two components could improve the degraded and baseline interfaces to the greatest extent and should be considered first.

Table 5.3 Modified GEDIS-UAV Scores

	Enhanced interface	Baseline interface	Degraded interface
Display Layout (DL)	0.83	0.75	0.65
Information Presentation (IP)	0.91	0.86	0.69
Color (C)	0.90	0.90	0.79
Text (T)	0.93	0.93	0.93
Map and Navigation (MN)	0.84	0.55	0.26
Status and Devices (SD)	0.90	0.86	0.82
Data Entry Command (DEC)	0.63	0.63	0.63
Alarm (A)	0.85	0.62	0.33
Physical Control (PC)	0.92	0.92	0.92
Global Evaluation Index	0.86	0.78	0.67

While the MCH-UVD is completed by participants during interface evaluations, the Modified GEDIS-UAV is completed by human factors expert analysts. Therefore, the Modified GEDIS-UAV may serve as a complimentary interface analysis tool to a participant based approach or as a precursor to participant evaluations. Zhang et al. (in review) also demonstrated that the Modified GEDIS-UAV was sensitive to interface feature manipulations and had good reliability among expert analysts.

6. Conclusion

6.1 Research summary

In conclusion, the objectives of the present research were to investigate how UAV control interface design may impact operator cognitive workload in fundamental control operations, and to demonstrate use of physiological measures of cognitive workload in simulated UAV operations. To achieve these objectives, three UAV interface design variations (based on the ArduPilot MP interface) were prototyped, including baseline, enhanced and degraded usability designs. Several approaches were used to verify the usability and functionality differences among the interface designs, including the Modified GEDIS-UAV evaluation, cognitive modeling and usability heuristics analysis. These tools were also used to gain insight into the potential capability for the interfaces to effectively support UAV control operations. Subsequently, an experiment was conducted to assess participant cognitive workload during task performance with the different interface designs. Results revealed the enhanced interface to be superior to the baseline and degraded interfaces in facilitating UAV control tasks and moderating operator cognitive workload. Moreover, the enhanced interface was more robust in moderating operator effort under highly demanding task circumstances.

Regarding cognitive workload assessment methods, this study demonstrated application of a variety of physiological workload measures in UAV control operations as well as administration of subjective workload ratings. Results revealed consistent effects of experimental manipulations across eye-tracking measures (i.e., BD decrease ratio) and workload ratings (i.e., NASA-TLX). However, these measures were not significantly correlated. Cardiac responses measures, including HR and HRV, did not show sensitivity for

detecting differences in workload among the three interface designs under various levels of task demand. In general, eye-tracking measures appear to have utility for UAV operator workload assessment. The lack of sensitivity of cardiac responses might have been limited due to specific tasks or system as part of this study. However, their reliability still needs to be further explored.

The present study also aimed to assess the validity of the Modified GEDIS-UAV tool as a means of operator cognitive workload prediction. The interface design variations were primarily driven by the Modified GEDIS-UAV evaluation on original MP interface: the enhanced interface was designed to achieve the highest scores with the tool and the degraded interface was designed to pose violations to the guidelines as part of the evaluation tool. The effectiveness of the Modified GEDIS-UAV was supported by the fact that the enhanced interface resulted in the lowest cognitive workload for operators and showed robustness under highly demanding task scenarios. The validity of the interface evaluation tool was also supported by consistency of results with usability heuristics analysis and cognitive task performance modeling results.

6.2 Applications

The results of the present study can be applied to UAV interface design processes. In order to moderate operator workload in UAV operations, designers need to provide additional functionalities as part of control interfaces by implementing advanced automation, including technologies for automatic detection of abnormal vehicle status. Moreover, optimal visual information presentation should be pursued by improving interface usability features (e.g., color-coding and logical layout of information). The study also showed that the Modified

GEDIS-UAV is an effective tool for identifying UAV interface design deficiencies and providing guidance for redesign. Last but not least, the Modified GEDIS-UAV interface evaluation approach can be modified and applied to a variety of machine-system platforms (e.g., UGVs, self-driving cars, robots, etc.) to assess and improve human-automation interaction.

6.3 Limitations and future work

In this study, UAV interface design variations were based on the ArduPilot MP interface and only represented three types of design. However, UAV interfaces with different layouts, complexity and design styles exist in the market. In the future, a greater variety of interface designs that yield a broader range of Modified GEDIS-U
AV scores should be investigated for user cognitive workload implications. In addition to the use of semi-functional prototypes, actual UAV platforms should also be tested in order to promote the generalizability of research results.

One limitation of the Modified GEDIS-UAV is that this tool requires substantial training in advance of application. Some knowledge of human factors issues in systems design is necessary for accurate understanding the guidelines included in the evaluation tool. At present, there is also no guidance on how many analysts should apply the interface evaluation tool to ensure reliability of results. Beyond this, although the Modified GEDIS-UAV showed utility for identifying UAV control interface design deficiencies, a scoring criteria for “acceptable” interface design has yet to be established. Ponsa, Diaz (2007) and Lorite et al. (2010) have implied that a well-designed interface should reach a score of 4 out of 5 (i.e., 80% satisfaction of design guidelines) for the original GEDIS and GEDIS-UAV.

They also recommended that any interface design process should attempt to produce a score as close to 5 as possible (i.e., 100% satisfaction of design guidelines) by referring to the content of the evaluation tool. However, this threshold has not been supported by any empirical study. Future investigations need to be conducted on the relationship of Modified GEDIS-UAV scores with operator performance and workload. “Knee-points” (i.e., a significant decrease or increase) in performance and workload should be identified as criteria for labeling interface designs as “acceptable”. Such criteria could be helpful for regulating the UAV control interface design industry.

Another limitation of the present work is that all participants were recruited from a university campus and surrounding community. Participants were not familiar with any UAV control operations. As a result, their behavior may not be representative of professional UAV operators. In addition, participants received limited training (participants were only trained with one mission scenario and the training was less than 1 hour) as compared to actual UAV operators (hours to days of training). The lack of experience and limited practice of study participants might have also influenced results. For future studies, formally trained UAV operators are recommended as participants. If the general population is used, training with various mission scenarios and map features should be provided in order to build expertise and confidence for participants prior to testing.

Last but not least, the interface design variations tested in this study were presented as a between-subject factor in order to prevent carry-over effects and to limit participant confusion in interface use, as a result of switching among several interfaces across test trials. This approach led to some difficulty in comparing physiological responses among interface

groups. Future studies should give priority to a within-subject design in which physiological responses are considered primary workload indicators.

REFERENCES

Abich, J., Matthews, G., & Reinerman-Jones, L. (2015). Individual Differences in UGV Operation A Comparison of Subjective and Psychophysiological Predictors. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 59, pp. 741–745). SAGE Publications. Retrieved from http://pro.sagepub.com/content/59/1/741.short

Adams, J. A. (2002). Critical considerations for human-robot interface development. In *Proceedings of 2002 AAAI Fall Symposium* (pp. 1–8). Retrieved from http://www.aaai.org/Papers/Symposia/Fall/2002/FS-02-03/FS02-03-001.pdf

Afergan, D., Peck, E. M., Solovey, E. T., Jenkins, A., Hincks, S. W., Brown, E. T., … Jacob, R. J. (2014). Dynamic difficulty using brain metrics of workload. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 3797–3806). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2557230

Ahlstrom, U., & Friedman-Berg, F. J. (2006). Using eye movement activity as a correlate of cognitive workload. *International Journal of Industrial Ergonomics*, *36*(7), 623–636. https://doi.org/10.1016/j.ergon.2006.04.002

Ayaz, H., Shewokis, P. A., Bunce, S., Izzetoglu, K., Willems, B., & Onaral, B. (2012). Optical brain monitoring for operator training and mental workload assessment. *NeuroImage*, *59*(1), 36–47. https://doi.org/10.1016/j.neuroimage.2011.06.023

Backs, R. W., & Walrath, L. C. (1992). Eye movement and pupillary response indices of mental workload during visual search of symbolic displays. *Applied Ergonomics*, *23*(4), 243–254.

Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychological Bulletin*, *91*(2), 276–292. https://doi.org/10.1037/0033-2909.91.2.276

Beatty, J., & Lucero-Wagoner, B. (2000). The Pupillary System. In J. T. Cacioppo, L. G. Tassinary, & G. Berntson, *Handbook of Psychophysiology* (Vol. 2, pp. 142–162). Cambridge University Press.

Benedetto, S., Pedrotti, M., Minin, L., Baccino, T., Re, A., & Montanari, R. (2011). Driver workload and eye blink duration. *Transportation Research Part F: Traffic Psychology and Behaviour*, *14*(3), 199–208. https://doi.org/10.1016/j.trf.2010.12.001

Brodal, P. (2004). *The Central Nervous System: Structure and Function*. Oxford University Press, USA.

Brookings, J. B., Wilson, G. F., & Swain, C. R. (1996). Psychophysiological responses to changes in workload during simulated air traffic control. *Biological Psychology*, *42*(3), 361–377. https://doi.org/10.1016/0301-0511(95)05167-8

Brown, I. D. (1978). Dual task methods of assessing work-load. *Ergonomics*, *21*(3), 221–224.

Bruce, A. S., Atchison, D. A., & Bhoola, H. (1995). Accommodation-convergence relationships and age. *Investigative Ophthalmology & Visual Science*, *36*(2), 406–413.

Çakır, M. P., Şenyiğit, A. M., Akay, D. M., Ayaz, H., & İşler, V. (2012). Evaluation of UAS Camera Operator Interfaces in a Simulated Task Environment: An Optical Brain Imaging Approach. In *International Conference on Brain Inspired Cognitive Systems*

(pp. 62–71). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-31561-9_7

Card, S., Newell, A., & Moran, T. (1983). *The Psychology of Human-Computer Interaction*. L. Erlbaum Associates Inc. Retrieved from http://portal.acm.org/citation.cfm?id=578027

Chen, J. Y. C., Barnes, M. J., & Harper-Sciarini, M. (2011). Supervisory Control of Multiple Robots: Human-Performance Issues and User-Interface Design. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *41*(4), 435–454. https://doi.org/10.1109/TSMCC.2010.2056682

Crescenzio, F. D., Miranda, G., Persiani, F., & Bombardi, T. (2009). A first implementation of an advanced 3d interface to control and supervise uav (uninhabited aerial vehicles) missions. *Presence: Teleoperators and Virtual Environments*, *18*(3), 171–184.

Cummings, M. L., Myers, K., & Scott, S. D. (2006). Modified Cooper Harper evaluation tool for unmanned vehicle displays. In *Proceedings of UVS Canada: Conference on Unmanned Vehicle Systems Canada*. Retrieved from http://www.eng.uwaterloo.ca/~s9scott/wiki/uploads/Main/cummings_uvs2006.pdf

Cummings, M. L., Pina, P. E., & Donmez, B. (2008). *Selecting metrics to evaluate human supervisory control applications*. MIT Humans and Automation Laboratory. Retrieved from http://dspace.mit.edu/handle/1721.1/46743

De Joux, N., Russell, P. N., & Helton, W. S. (2013). A functional near-infrared spectroscopy study of sustained attention to local and global target features. *Brain and Cognition*, *81*(3), 370–375. https://doi.org/10.1016/j.bandc.2012.12.003

Demir, K. A., Cicibas, H., & Arica, N. (2015). Unmanned Aerial Vehicle Domain: Areas of Research. *Defence Science Journal*, *65*(4), 319–329.

Dix, A. (2009). *Human-computer interaction*. Springer. Retrieved from http://link.springer.com/10.1007/978-0-387-39940-9_192

Dixon, S. R., & Wickens, C. D. (2003). *Control of multiple-UAVs: A workload analysis*. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA446844

Donmez, B., Cummings, M. L., Graham, H. D., & Brzezinski, A. S. (2010). Modified Cooper Harper scales for assessing unmanned vehicle displays. In *Proceedings of the 10th Performance Metrics for Intelligent Systems Workshop* (pp. 235–242). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2377620

Drury, J. L., Scholtz, J., & Kieras, D. (2007). Adapting GOMS to Model Human-robot Interaction. In *Proceedings of the ACM/IEEE International Conference on Human-robot Interaction* (pp. 41–48). New York, NY, USA: ACM. https://doi.org/10.1145/1228716.1228723

Eckstein, M. K., Guerra-Carrillo, B., Miller Singley, A. T., & Bunge, S. A. (2016). Beyond eye gaze: What else can eyetracking reveal about cognition and cognitive development? *Developmental Cognitive Neuroscience*. https://doi.org/10.1016/j.dcn.2016.11.001

Fahlstrom, P., & Gleason, T. (2012). *Introduction to UAV systems*. John Wiley & Sons. Retrieved from

https://books.google.com/books?hl=en&lr=&id=uLsNtm99IWYC&oi=fnd&pg=PT9&dq=Introduction+to+UAV+systems&ots=lPKTzbaXLA&sig=L2tkHoePRw_UJh0AWS-NIYBsjtI

Fong, T., & Thorpe, C. (2001). Vehicle teleoperation interfaces. *Autonomous Robots*, *11*(1), 9–18.

Foyle, D. C., Hooey, B. L., Byrne, M. D., Corker, K. M., Deutsch, S., Lebiere, C., … Wickens, C. D. (2005). Human Performance Models of Pilot Behavior. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *49*(12), 1109–1113. https://doi.org/10.1177/154193120504901202

Fuchs, C., Borst, C., de Croon, G. C., van Paassen, M. R., & Mulder, M. (2014). An ecological approach to the supervisory control of UAV swarms. *International Journal of Micro Air Vehicles*, *6*(4), 211–229.

Gabriel, G., Ramallo, M. A., & Cervantes, E. (2016). Workload perception in drone flight training simulators. *Computers in Human Behavior*, *64*, 449–454.

Galster, S. M., Knott, B. A., & Brown, R. D. (2006). Managing multiple UAVs: Are we asking the right questions? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, pp. 545–549). SAGE Publications. Retrieved from http://pro.sagepub.com/content/50/3/545.short

Grant, R. C., Carswell, C. M., Lio, C. H., & Seales, W. B. (2013). Measuring surgeons' mental workload with a time-based secondary task. *Ergonomics in Design: The Quarterly of Human Factors Applications*, *21*(1), 7–11.

Gratton, G., & Fabiani, M. (2006). Optical Imagiing of Brain Function. In R. P. P. of P. and D. of the C. S. L. T. C. U. of America & D. of the D. of N. M. R. P. of N. Iowa and Professor of Industrial Engineering and Public Policy The University of, *Neuroergonomics : The Brain at Work: The Brain at Work* (pp. 65–81). Oxford University Press, USA.

Gupta, S. G., Ghonge, M. M., & Jawandhiya, P. M. (2013). Review of unmanned aircraft system (UAS). *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, *2*(4). Retrieved from https://www.researchgate.net/profile/Mangesh_Ghonge/publication/249998592_Review_of_Unmanned_Aircraft_System_(UAS)/links/02e7e51e8ef1668ce8000000.pdf

Harriott, C. E., Buford, G. L., Zhang, T., & Adams, J. A. (2015). Mental workload and task performance in peer-based human-robot teams. *Journal of Human-Robot Interaction*, *4*(2), 61–96.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. H. and N. Meshkati (Ed.), *Advances in Psychology* (Vol. 52, pp. 139–183). North-Holland. Retrieved from http://www.sciencedirect.com/science/article/pii/S0166411508623869

Helander, M. G. (2014). *Handbook of Human-Computer Interaction*. Elsevier.

Hess, E. H., & Polt, J. M. (1964). Pupil Size in Relation to Mental Activity during Simple Problem-Solving. *Science*, *143*(3611), 1190–1192. https://doi.org/10.1126/science.143.3611.1190

Higgins, J. (2003). *Introduction to Modern Nonparametric Statistics*. Cengage Learning. Retrieved from http://www.amazon.ca/exec/obidos/redirect?tag=citeulike09-20&path=ASIN/0534387756

Hill, S. G., Iavecchia, H. P., Byers, J. C., Bittner, A. C., Zaklade, A. L., & Christ, R. E. (1992). Comparison of Four Subjective Workload Rating Scales. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(4), 429–439. https://doi.org/10.1177/001872089203400405

Hockey, G. R. J., Briner, R. B., Tattersall, A. J., & Wiethoff, M. (1989). Assessing the impact of computer workload on operator stress: the role of system controllability. *Ergonomics*, *32*(11), 1401–1418. https://doi.org/10.1080/00140138908966914

Hollnagel, E. (2003). *Handbook of cognitive task design*. CRC Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=dElPH0ruR-sC&oi=fnd&pg=PP1&dq=Handbook+of+Cognitive+Task+Design&ots=EGmVq44LDG&sig=JDYSbTWmGljZsYB30icGrCtQC5E

Humphrey, C. M., & Adams, J. A. (2011). Analysis of complex team-based systems: augmentations to goal-directed task analysis and cognitive work analysis. *Theoretical Issues in Ergonomics Science*, *12*(2), 149–175.

Hwang, S.-L., Yau, Y.-J., Lin, Y.-T., Chen, J.-H., Huang, T.-H., Yenn, T.-C., & Hsu, C.-C. (2008). Predicting work performance in nuclear power plants. *Safety Science*, *46*(7), 1115–1124. https://doi.org/10.1016/j.ssci.2007.06.005

Irizarry, J., Gheisari, M., & Walker, B. N. (2012). Usability assessment of drone technology as safety inspection tools. *Journal of Information Technology in Construction, 17*, 194–212.

Izzetoglu, K., Bunce, S., Onaral, B., Pourrezaei, K., & Chance, B. (2004). Functional Optical Brain Imaging Using Near-Infrared During Cognitive Tasks. *International Journal of Human–Computer Interaction, 17*(2), 211–227. https://doi.org/10.1207/s15327590ijhc1702_6

John, B. E., & Kieras, D. E. (1996). The GOMS Family of User Interface Analysis Techniques: Comparison and Contrast. *ACM Trans. Comput.-Hum. Interact., 3*(4), 320–351. https://doi.org/10.1145/235833.236054

Jorna, P. G. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological Psychology, 34*(2), 237–257.

Kaber, D. B., Kim, S. H., & Wang, X. (2011). Computational Cognitive Modeling of Human-Robot Interaction Using a GOMS Methodology. In X. Wang (Ed.), *Mixed Reality and Human-Robot Interaction* (pp. 53–75). Springer Netherlands. https://doi.org/10.1007/978-94-007-0582-1_4

Kaber, David B., Riley, J. M., & Tan, K.-W. (2002). Improved usability of aviation automation through direct manipulation and graphical user interface design. *The International Journal of Aviation Psychology, 12*(2), 153–178.

Kaber, Segall, N., Green, R. S., Entzian, K., & Junginger, S. (2006). Using multiple cognitive task analysis methods for supervisory control interface design in high-throughput biological screening processes. *Cognition, Technology & Work, 8*(4), 237–252.

Kieras, D. (1994). A guide to GOMS task analysis. *Unpublished Manuscript, University of*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.51.109&rep=rep1&type=pdf

Kieras, D. (1999). A guide to GOMS model usability evaluation using GOMSL and GLEAN3. *University of Michigan*, (313). Retrieved from http://acs.ist.psu.edu/ist331/GOMSL_Guide.pdf

Kramer, A. F. (1990). *Physiological Metrics of Mental Workload: A Review of Recent Progress* (Interim No. NPRDC-TN-90-23). Navy Personnel Research and Development Center.

Kramer, A. F., Trejo, L. J., & Humphrey, D. G. (1996). Psychophysiological measures of workload: Potential applications to adaptively automated systems. *Automation and Human Performance: Theory and Applications*, 137–162.

Lerch, F. J., Mantei, M. M., & Olson, J. R. (1989). Skilled Financial Planning: The Cost of Translating Ideas into Action. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 121–126). New York, NY, USA: ACM. https://doi.org/10.1145/67449.67474

Liu, D., Peterson, T., Vincenzi, D., & Doherty, S. (2016). Effect of time pressure and target uncertainty on human operator performance and workload for autonomous unmanned aerial system. *International Journal of Industrial Ergonomics, 51*, 52–58. https://doi.org/10.1016/j.ergon.2015.01.010

Lorite, S., Muñoz, A., Tornero, J., Ponsa, P., & Pastor, E. (2013). Supervisory control interface design for unmanned aerial vehicles through GEDIS-UAV. In *International Conference on Human-Computer Interaction* (pp. 231–240). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-39232-0_26

Lu, J.-L., Horng, R.-Y., & CHAO, C.-J. (2013). DESIGN AND TEST OF A SITUATION-AUGMENTED DISPLAY FOR AN UNMANNED AERIAL VEHICLE MONITORING TASK 1. *Perceptual & Motor Skills*, *117*(1). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=00315125&AN=92036019&h=cK9JOtrvTO6ijRgAYVGJAPUh3ytj0RVluePTWRZ6Abt6O85DdttqGA0wLDq%2FvlkPxwQg02QJ8wO%2FF5Qyn1bRsw%3D%3D&crl=c

Marshall, S. P. (2002). The Index of Cognitive Activity: measuring cognitive workload. In *Proceedings of the 2002 IEEE 7th Conference on Human Factors and Power Plants, 2002* (pp. 7-5-7–9). https://doi.org/10.1109/HFPP.2002.1042860

Matthews, G., Reinerman-Jones, L. E., Barber, D. J., & Abich, J. (2015). The Psychometrics of Mental Workload Multiple Measures Are Sensitive but Divergent. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(1), 125–143. https://doi.org/10.1177/0018720814539505

Maza, I., Caballero, F., Molina, R., Peña, N., & Ollero, A. (2010). Multimodal interface technologies for UAV ground control stations. *Journal of Intelligent and Robotic Systems*, *57*(1–4), 371–391.

Menda, J., Hing, J. T., Ayaz, H., Shewokis, P. A., Izzetoglu, K., Onaral, B., & Oh, P. (2011). Optical brain imaging to enhance UAV operator training, evaluation, and interface development. *Journal of Intelligent & Robotic Systems*, *61*(1–4), 423–443.

Miller, C. A., & Parasuraman, R. (2007). Designing for Flexible Interaction Between Humans and Automation: Delegation Interfaces for Supervisory Control. *Human Factors*, *49*(1), 57–75. https://doi.org/10.1518/001872007779598037

Molich, R., & Nielsen, J. (1990). Improving a Human-computer Dialogue. *Commun. ACM*, *33*(3), 338–348. https://doi.org/10.1145/77481.77486

Montgomery, D. C. (2017). *Design and Analysis of Experiments*. John Wiley & Sons.

Nielsen, C. W., & Goodrich, M. A. (2006). Comparing the usefulness of video and map information in navigation tasks. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 95–101). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1121259

Nielsen, J. (1994). *Usability Engineering*. Elsevier.

Nielsen, J. (1996). Usability metrics: Tracking interface improvements. *IEEE Software; Los Alamitos*, *13*(6), 12–13. https://doi.org/http://dx.doi.org/10.1109/MS.1996.10031

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 249–256). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=97281

Norman, D. A. (2013). *The Design of Everyday Things: Revised and Expanded Edition*. Basic Books.

Nygren, T. E. (1991). Psychometric Properties of Subjective Workload Measurement Techniques: Implications for Their Use in the Assessment of Perceived Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *33*(1), 17–33. https://doi.org/10.1177/001872089103300102

Olson, W. A., & Wuennenberg, M. G. (2001). Autonomy based human-vehicle interface standards for remotely operated aircraft. In *Digital Avionics Systems, 2001. DASC. 20th Conference* (Vol. 2, p. 7D3–1). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=964199

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating Cognitive Load Using Remote Eye Tracking in a Driving Simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 141–144). New York, NY, USA: ACM. https://doi.org/10.1145/1743666.1743701

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354

Peshkova, E., Hitz, M., & Kaufmann, B. (2017). Natural Interaction Techniques for an Unmanned Aerial Vehicle System. *IEEE Pervasive Computing*, *16*(1), 34–42. https://doi.org/10.1109/MPRV.2017.3

Ponsa Asensio, P., Amante García, B., & Díaz Boladeras, M. (2010). Ergonomic design applied in a sugar mill interface. *Latin American Applied Research*, *40*(1), 27–33.

Ponsa, P., & Díaz, M. (2007). Creation of an ergonomic guideline for supervisory control interface design. In *International Conference on Engineering Psychology and Cognitive Ergonomics* (pp. 137–146). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-73331-7_15

Prewett, M. S., Johnson, R. C., Saboe, K. N., Elliott, L. R., & Coovert, M. D. (2010). Managing workload in human–robot interaction: A review of empirical studies. *Computers in Human Behavior*, *26*(5), 840–856.

Putze, F., & Schultz, T. (2014). Adaptive cognitive technical systems. *Journal of Neuroscience Methods*, *234*, 108–115. https://doi.org/10.1016/j.jneumeth.2014.06.029

Reid, G. B., & Nygren, T. E. (1988). The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. In P. A. H. and N. Meshkati (Ed.), *Advances in Psychology* (Vol. 52, pp. 185–218). North-Holland. Retrieved from http://www.sciencedirect.com/science/article/pii/S0166411508623870

Reiner, M., & Gelfeld, T. M. (2014). Estimating mental workload through event-related fluctuations of pupil area during a task in a virtual world. *International Journal of Psychophysiology*, *93*(1), 38–44. https://doi.org/10.1016/j.ijpsycho.2013.11.002

Roscoe, A. H. (1992). Assessing pilot workload. Why measure heart rate, HRV and respiration? *Biological Psychology*, *34*(2), 259–287.

Ruff, H. A., Narayanan, S., & Draper, M. H. (2002). Human Interaction with Levels of Automation and Decision-Aid Fidelity in the Supervisory Control of Multiple

Simulated Unmanned Air Vehicles. *Presence: Teleoperators and Virtual Environments, 11*(4), 335–351. https://doi.org/10.1162/105474602760204264

Schraagen, J. M., Chipman, S. F., & Shalin, V. L. (2000). *Cognitive task analysis*. Psychology Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=gyt5AgAAQBAJ&oi=fnd&pg=PP1&dq=Cognitive+Task+Analysis&ots=cfI2wxN5ta&sig=L9Lo_HzsdCO9sfuId5MtjrfT5p8

Schulz, C. M., Schneider, E., Fritz, L., Vockeroth, J., Hapfelmeier, A., Wasmaier, M., … 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. https://doi.org/10.1093/bja/aeq307

Scribner, D. R., & Gombash, J. W. (1998). *The Effect of Stereoscopic and Wide Field of View Conditions on Teleoperator Performance.* DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA341218

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3/4), 591–611.

Shneiderman, B., & Plaisant, C. (1987). Designing the user interface: Strategies for effective human-computer interaction. *ACM SIGBIO Newsletter*, *9*(1), 6.

Snedecor, G. W., & Cochran, W. G. (1989). *Statistical Methods* (8 edition). Ames: Iowa State University Press.

Song, J., Zhuang, D., Song, G., & Wanyan, X. (2011). Pilot mental workload measurement and evaluation under dual task. In *2011 4th International Conference on Biomedical Engineering and Informatics (BMEI)* (Vol. 2, pp. 809–812). https://doi.org/10.1109/BMEI.2011.6098376

Squire, P., Trafton, G., & Parasuraman, R. (2006). Human control of multiple unmanned vehicles: effects of interface type on execution and task switching times. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 26–32). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1121248

Swangnetr, M., & Kaber, D. B. (2013). Emotional state classification in patient–robot interaction using wavelet analysis and statistics-based feature selection. *IEEE Transactions on Human-Machine Systems, 43*(1), 63–75.

Swangnetr, M., Kaber, D. B., Vorberga, E., Fleischera, H., & Thurowa, K. (2014). Identifying automation opportunities in life science processes through operator task modeling and workload assessment. In K. Stanney & K. S. Hale, *Advances in Cognitive Engineering and Neuroergonomics* (pp. 215–222). AHFE Conference.

Tarvainen, M. P., Niskanen, J.-P., Lipponen, J. A., Ranta-aho, P. O., & Karjalainen, P. A. (2014). Kubios HRV – Heart rate variability analysis software. *Computer Methods and Programs in Biomedicine, 113*(1), 210–220. https://doi.org/10.1016/j.cmpb.2013.07.024

Task Force of the European Society of Cardiology. (1996). Heart rate variability standards of measurement, physiological interpretation, and clinical use. *Eur Heart J*, *17*, 354–381.

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Wierwille, W. W., & Eggemeier, F. T. (1993). Recommendations for mental workload measurement in a test and evaluation environment. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *35*(2), 263–281.

Williams, K. W. (2004). *A summary of unmanned aircraft accident/incident data: Human factors implications*. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA460102

Williams, K. W. (2007). *An assessment of pilot control interfaces for unmanned aircraft*. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA465657

Yesilbas, V., & Cotter, T. S. (2014). Structural Analysis of Hfacs in Unmanned and Manned Air Vehicles. In *Proceedings of the International Annual Conference of the American Society for Engineering Management.; Huntsville* (pp. 1–7). Huntsville: American Society for Engineering Management (ASEM). Retrieved from http://search.proquest.com/docview/1705171380/abstract/5BCAB252E0054797PQ/1

Young, M. S., & Stanton, N. A. (2004). Mental Workload. In N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, & H. W. Hendrick, *Handbook of Human Factors and Ergonomics Methods* (p. 39–1 to 199). CRC Press.

Zhang, W., Feltner, D., Shirley, J., & Kaber, D. (2016). Unmanned Aerial Vehicle Control Interface Design and Cognitive Workload: A Constrained Review and Research Framework. Presented at the Systems, Man, and Cybernetics, IEEE International Conference, Budapest, Hungary.

Zhang, W., Feltner, D., Shirley, J., Kaber, D., & Swangnetr, M. (Submitted). Enhancement and application of an UAV supervisory control interface evaluation technique: Modified GEDIS-UAV. *International Journal of Human-Computer Studies*.

Zhang, W., Ma, W., Brandao, M., Kaber, D. B., Bloomfield, P., & Swangnetr, M. (2016). Biometric validation of a virtual reality-based psychomotor test for motor skill training. *Assistive Technology*, *28*(4), 233–241.

Zhang, W., White, M., Zahabi, M., Winslow, A. T., Zhang, F., Huang, H., & Kaber, D. (2016). Cognitive workload in conventional direct control vs. pattern recognition control of an upper-limb prosthesis. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 002335–002340). https://doi.org/10.1109/SMC.2016.7844587

Zhao, G., & Wu, C. (2013). The Effects of Image Set Size and Time Pressure on Human Performance and Mental Workload in a Cyber Image Search Task. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *57*(1), 2066–2070. https://doi.org/10.1177/1541931213571461

APPENDICES

Appendix A

Demographic Questionnaire

Participant #:_____

(1) Please circle your gender: Male Female Prefer not to answer

(2) What is your age:

(3) What is your current corrected vision:

(4) Do you have full color vision?

(5) To what extent do you use a computer in daily life?

Very little Occasionally Frequently Extensively

(6) Rate your video gaming experience

None Very little Some Moderate Expert

(7) Rate your manned flight experience

None Very little Some Moderate Expert

Hours: _____ Type Rating: _____

(8) Rate your flight simulation experience

None Very little Some Moderate Expert

Appendix B

Usability Questionnaire

Participant #:_____

Please indicate if you agree or disagree with each of the following statement.

	Strongly				Strongly agree
	1	2	3	4	5
Overall, the UAV control interface was easy to use.					
It was easy to find information I needed.					
The interface was effective in helping me completing the tasks.					
The UAV control tasks were easy to accomplish.					

What comments/complaints do you have about the system interface?

What recommendations for improvement do you have for the interface?

Appendix C

North Carolina State University
INFORMED CONSENT FORM for RESEARCH

Title of Study: Cognitive Workload Effects of Unmanned Aerial Vehicle interface Design
Principal Investigator: Wenjuan Zhang, David Feltner Faculty Sponsor: Dr. David Kaber

What are some general things you should know about research studies?

You are being asked to take part in a research study. Your participation in this study is voluntary. You have the right to be a part of this study, to choose not to participate or to stop participating at any time without penalty. The purpose of this research is to gain a better understanding of how unmanned aerial vehicle interface design impact user workload and performance. You are not guaranteed any personal benefits from being in this study. The study only poses minimal risks to those that participate. In this consent form you will find specific details about the research. If you do not understand something in this form it is your right to ask the researcher for clarification or more information. A copy of this consent form will be provided to you. If at any time you have questions about your participation, do not hesitate to contact the researcher(s) named above.

What is the purpose of this study?

The purpose of this study is to investigate Unmanned Aerial Vehicle (UAV) supervisory control interface design.

What are the inclusion criteria of this study?

To be eligible for this study, you need to (1) be between 18-64 years of age; (2) have 20/20 or corrected vision and no color-impairment due to the study requirement for visual perception of colored items/text on a display screen; (2) have general familiarity with use of computers as the study involves using an interface simulation presented on a desktop/laptop computer; and (3) have no experience with UAV supervisory control operations (so as not to bias performance in control tasks).**What will happen if you take part in the study?**

If you agree to participate in this study, you will be asked to:

1. Complete a demographic questionnaire requesting information about your age, gender, eye sight, computer usage, and any experience in UAV supervisory control.
2. Participate in a brief training session to learn some basic UAV operations with a simulation, for example, launching a UAV and changing flight parameters.
3. Participate in calibration of an eye-tracking device so that your eye movements (e.g., pupil size, blink rate) can be recorded during experiment trials.
4. Don a chest band so that your heart rate can be monitored during experiment trials.
5. Perform training and 4 test trials of simulated UAV operation.
6. You will be asked to complete a short survey and be provided time to rest after each trial.

These steps will take place in the Human Factors and Ergonomics Lab (Daniels Hall, Room 457). In total, the experiment is expected to take approximately 2 hours of your time.

7. Risks

You may experience eyestrain during the interaction with computer interfaces. However, you will be provided with rest after each trial. You may also experience slight discomfort with a chest strap as part of a heart rate monitoring system.. There are no known risks associated with the eye-tracker. Overall, the risks in the experiment are minimal.

8. Benefits

The results of this research are expected to be beneficial for UAV interface design. There is no direct benefit to you as a result of participation in this experiment.

9. Confidentiality

The information in the study records will be kept confidential to the full extent allowed by law. Data will be stored securely on the hard drives of a laboratory computer or researcher computers. In addition, video recordings will be of your use of control interface in all trials along with computer screen captures. Neither your face nor any other distinguishing physical features will be captured in recordings. Videos will be destroyed at the conclusion of the study. No reference will be made in oral or written reports that could link you to the study.

Compensation

For participating in this study you will receive payment at the rate of $15 per hour. If you withdraw from the study prior to its completion, you will be paid for the amount of time you spent the experiment.

What if you are a NCSU student?

Your performance in this study will not affect your class standing or grades at NC State.

What if you are a NCSU employee?

Participation in this study is not a requirement of your employment at NCSU and your participation, or lack thereof, will not affect your job.

10. What if you have questions about this study?

If you have questions at any time about the study or the procedures, you may contact the researcher, Wenjuan Zhang, at wzhang28@ncsu.edu.

What if you have questions about your rights as a research participant?

If you feel you have not been treated according to the descriptions in this form, or your rights as a participant in research have been violated during the course of this project, you may

contact Deb Paxton, Regulatory Compliance Administrator, Box 7514, NCSU Campus (919/515-4514).

Prior Knowledge of the Study and its Goals

If you have prior knowledge of the experiment and any goals, expected results, or other information that may affect the integrity of the experiment (e.g., based on conversations with other participants or experimenters), please let the experimenter know before signing below.

11. Consent to Participate

"I have read and understand the above information. I have received a copy of this form. I agree to participate in this study with the understanding that I may choose not to participate or to stop participating at any time without penalty or loss of benefits to which I am otherwise entitled."

Subject's signature_____ **Date** _____

Investigator's signature_____ **Date** _____

Appendix D

Experiment Instructions (Baseline)

(Note: [ ] indicate required actions by an experimenter.)

(Note: An experimenter needs to read to participants the text in *italics*, as presented below.)

1. **Checklist of materials prior to experiment**

   Forms:

   - Consent forms
   - Payment forms (2 copies)
   - Demographic survey
   - TLX ranking form
   - TLX rating forms (4 copies)
   - Training materials
     - Familiarization mission, maneuver, status and alarm documents
     - Training mission, maneuver, status and alarm documents
     - Timeline
     - Gradesheet
   - Testing materials
     - Scenario 1 and Scenario 2 mission, maneuver, status and alarm documents
     - Prototypes 5-8 timelines
     - Prototypes 5-8 Gradesheet
   - Closing questionnaire (Usability + MCH-UVD)

   Equipment:

   - Desktop computer with extra monitor
   - Eye-tracking cameras
   - Interface simulations open and minimized
   - Hide desktop tool bar
   - Screen capture software and microphone working and ready for use
   - HR monitor is working and ready for use
   - Paper sheet protector
   - Stop watch
   - Pens
   - Door sign “Experiment in Progress – Do Not Disturb” is ready
   - Audio Recordings (phone on airplane mode)

2. **Orientation**

   a. **Introduction**

   [Record the time at which the participant arrives]

Thank you for agreeing to participate in this experiment. The objective of this experiment is to assess how Unmanned Aerial Vehicles (UAV) control interface design features may impact operator performance and cognitive workload in fundamental control operations. You will complete a training session and 4 testing trials in total. The study will last approximately 2 hours, for which you will be compensated at a rate of $15 per hour. Your heart rate and eye movement will be recorded during the experiment. A screen capture software will record your performance in using a testing computer and your audio interaction with the experimenter, but every step will be taken to preserve your anonymity in the recordings. Once the study is complete, all video and audio recordings will be destroyed. At this time, I ask that you turn off your cell phone or any other electronic device that may be a distraction to you during the experiment.

b. Informed Consent form

[Sit participant. Present both copies of the informed consent form and a pen.]

This is an informed consent form. It summarizes everything you need to know about the experiment, including your compensation, any potential risks, and your rights as an experiment participant. Please take the time to read the form carefully. Please inform an experimenter of any questions you might have. If you consent to participate, please sign and date both copies of the form. One form will be for your records and one for our records.

[Allow the participant time to read and sign the form.]

c. Demographic Questionnaire (DQ)

Now we ask that you complete a questionnaire requesting general information about your background. Please answer all questions as accurately as possible. As stated in the informed consent form, all of your answers will be kept confidential and none of this information will be published in any form that might reveal your identity.

[Allow the participant time to complete the DQ.]

3. Eye-tracking calibration

Now we need you to assist in the eye-tracking equipment calibration. Please sit in front of the monitors. The location and position of the monitors are very important for our measurement. Please do not touch or adjust the monitors. However, feel free to adjust your chair position and make sure you are comfortable viewing the screens, and using the mouse and keyboard. I will mark the position of your chair once you have made adjustments. You will be required to maintain the same body position and posture during all experiment trials.

[Sit participant in front of simulation]

a. **Open FaceLab from the computer desktop.**
b. **Click "Recalibrate" from the bottom-left window. Click through the wizard and follow the instructions.**
c. **Click "New head model" → "Manual". Click through the wizard and follow the instructions.**
d. **Attach calibration sheets to the monitor on the left. Click on the left plane and then click "show SID". Instruct the participant to look at the calibration dots in sequence.**
e. **Repeat Step (d) for the monitor on the right.**

4. **Familiarization**

[Ensure that the sign on the door indicates experiment in progress.]
[Sit the participant in front of the simulation]
Now that the calibration is complete. We will proceed to the training session.

a. Interface Familiarization

The monitor in front of you presents the UAV control interface that you will use during this experiment. This interface has a few components including a primary flight status display (PFD) in the upper left corner. I will now go through each component one by one and provide information for your use. Please follow my instructions to interact with the interface and do not click unnecessary buttons. However, please feel free to ask any questions.

(1) *UAV control buttons: Below the PFD are several UAV control buttons. Not* every button is *active, but all buttons that you need to use to accomplish tasks can be selected.*

- *"Arm/Disarm" button. This button allows you to arm or disarm the vehicle for flight. (The button does not refer to weapon system use.) It is necessary to arm the UAV before you launch it. Otherwise, the UAV will not launch. Now, please left click the Arm/Disarm button.* ***[Wait for participant to click Arm/Disarm button]*** *The UAV is now prepared to launch.*
- *"Launch UAV" button: Once you have successfully armed the UAV, please click "Launch UAV" and an UAV icon will appear on the Navigation Display at the launch point (LP) and start on its course.*
- *"RTL" Button: RTL stands for Return to Launch Point. This command takes the UAV back to the LP. You may need to do this at a designated time in the mission. An order to RTL will come from your headquarters; an audio message during a mission.*

  [Instruct participant to click RTL one UAV reaches WP3]
- *Any questions on the UAV control buttons?*

(2) *Quick display: Click the Quick tab. During the mission, you need to monitor the status of vehicle air speed, distance traveled, ground course, altitude, battery remaining, and ground speed. All of these parameters are displayed under the quick tab, and they will change during the UAV's flight. Under some parameter headings, you may see a "Norm" or "Acceptable" range. These displays identify the normal or acceptable range of parameters. All ranges are summarized in this table.* **[Point out Norms on paper]** *I will also talk you through all the information presented under the Quick tab.*

- *Air Speed: Now, please look at the top left number in purple; this is air speed indicator. Air Speed is how fast the UAV is flying through the air. It is measured in knots. Under the indicator, you can see "Norm [45,55]". This means that it is normal for this UAV to fly between 45 and 55 knots.*
- *Distance Traveled: Look at the top right number in orange; this is the distance in meters that the UAV has traveled during the mission. It does not accumulate from previous missions. Under this indicator, you can see "Target < 10,000". This means that your distance traveled should be less than 10,000 meters for a mission.*
- *Ground Course: Look at the middle number on the left side in red; this is the ground course, which is measured in degrees. Its value can range between 0 to 359 degrees. You do not need to worry about how this number is obtained. You only need to monitor the status during the flight.*
- *Altitude: Look at the middle number on the right-hand side; this is the UAV's altitude, which is measured in feet. The normal parameter range is from 40 to 55 feet, as shown below the altitude indicator.*
- *Battery Remaining: Look at the bottom left number in yellow; this is the UAV's battery life measured in percent remaining. The acceptable range for Battery Remaining is from 20 to 100 percent.*
- *Ground Speed: Look at the bottom right number in blue; this is the UAV's ground speed measured in meters per second. The normal parameter range is from 0 to 10.*
- *Parameter deviation: When any of the above parameters deviate outside the normal or acceptable range, please immediately notify an experimenter by saying "Warning" then the parameter that is out of tolerance. For example, you may say "Warning: Altitude" or "Warning: Battery remaining". Now please identify all deviations from the current display.* **[If any error, provide correct answer and explanation]**
- *Any questions on the quick display?*

(3) *PFD: On the top left of the screen, you see the PFD. It provides some vehicle status information, including ground course, air speed, and altitude. The display is not continuous and updates about every 10-15 seconds. The ground course is shown by*

the bar at the top, the black arrow will point to the current ground course which will match the quick display number. On the left side is a grey box labeled AS for Air Speed; the black arrow will point to the vehicle's current air speed, which will match the quick display number. On the right side is a grey box labeled ALT for Altitude; the black arrow will point to the vehicle's current altitude, which will match the quick display number.

(4) *Mission Planning tool: At the bottom right of the screen you will see a listing of flight waypoints (WPs)and some of their parameters. During the mission, you may be asked to make adjustments to the UAV's flight path. For example, you may need to change the altitude for WP2. To do this, simply find the Altitude column and the corresponding Waypoint row. Click once in the box for the input field to show up, and then click again to type in the new altitude.* **[Wait for participant to click.]** *Now you can type in the desired altitude in feet, for example, 99. After that, you need to click the "Write WP" button on the right-hand side of the screen* **[Point at the button]** *to confirm the change.* **[Wait for participant to click.]** *Once you click the button, a message window will show up to confirm your change. The window will disappear by itself. Now the updated altitude can be seen in the table.*

(5) *Navigation display: The top-right portion of the interface presents a navigation display.*
   - *Northing & Easting: On the top and left sides of the map, you see a few boxes with numbers. They are called Eastings and Northings. You can use them to determine distances and locations on the map. I will explain how to read them in just a few minutes.*
   - *Launch Point (LP): The red circle is the home station for the UAV, and it is where you will launch from and land. You may be responsible for determining the vehicle coordinates, or distance between it and another object, at any time during a flight. The center point of the circle is where all measurements are taken from.*
   - *Waypoints (WPs): The red icons with a pointy tip are waypoints for this mission. The vehicle will fly through the waypoints in numerical order. You may be responsible for determining WP coordinates, or the distance between two WPs. The bottom tip of a WP is where all measurements should be taken from. Waypoints also provide an indicator of the degree of mission completion. For example, you now see 3 waypoints on the screen now (the launch point does not count). If you are past WP1 but before WP2, then you are 33% complete with the mission.*
   - *Targets: The orange triangles represent targets on the map. You may be asked to count targets, obtain their coordinates or determine the distance between two targets. The center point is where all measurements are taken from.*

- *Any questions on the navigation display?*

(6) *Map action button: On the top right corner of the navigation display, you see a yellow button called "Map Actions". Please click on the Map Action button.* **[Wait for participant to click]** *This menu provides you a way to interact with the UAV. Not every option in this menu is active, but all buttons that you need to use to accomplish tasks can be selected. I will highlight every option that is active. All others are inactive.*

[Make sure the participant is able to locate the menu options. Demonstrate if necessary]

- *Waypoint (WP): "WP" is used as an acronym for waypoint. There are four options related to waypoints in the menu: Insert, Load, Edit, and Delete. These buttons are not active but you may be asked about them.*
- *Drop Payload: The "drop payload" button is active and allows you to drop your payload at a designated location. Payload is the cargo that the UAV is carrying that you must drop at a specific location. During a mission, your headquarters may require you to drop a payload at WP2. To do this, all you have to do is click the menu option, as soon as possible, when you receive the order. Now please click "Drop Payload". Once you click the button, a payload icon will appear at WP2. The menu is currently blocking the dropping location. Please click the "Map Action" button to hide the menu. Now you can see the payload icon. The dropping location has been preprogrammed. Optimally, you should drop the payload within 3 seconds after receiving the order. Please click "Map Action" and show the menu again. We will continue with other menu options.*
- *Loiter: There are three options for loiter: Forever, Time and Circles. These buttons are not active but you may be asked about them.*
- *Jump To: There are three options for "Jump To": Start, WP #, and LP.*
- *Overlays: There are three options for "Overlay": Create, Edit, and Delete.*
- *Draw: There are three options for "Draw": Line, Polygon, and Route.*
- *Commands: There are 5 options of "Commands" in the menu: Take off, Altitude, Speed, Land and RTL. Only RTL is an active option.*
- *Clear Mission: There is only one option related to "Clear Mission". You can find it at the bottom of the menu.*
- *Any questions on the menu? If not, please click the Map Action button to close the menu.* **[Wait for participant to close Map Action menu]**

(7) *AOI filters: To the left of the Map Action button, there is a button named "AOI filters". AOI stands for "area of interest". This button helps you locate areas of interest on the map. Now please click the AOI filter button.* **[Wait for participant to click AOI Filter button].** *The dotted outlines specify the AOIs for this mission. The*

name of AOIs will be shown in the shapes. Now please click the button again to hide the AOIs.

(8) *Alarms: During the UAV mission, you may see alarms below the Waypoints table at different times. Please pay close attention to the following information because you will be responsible for handling these alarms. There are two types of alarms.*
   - *The first type of alarm requires you to assign priority levels to alarm events.* **[Point out alarm on interface]** *There are 3 levels of Alarms: Alerts, Warnings, and Advisories.* **[Point out Alarm Priorities on paper]** *Alerts are the highest priority, Warnings are medium, and advisories are lowest priority. Please take a minute to read this. When this alarm box shows up, you will be required to prioritize the alarms from 1 to 3, with 1 representing highest priority. In this example, Stall impending is an alert, so you need to type "1" in the input box.* **[Point to the input box for the participant]**. *"Navigation active" is an advisory so please type "3". "Maximum air speed reached" is a warning so please type "2". Now that you finished assigning priority levels, please click the "Done" button. Then this alarm box will disappear.* **[Wait for participant to click]**
   - *The second type of alarm requires you to select an action button to fix the issue.* **[Point out alarm on interface]** *The solutions to all possible alarms are summarized in this table.* **[Point out how to resolve alarms table on paper]** *In this example, the alarm says "Crash Imminent". To fix this, you need to "pull up". The button for this can be found on the right side of the interface under Emergency Controls. Please click it. Now that the issue is resolved, the alarm box will disappear and you will see a confirmation box saying the crash has been averted.*
   - *Remember that these alarms require you to do something for them to be resolved. The buttons to resolve these alarms can be found under the "Emergency Control". They are different from the immediate commands under the "Action" tab, which are only for normal vehicle operation. For the deviation of vehicle status parameters, as shown in the quick display, you only have to announce verbally. Any questions on the alarms?*

(9) *Any other questions on the interface?*

b. Mission Familiarization

Now that you are familiar with the UAV control interface. Let's move on to your mission.
[Hand participant Training Mission Familiarization sheet]
This paper summarizes important information on your mission, including an acronym list, explaining commonly used acronyms. This is where you can find the norms for the system statuses we discussed earlier as well as the Alarm priorities and how to resolve alarms.

[Allow participant to read]

[If no questions, hand participant Familiarization sheet]

This paper presents you the mission map and steps to vehicle maneuver. The map is a to-scale representation of the area you will see presented on the interface.

- *The map presents a Military Grid Reference System (MGRS) to determine an object's location. All major horizontal and vertical grid lines are 1,000 meters apart, making each box a 1,000 m x 1,000 m square. The grid lines 10-15 are Eastings, which provide a designator as to how far East an object is. For example, the Easting WP2 is 149. To get 149, you read the 2 digit Easting that WP2 is past and estimate the third digit. Waypoint 2 is further East than 14 and is 90% of the way between 14 and 15, which gives you 149* ***[Point to line on familiarization]****. The grid lines 20-25 are Northings, which provide a designator as to how far North an object is. The Northing of WP2 is 228. Similarly, to get 228, you read the 2 digit Northing that WP2 is past and estimate the third digit. Waypoint 2 is further North than 22 and is 80% of the way between 22 and 23, which gives you 228* ***[Point to on familiarization]****. Taken together, these readings create a unique designation of 149,228 – the three digits of the Easting comes first then the three digits of the Northing. Now please report the coordinates for WP1.* **[Correct answer: 127 243; If incorrect, provide explanation and ask for WP3 location; Correct WP3 location: 125 203]**
- *Every mission map and interface will present the numbered boxes, but the actual grid lines themselves may not be there, and the tick marks will not be there. The distance between objects can be estimated based on the grid lines or using the hypotenuse between the two objects. For example, the distance between WP3 and the LP is about 2.7 grids. Since each grid is 1000 meters, your estimation would be 2700 meters. Now please report the distance between WP2 and WP3* **[Correct answer: 3,400 meters; If incorrect, provide explanation and ask for distance between WP2 and LP; correct answer: 3,600 meters]**
- *At the bottom right of the map is a compass rose, which means that North is to the top, East is to the right, West is to the left, and South is towards the bottom.*
- *The LP is the red circle.* **[Point out]** *You will launch the UAV here and return to the LP at the end of the mission.*
- *To the Northeast of the LP, you see WP1. Similar to what you see on the interface, the Waypoints are numbered and the UAV will follow the WPs in numerical order.*
- *The next two grey shapes are AOIs. They represent a physical area of special interest to your mission. On the way to WP1, you will pass through a grey arrow named Axis Nova. On the Eastern side of the map, you see NAI nail. NAI stands for "Named Area of Interest".*
- *The areas of interest, the LP, the WPs, the compass, and the grid numbers will be on every mission map.*
- *On the left side of the sheet is a scheme of maneuver. This is a list of tasks you will be required to complete during the mission. It also indicates the order of your tasks. For example, in this mission, you will…* **[Read scheme of maneuver for participant]** *During*

the mission, headquarters will provide a verbal cue before a task must be executed. For example, … **[Play a single audio recording as example]** *Please wait for the verbal cue before you execute any tasks.*

- *Any questions on the mission map or tasks?*

5. HR monitor

Congratulations on completing the training. Now we ask that you don this heart rate (HR) monitor during the experiment. You can use the restroom to put it on. You need to moisten (but not soak) the sensor with water and secure the monitor around your chest very tightly. The sensor should be at the middle of the front of your chest and at the same height of your heart, just like in this picture. An experimenter can accompany you for donning the monitor.

[Show participant picture in watch manual, and escort him/her to the restroom.]

[When participant is back, test HR and RR interval signal.]

[To measure RR intervals:

- **The default display of Polar watch is a time display.**
- **Press "Up"/ "Down" button until you see "Tests". Press the red button to select.**
- **Press "Up"/ "Down" button until you see "RR recording". Press the red button to select.**
- **Select "Start recording" and press the red button. The watch will start searching for HR data from the sensor.]**
- **If the watch says "No HR found", ask the participant to wet the sensor and adjust the strap again.**
- **To stop recording, hold the bottom left button for 3 seconds.]**

6. Training scenario

[Re-simulate the training prototype]

Now that you are familiar with the interface and the mission, let's go through a mission scenario. This mission is also designed to help you learn the system. Before we start, I'm going to ask you a few questions on what we covered during the familiarization session. If you don't know the answer to a question, feel free to let me know, and I will tell you the answer and provide an explanation in terms of the interface content or mission information.

[Ask familiarization queries – refer to Timeline at Time 0]

[Provide answers or point to related material/interface if necessary]

[After Time 0 questions]

During the following training mission, you will be asked similar questions about the system, the task, and the environment. Do your best to answer these questions as quickly and accurately as possible; however it is ok to say that you do not know the

answer to a question. All questions will be presented verbally and please respond with answers, verbally.
The mission tasks are listed in the Scheme of Maneuver. Please read it carefully. **[Wait participant to read]** *Again, please wait for verbal cues from your headquarters (an audio message) before executing any steps. As a reminder, you need to pay close attention to the UAV parameters under the quick display. The normal or acceptable ranges can be found on this document.* **[Point to paper]** *You need to verbally report any deviations but there is no action required. Finally, not all buttons within the interface are active, but all buttons that you need to use to accomplish tasks can be selected. Any questions?*

Please launch when ready.
[Play audio recording to deliver scheme of maneuvers and ask SA questions]
Now you have completed the training mission. The experiment test missions will be similar to this mission. Do you feel comfortable with the mission procedure? If not, we can go through this mission again. **[If needed, go through just the training scenario without the first 14 SA questions]**

7. TLX Ranking

As I mentioned earlier, during this study we want to measure the degree of cognitive workload you perceive in using the UAV control interface for various mission trials. Related to this, we will ask you to complete workload surveys during the experiment. These surveys will require that you rate various task demands, including mental, temporal, effort, physical, frustration and performance. However, before you make ratings, we also want to ask you to rank these various demands in terms of importance to the UAV control task. Based on your training experience, we would now like to ask you to complete the first portion of the survey instrument. Please follow the instructions on the paper. Remember, to consider the mission you just completed as a basis for your answer.
[Hand participant TLX definitions and ranking sheet]
Let me know if you need any clarification on the demand definitions or what aspects of the task you should consider in making rankings. Please note that you will subsequently make ratings after each test trial.

8. Experiment trials

a. Initial Test Trial

[Start screen capture software]

We will now begin the experimental testing. You will use the same UAV control interface. The mission will be similar to the training, but with different tasks.

[New trial starting point] *Please read through the scheme of maneuver before we begin.*
[Allow time to read] *Any questions? Are you ready for this mission?*

Again, remember to verbally report parameter deviations. Please also try to maintain your posture during the mission and limit your head movement so the eye tracker can capture your gaze pattern.

[Start Camtasia]

[Start RR recording]

[Start Eye-tracking]

Ok, the mission will begin. Launch when ready.

[Play audio recording once participant correctly Launches the UAV]
[Deliver SA queries according to Prototype Timeline; pay close attention]
[Second experimenter to record subtask performance and SA responses in gradesheet]

[Stop RR recording upon mission completion]
[Stop eye-tracking; rename files]
[Stop screen capturing upon mission completion]

b. Rest & TLX rating
[Save all files. Name properly, e.g., "P1_Trial1"]

You have completed this mission. Now we ask that you complete the second portion of the workload questionnaire, where you need to rate the workload demand components separately. Please follow the instructions on this paper. **[Hand participant TLX rating sheet]** *Let me know if you need any clarifications on the definitions of the demand components or the aspects of the task that you should consider in your ratings.*

[If polar watch start to lose HR data, ask participant to wet the monitor again]

Thank you for completing the questionnaire. Now, please take a 2 min to rest break.

[Time 2 min]

Your resting period is now complete. We will now proceed to your next trial.

c. Repeat a-b 3 times. (4 trials in total for each participant).

9. Closing questionnaire
Usability: Please fill this questionnaire based on the interface you used.
MCH-UVD: Please start with this point **[point on paper]** *and follow the flow chart to give your rating.*

10. Baseline measurement
Now you have completed all experiment trials. Before we close the experiment, we need to take some additional measurements with the eye-tracking system and the HR monitor.

Please look anywhere on the UAV control screen. I need you to do this for 5 minutes. You do not need to perform any tasks but please do not direct your eyes away from the screen. Stay relaxed and just blink normally.

[Record eye-tracking and RR-interval for 5 mins.]

11. Debrief

a. Complete Payment Form

[Hand payment form to participant.]

[Calculate the participant's compensation.]

Here is the form for your compensation for participation in the study. For your time today, you will receive $___.___.

b. Departure and Thank You

[Give the participant a copy of his or her signed informed consent form as well as the original payment form.]

The experiment is now complete. The data we collected today will be used to investigate the effect of interface design on workload in UAV operations. You will not be personally identified in any of the data analyses or reports based on this study. If you are interested in future information about this experiment, you may contact Dr. David Kaber, whose contact information is included in the informed consent form.

12. Post-Experiment Procedure

a. Organize all data sheets in participant folder
b. Record subtask performance accuracy and times in spreadsheet; verify with videos if necessary
c. Record TLX responses in spreadsheet
d. Save video file with Participant #_Trial #
e. Export HR data to txt files and name properly.
f. Backup files
g. Charge the Polar watch