ARMY RESEARCH LABORATORY

Examining the Convergent and Discriminant Validity of Visual and Mental Workload Using Ocular Activity Variables

by Michael Sage Jessee

ARL-TR-5132 **March 2010**

NOTICES

Disclaimers

Army Research Laboratory
Aberdeen Proving Ground, MD 21005-5066

ARL-TR-5132 **March 2010**

Examining the Convergent and Discriminant Validity of Visual and Mental Workload Using Ocular Activity Variables

Michael Sage Jessee
Human Research and Engineering Directorate, ARL

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*
March 2010

2. REPORT TYPE

3. DATES COVERED (From - To)

4. TITLE AND SUBTITLE
Examining the Convergent and Discriminant Validity of Visual and Mental Workload Using Ocular Activity Variables

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Michael Sage Jessee

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
ATTN: RDRL-HRM-DI
Aberdeen Proving Ground, MD 21005-5425

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-5132

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSOR/MONITOR'S ACRONYM(S)

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

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

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The Mental Workload construct is examined along with Visual Workload in order to discriminate the relationship between the converging constructs. Six ocular activity variables were measured in order to test their diagnostic and sensitivity properties with regards to Visual and Mental Workload in a UH-60M upgrade simulated flight test. Three pilot crews flew six flight scenarios in which subjective and physiological mental workload measures were implemented across Task Difficulty and Task Differences (pilot on-controls versus pilot off-controls). Results indicate less subjective mental workload for the Hover task compared to the Action on Contact task and greater fixation duration variability for the lower difficulty task. Blink interval was greater for the pilot on-controls and saccadic extent was greater for the pilot off-controls. Results suggest that blink interval and saccadic extent are diagnostic of different aspects of visual workload, whereas fixation duration variability is sensitive to mental workload.

15. SUBJECT TERMS
Mental workload, visual workload, ocular activity, UH-60M

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Michael Sage Jessee
Unclassified	Unclassified	Unclassified	UU	68	19b. TELEPHONE NUMBER (*Include area code*) (256) 842-8830

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

Contents

List of Figures

List of Tables

INTENTIONALLY LEFT BLANK.

1. Introduction

1.1 Statement of the Problem

The purpose of this research is to examine the convergent and discriminative validity of Visual and Mental Workload using eye tracking metrics. We collected ocular activity data that have previously been correlated with mental workload (MWL) as a measure of Visual and Mental Workload in similar experimental circumstances. Much of the previous research that has been conducted in this area has attempted to identify evidence that ocular activity variables provide an unobtrusive objective measure of operator states that highly correlate with performance, namely MWL. Although some of these variables have successfully demonstrated a link to MWL, they are often fairly task-specific and have not been linked specifically to visual workload (VWL). To date, this author knows of no data that have been collected that evaluate ocular activity measures of task differences while holding task difficulty constant. More specifically, these data have not been collected under conditions where operator freedoms are somewhat unconstrained in terms of very specific ocular activity behavior.

The importance of collecting these data has many benefits. First, the nature of many complex systems allows operators the freedom to perform tasks in a great number of ways while maintaining constant performance levels. For example, while driving a car, the driver is required to look outside the window in order to monitor their position and trajectory. Although there is a known area of interest toward which drivers direct their gaze while looking outside the window (Land & Lee, 1994), the specific fixation frequency, fixation duration, and number of transitions into that area of interest can vary substantially. Thus, finding ocular activity measures that are sensitive to visual and/or mental demands is one of the primary challenges of this research.

The second benefit of this research is that the measure will provide continuous data about operator Visual and Mental Workload. Currently, some very commonly used MWL measurement strategies—subjective measurements—are insensitive to subtle and quick shifts in MWL, which a continuous measure is capable of capturing. In a multi-tasked environment, such as driving a car or piloting a helicopter, MWL can quickly increase to unsatisfactory levels for short periods of time. Global (off-line) assessment techniques tend to not be sensitive to those events, but continuous physiological data can provide information about operator states related to MWL and VWL. This would be beneficial to researchers who seek to identify measures that are operator state specific and perceptual processing structure specific. The state of the operator can be inferred from MWL levels, and the perceptual system being used can be gathered from VWL data. Additionally, it would provide developers and evaluators the means to apply a strategy of operator VWL and MWL measurement to virtually any operator task.

The third benefit of this research is the use of an interval measurement scale. Using interval data (physiological measure) rather than ordinal data (subjective measure) is advantageous because interval data has fixed and defined intervals, whereas ordinal data lacks sensitivity between data points. An ordinal scale consists of numbers that can be meaningfully ordered. This is achieved by ranking the data. An interval scale can also be ordered from lowest to highest, but the magnitude between each ranking is fixed. This means that all the information available in an ordinal measurement scale is also available in an interval scale, i.e., the ability to order the data. However, ordinal measurement scales lack known intervals between the data points. The motivation for the current research is to validate an interval measurement scale that is more sensitive and diagnostic of Visual and Mental Workload.

The research question for this study is, what measures of VWL, as measured via ocular metrics, converge and diverge with a traditional subjective measure of MWL in a multi-tasked rotary wing simulation? In order to build the case for such an inquiry, we will first briefly review the MWL concept and measurement strategies. Then, we will discuss VWL in terms of specific ocular activity variables, followed by a presentation of previous research related to ocular activity as a measure of MWL. Once this is complete, specific hypotheses will be presented.

1.2 Mental Workload

MWL is a multidimensional construct that has no universally accepted definition. Along with the diverse set of measurement techniques, MWL has been conceptualized in various formats. For example, Jex (1988) suggested that MWL is the operator's evaluation of the attentional load capacity between their motivation and current task demands while achieving sufficient task performance. In other words, MWL is the operators' motivation to distribute attentional resources (limited) to current task demands in order to maintain operator performance. Similarly, Wierwille (1979) defines MWL as the operator's subjective experience of cognitive effort. In time sharing tasks, MWL is the ratio of the total time required to complete a task to the time available to complete the task (Hendy, Liao, & Milgram, 1997), but this definition is primarily implemented for mentally loaded tasks as opposed to physical tasks, such as running 50 yards. More generally, MWL can be defined as the ratio of resources required of a task to the total resources available to perform the task (Wickens, Lee, Liu, & Gordon-Becher, 2003). Orlady and Orlady (1999) defined MWL as the cost incurred by a human operator in achieving a specified level of performance. A definition proposed by Young and Stanton (2001) suggests that MWL represents the level of resources required to meet both objective and subjective performance criteria, which are mediated by task demands, external support, and past experience. Although this summary of MWL definitions is not a comprehensive list, it captures most of the prominent aspects of this multifaceted construct.

Jex (1988) and Wierwille (1979) focus the scope of MWL to a subjective estimate of effort, which also provides information about how to measure it, namely with a subjective report. However, subjective measures have several limiting factors. Responses can be fabricated if the participant desires to be perceived in a certain way; the subject may be unaware of important cognitive events that are related to MWL, or the participant may be unable to report such events (Knust, Marshall, & Ishizaka, 2000). The last four definitions of the previous paragraph emphasize that MWL is a ratio of the costs incurred from performing a task over the cost required to perform the task. Hendy and colleagues (1997) provided a time-ratio definition, previously mentioned, but it does not capture task complexity and is not diagnostic of particular processing structures involved in task completion. Wickens and associates (2003), as well as Orlady and Orlady (1999), emphasize the measurement of the cost or resources required to perform a given task, but typically researchers do not explicitly specify the nature of those resources. Young and Stanton (2001) propose a similar definition in terms of measuring resources, and the strength is that their definition outlines three primary contributing factors, namely, task demands, operator histories, and environmental influences. For example, a task demand sets the limit on what the resource requirements are for goal completion. Operator history influences the availability of resources through automation, fatigue, cognitive appraisals, and other factors unique to an individual. Environmental influences mediate the availability of external resources. For example, the availability of an airplane greatly increases a human's ability to fly by providing a resource for a goal that would otherwise be unobtainable. Further detail will now be provided about these mediating factors in light of specific research that has manipulated MWL by altering these primary factors.

1.2.1 Antecedents of Mental Workload: Task Demands, Operator History, and External Support

1.2.1.1 Task Demands

Given that MWL is mediated by task demands, external support, and operator history, the amount of available resources required to achieve a goal can be influenced in several ways. For example, previous research has manipulated task demands by increasing task complexity (Veltman & Gaillard, 1998). Participants flew through a simulated tunnel that varied by horizontal and vertical maneuver requirements, while performing a continuous memory task. During the more simple flight tasks, participants were required only to navigate, but during the more complex tasks, participants combined navigation with speed modulation. The continuous memory task required participants to respond with a single button press when target letters were heard and respond with two button presses when non-target letters were heard. Task complexity was manipulated by increasing the number of target letters the participant tracked. MWL was measured with heart period, continuous blood pressure, respiration, and eye blinks.

Brookings, Wilson, and Swain (1996) manipulated MWL by varying task demands, via complexity and air traffic volume (temporal component). Air Traffic Controllers (ATCs) performed on a Terminal Radar Approach Control (TRACON) simulation under three different scenarios. The first scenario manipulated task demands by varying the volume of traffic to be handled, which increased the temporal constraints on the operator. The second scenario varied traffic complexity by changing the arrival to departure ratios, as well as mixing the aircraft types. The third scenario, overload, also increased the temporal constraint of task demands by requiring the participants to handle a large number of aircraft in a short amount of time. The effects of MWL manipulation were recorded from performance, subjective (NASA-TLX), and several physiological measures.

In a dual task scenario, Ryu and Myung (2005) manipulated MWL via task demands. For the primary task of a simulated instrument landing on a final approach, participants were required to use a mouse to control the pitch of an aircraft for glide slope corrections. MWL for this task was manipulated by changing the velocity of the target. Mental arithmetic was implemented as a secondary task in which participants mentally computed the sum of two numbers. Difficulty was manipulated by changing the number of digits in the arithmetic question. For the easier task, the sum of two double-digit numbers in which participants were not required to carry a number was computed, but during the more difficult arithmetic task, the sum of two triple digit numbers had to be computed in which a number also had to be carried. Van Orden, Limbert, Makeig, and Jung (2001) manipulated MWL during a mock air warfare task where participants identified oncoming targets as either friend or foe via a look-up table. Ocular activity was used as a measure of MWL. Task difficulty was manipulated by increasing the number of targets on the screen. In addition to task demand manipulations of MWL, performance variability has also been examined under various operator state constraints.

1.2.1.2 Operator History

As an example of operator state constraints, previous research conducted on driver efficiency by Wikman, Nieminen, and Summala (1998) showed that differences in performance varied as a function of novice versus expert drivers. Although MWL is not explicitly mentioned as a manipulation in their article, they suggest that experience and automation relate to the efficiency of resource allocation, which can be considered an antecedent of MWL because the on-board resources used during the allocation process are decreased due to automation. This is similar to manipulating the effectiveness of a hypothesized central executive (Baddeley, 1986). The central executive is a theorized component of the operators' working memory model that is responsible for the selection, initiation, and termination of processing routines; these factors can change the amount of effort required to perform a task, thus influencing MWL. For example, the central executive is seen as the strategy planner; a central executive of the novice driver will, therefore, tend to be less efficient and experience a higher level of MWL since more resources must be spent in order to plan the appropriate strategy.

Van Orden, Jung, and Makeig (2000) examined fatigue-related changes in performance while participants controlled a simulated target disk with a trackball. Participants were required to keep the target disk within an annulus for 53 min, while pseudorandom wind forces destabilized it. Over the course of the trials, the amount of effort required to perform the task was mediated by the state of the operator. Thus, MWL increased across time because continuous performance requires an increase in the amount of resources in order to sustain performance. Although practice typically leads to expertise and a reduction in MWL, longer periods of continuous time on task periods can lead to an increase in MWL because of fatigue effects, which was the purpose in Van Orden, Jung, and Makeig's study. In a study where cognitive interference and self-efficacy were examined, Knust, Marshall, and Ishizaka (2000) had participants perform a dual-attention task during which they monitored six gauges and evaluated arithmetic expressions. Knust et al. found that participants with higher self-efficacy (pre-task screening) also performed better. Specifically, the cognitive workload index (CWI) was positively correlated with measures of task-related intrusive thoughts. This provides evidence that reports of MWL are mediated by variables related to individual differences. Aside from variability in task demands and operator history, the amount of external support an operator has can also moderate his or her level of MWL (Ahlstrom & Friedman-Berg, 2006; Backs & Walrath, 1992).

1.2.1.3 External Support

The amount of resources an operator has available is mediated by the environment in which they operate. For example, environmental influences on MWL can include, but are not limited to, motion (low and high frequency vibrations), temperature, luminosity, sound intensity (Wickens et al., 2003), and even air quality (Kramer, Coyne, & Strayer, 1993).

In a study investigating ATC operations and workload, Ahlstrom and Friedman-Berg (2006) report significant differences in blink duration, as a measure of MWL, across various weather displays (external support manipulation). Researchers implemented an auxiliary presentation of weather, which was located above the operator workstation, a weather condition, where the weather information was overlaid on the controller workstation, and a control condition, where no weather display was used. Although subjective reports of operator MWL were not significantly different across weather display conditions, blink duration, which has shown to be a sensitive measure of MWL (Van Orden, Limbert, and Makeig, 2001), was significantly different across weather display conditions. This suggests that MWL was higher during the control condition compared to both other conditions in which external support was available. In terms of how external support mediates MWL, these data imply that the availability of information aids to the operator lowers the amount of effort required to perform a task. In a similar study in which ocular activity measures were taken to evaluate the addition of color coding to a symbolic tactical display, Backs and Walrath (1992) found that multiple ocular measures related to MWL varied as a result of color-coded displays versus monochrome displays. Again, these research findings indicate that the quality of the external support mediates the amount of effort required to perform a task.

1.3 Consequences of Mental Workload

One of the primary reasons the MWL construct is studied is because of its relationship to performance, which can most generally be characterized by O'Donnell and Eggemeier's (1986) hypothetical relationship between primary task performance and operator MWL. Figure 1 displays operator performance under conditions of low MWL (region 1), during which the performance of the operator is unhindered by the level of workload being experienced. Region 2 demonstrates how the performance of the operator tends to decrease after a threshold where the level of workload starts to impair performance. Region 3 shows the reduced operator performance under levels of high MWL. Although MWL is a multifaceted construct, this simplified description suggests the consequences and general motivation for measuring MWL and its associated performance profile. However, there are other MWL consequence descriptions that emphasize capacity regions and operator state descriptions—specifically, levels of arousal.

Figure 2 illustrates the hypothetical relationship between resource demands and the level of MWL that is experienced. Within the spare capacity region, the operator has enough resources to perform at a stable level until the overload region is reached. In this paradigm, MWL is defined as the ratio of the time required (TR) to perform the task over the time available (TA) to perform the task (TR/TA), and essentially characterizes urgency. When the ratio passes 1, performance begins to decrease, while the subjective experience of MWL by the operator continues to increase (Hendy, Liao, & Milgram, 1997). Similarly, the relationship between MWL and performance can also be likened to the Yerkes-Dodson law (Yerkes & Dodson, 1908). Under low levels of arousal, performance tends to suffer and then reach a peak where arousal and performance are optimized. After the optimal level of arousal is passed, performance begins to decrease.

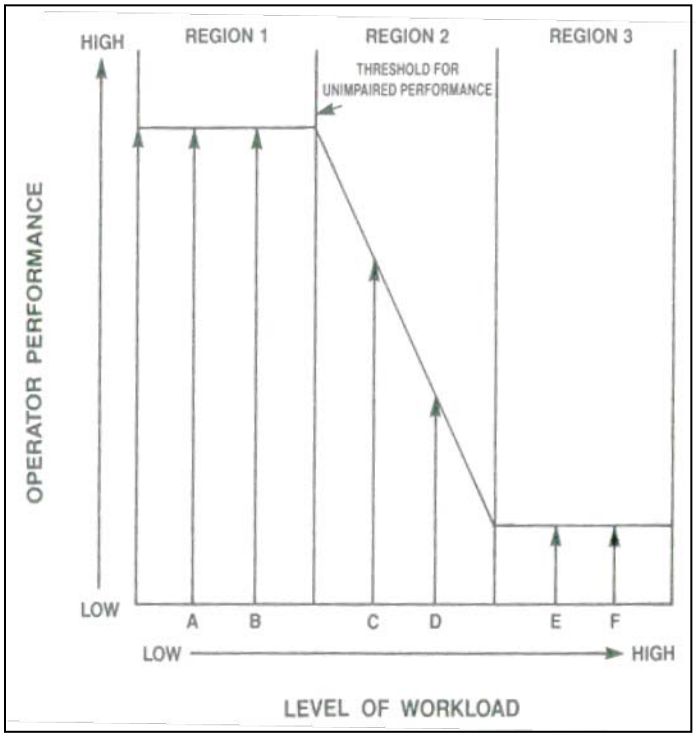


Figure 1. Hypothetical relationship between MWL and primary task performance (O'Donnell & Eggemeier, 1986).

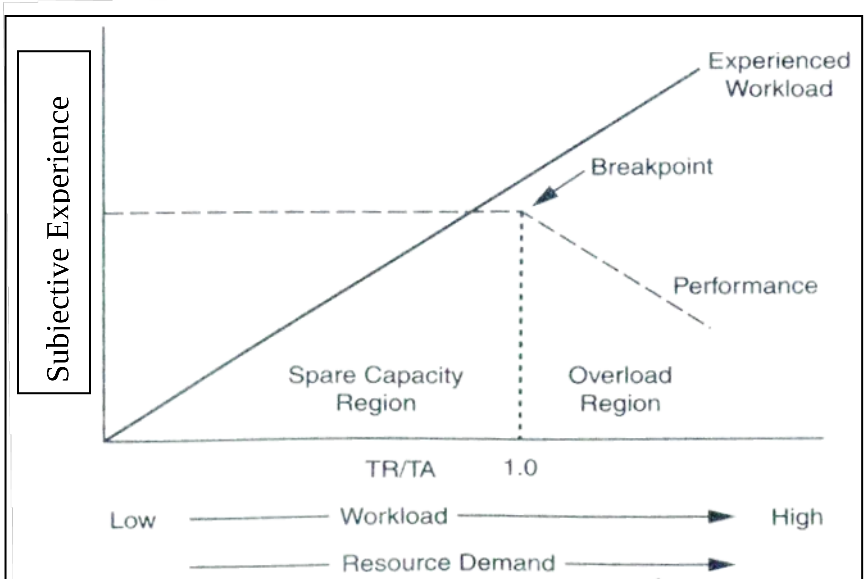


Figure 2. Hypothetical relation between MWL imposed by task and MWL experienced (Wickens et al., 2003).

Once operators reach an overload region, aspects of their performance begin to decrease. Performance decrements can take many forms, but Edland and Svenson (1993) outline several general operator tendencies that occur under time-sensitive tasks. Typically, operators give more weight to more important sources of information, are more selective about their input, inadvertently decrease their accuracy of input, decrease the use of strategies where heavy mental computation is required, and lock into a single strategy, often referred to as cognitive tunneling (Cook & Woods, 1994).

1.4 Measuring Mental Workload

Wickens and colleagues (2003) suggest that measuring MWL can be useful for three different purposes. First, assessing MWL during task execution (continuous measurement) can provide feedback to a system controller who can then allocate workload to team members such that optimal operation can be achieved. For example, if an individual operator in a group of ATCs begins to experience high levels of MWL because of an increased number of aircraft in an assigned sector, other operators on the team can assist with his task demands. Second, MWL can be assessed in usability analyses. For example, if workload is found to be too high during usability testing, then the system design may need to be changed so that operators can achieve optimal performance more efficiently. Lastly, assessing workload can contribute to predictive models of MWL. These can be used as aids to develop and model human performance for simulation and for test and evaluation purposes. In order to successfully measure MWL, researchers must be able to characterize the measurement properties for each of four general MWL measurement techniques—primary task analysis, secondary task analysis, subjective measures, and physiological measures.

As reported by Wickens and Hollands (2000), a primary task measurement is an evaluation of performance on a given system. For example, if one were evaluating a new type of navigation device, then a pilot may simply be asked to use it to navigate to a waypoint or landing zone. The dependent variable for this evaluation might be "time to waypoint," using the new system versus the old system. Additionally, if one were to evaluate Web site design, she may have a user perform a common task such as finding the contact information of the Web site owner or purchasing a product from the Web site. Here, the dependent variable might be the number of button presses or frequency of backtracks. These are all primary task measures because the users' performance is being evaluated based on tasks that are relevant to the goals informed by the system. However, there are four limitations of primary task measurement (Wickens & Hollands, 2000). First, if two tasks are being measured that lie in an under-load region (i.e., operator has enough resources to perform both tasks perfectly), then it is difficult to gather sensitive data regarding the MWL between the two tasks. Second, two primary tasks may be measured differently and may be different in what those measurements mean. For example, a pilot's performance on a revised communication system may produce more errors in terms of channel selections, but there may be no significant difference in subjective ratings of performance. Third, sometimes it is difficult to obtain a primary task measurement. A task might impose rigorous mental effort, but its outcome (correct or incorrect response) may not provide any insight into processes involved in the outcome. For example, a correct response on a math question does not necessarily provide exact information about the cognitive and decision-making processes underlying the response. Finally, primary tasks may differ in their performance measures, not because of MWL, but for many other reasons, such as the use of automated voice recognition systems as opposed to manual data-entry systems. Automated voice recognition systems may require fewer resources from the operator compared to manual systems, but often performance differences can be measured because of automated system reliability. In short, primary performance measures can differ for many other reasons than MWL changes.

A secondary task measurement technique occurs when a non-primary task is imposed upon the operator—in addition to the primary task—in order to examine how much "left over" resources the operator possesses. One limitation with this measurement technique is that the secondary task may not always draw from the same processing structures that the primary task depletes. For example, during a driving exercise the primary task is visual-spatial with a manual response; thus, the secondary task should also be visual-spatial with a manual response. If the secondary task does not access the same processing structures, then the resources required to fulfill secondary tasks may be independent of those required to fulfill the primary task (Schlege, Gilliland, & Schlegel, 1986). Essentially, this means that secondary tasks, as independent variables, do not always lead to measurements that are diagnostic of the factors of primary interest that influence MWL. Another limitation with secondary task measurement is that it may obstruct performance on the primary task. Additionally, secondary unrelated tasks lower the ecological validity of the experiment. In order to account for this, some researchers have used secondary tasks that are legitimate components of primary tasks (Raby & Wickens, 1994).

Another widely used MWL measure is the subjective MWL measure. Subjective MWL scales can come in many forms. However, they can typically be categorized as either unidimensional or multidimensional (Young & Stanton, 2005). Four of the commonly used MWL measures include the NASA Task Load Index (TLX) (Hart & Staveland, 1987), the Bedford Workload Scale (BWL) (Roscoe, 1984), the Subjective Workload Assessment Technique (SWAT) (Reid & Nygren, 1988), and the Cooper-Harper Rating Scale (Cooper and Harper, 1969). The NASA TLX is a multidimensional scale that assesses mental demand, physical demand, temporal demand, performance, effort, and frustration level, and requires an individual score for each dimension. The SWAT is also a multidimensional MWL scale, but measures only three dimensions—temporal load, mental effort, and psychological stress. The Cooper-Harper Rating Scale is a unidimensional decision tree-based scale that captures a general measure of MWL and is easy to administer. However, Roscoe's (1988) BWL was implemented during the current study for several reasons. First, it is quick and easy to administer. Each participant reported a workload measure for several Aircrew Training Manual (ATM) tasks, such that their responses to MWL questions about many individual tasks would have caused a fatigue affect. Additionally, the BWL has been shown to be a reliable and valid measure of MWL (Corwin et al., 1989), while Svensson, Angelborg-Thanderz, Sjoberg, and Olsson (1997) reported the reliability using Cronbach's alpha of the BWL to be +0.82.

Psychophysiological measures of MWL are often broken down into two primary categories, central nervous system (CNS) measures and autonomic nervous system (ANS) measures (Tsang & Wilson, 1997). CNS measures include electroencephalic activity (EEG), magnetic activity of the brain (i.e., FMRI), event related brain potentials (ERPs), and measures of brain metabolism. ANS psychophysiological measures of MWL include pupil diameter, cardiovascular activity, respiratory rates, and measures relating to visual scanning such as eye blinks and saccadic distances. For a good review of the psychophysiological measures unrelated to ocular activity, see Kramer (1991). For the current study, ocular activity refers to fixation duration variability, average fixation duration, saccadic extent, pupil size variability, average pupil size, and blink interval. A fixation is the period between saccades when the eye dwells upon visual information. A saccade is a quick discrete movement of the eye from one location to the next, and a dwell or fixation is time spent between saccades while the eye is relatively stationary. During this fixation period, the visual system extracts the most visual information. Saccadic extent refers to the spatial length between fixations and is measured in degrees via visual angle. Pupil size variability is a measure of how much the pupil size deviates from its mean. Blink interval is the average amount of time, in seconds, between each blink.

There are six primary workload assessment properties that must be considered in order to implement the most effective MWL assessment technique; these include sensitivity, diagnosticity, intrusiveness, reliability, implementation requirements, and operator acceptance (O'Donnell & Eggemeier, 1986). A MWL measure must be sensitive, in order to detect differences in varying levels of workload that are associated with task performance. For example, short and sudden increases in MWL need to be measurable in order to identify very specific subtasks or situations that may have catastrophic consequences to an operator if they are unable to adapt quickly. Second, diagnosticity refers to an instrument's capacity to discriminate between different causes of workload. For example, MWL can change because of increased task demands or operator fatigue. A diagnostic measure is capable of discriminating these two causes. Hart and Staveland's (1987) NASA TLX, for example, was designed to discriminate between physical demand, mental demand, temporal demand, performance, effort, and frustration level. Third, intrusiveness is the disruption of the primary task that the measurement imposes upon the operator. This is an important variable to consider, depending on the nature of the task being evaluated. An intrusive measure can alter primary task performance, thus limiting the data's applicability to the actual task, as well as how successfully the data generalize. MWL measures must also be reliable; a reliable measure is one that provides stable scores across similar situations for the same operator. Implementation requirements include any hardware or software that is used during the MWL evaluation. It also encompasses the prerequisite training required for proper application of the assessment technique. Largely, this is a matter of managing the logistics of administration and can often result in using a less favorable measurement technique because of situational and performance constraints. Lastly, operator acceptance can influence the measurement process. If an operator's perception of the utility of a measurement is unfavorable or physically uncomfortable, which can be the case with some physiological measures, then the required responses and precision of responses may be hampered. Each of these measurement properties can be brought to bear upon the previously mentioned MWL measurement categories.

Primary task analysis measures can be extremely broad. Thus, their measurement properties vary considerably, as well. They can be designed primarily for sensitivity or diagnosticity, but they must be reliable. Secondary task analysis can also vary widely in its sensitivity and diagnostic properties. However, these measures tend to be mostly hindered by their intrusiveness. Subjective measures tend to lack sensitivity and diagnosticity. The global nature of subjective measures reduces their sensitivity, and the lack of multi-dimensional questions in many subjective measures decreases their diagnostic properties. However, subjective measures tend to have favorable implementation requirements. Physiological measures of MWL can be sensitive and diagnostic, and have been shown to be reliable (Wilson, 2002). Their intrusiveness can often interfere with primary task responsibilities and the implementation requirements can hinder the logistics of managing larger operational tests. Lastly, operator acceptance can influence the quality of data that is being recorded.

1.5 Ocular Activity Variables as a Measure of Visual Workload

Wickens (1979) proposed a challenge to measure separate processing structures associated with MWL, instead of searching for a single measure capable of discriminating all aspects of MWL. Eventually, this research led to what is known as the multiple resource theory, presented by Wickens (1980) and expanded upon in Wickens and Hollands (2000). The purpose of this challenge was twofold. First, there seemed to be an ample amount of evidence that predicting interference between tasks was highly dependent upon structural factors. For example, requiring an operator to perform multiple responses does not necessarily hinder perceptual factors. Automobile drivers can successfully manage lateral movement and appropriate acceleration parameters, and listening to a radio does not appear to hinder this performance. This suggests that the resources required to perform a task are not completely undifferentiated. Second, collecting structure specific data—in this case, ocular activity data—can provide information about operators' ability to perform other tasks concurrently, which is important in a multi-tasked environment, such as piloting a helicopter. As previously alluded, this is the primary rationale for conceptualizing ocular activity as a measure of VWL that is capable of converging with aspects of the broader MWL construct.

Compared to MWL, the VWL construct has received considerably less attention. Although VWL has been mentioned in several studies (Backs & Walrath, 1992; Hancock & Desmond, 2001; van der Horst, 2004; Verwey & Veltman, 1996; Wickens, Helleberg, & Xu, 2002), the concept has yet to be clarified in a meaningful and concrete manner. Van der Horst (2004) studied VWL, in terms of occlusion, in order to measure visual information processing performance. Participants were fitted with a set of spectacles that were either transparent or light-scattering (milky). The glasses allowed participants to press a micro-finger switch as a request for visual information, thereby making the glasses transparent. The time between visual input requests was taken as a measure of VWL. Researchers suggest that increasing speed, reducing lane width, and driving at night increased VWL because the time between successive visual information requests decreased.

Research conducted by Wickens, Helleberg, and Xu (2002) examined VWL in terms of visual scanning behavior by characterizing the total dwell times spent gazing at the instrument panel, outside the window, and the cockpit display of traffic information (CDTI). It was found that, compared to flights with no CDTI, visual attention was primarily "stolen" from the outside the window scene in order to monitor the CDTI, while percentages of gaze duration stayed relatively stable for the instrument panel. This was considered acceptable, considering that the primary information for aviating (attitude and airspeed control) is located on the instrument panel, which supports the notion that primary task hierarchy of performance, aviate, navigate, and communicate (Wickens, Xu, Helleberg, & Marsh, 2001) was maintained.

Verwey and Veltman (1996) compared nine different workload assessment measures and suggest that eye blinks are diagnostic of VWL, because people tend to suppress eye blinks when they have to process visual information (Veltman & Gaillard, 1996). Additionally, Verwey and Veltman (1996) reported that Veltman (1993) suggested that VWL and MWL can diverge. During driving, VWL was relatively high in situations where turns and curves were executed, but MWL was affected much less. The current research measures VWL based on ocular activity variables previously mentioned and measured via video-based infrared eye-tracking technology. The following section will outline the proposed relationship between saccadic extent, blink interval, and pupil size variability, and how they measure VWL based on (a) the proposition that VWL is highly related to the amount of effort required to obtain visual information and (b) the importance of that information as reflected in how frequently it is occluded via blink activity.

1.5.1 Saccadic Extent

Saccadic extent refers to the distance between successive fixations, as measured in units of visual degrees. VWL refers to the effort required to obtain visual information; thus, longer saccades are indicative of higher VWL. As relevant visual information is dispersed across longer distances within the operator's field of view, VWL is increased. Although it is also more difficult to discriminate similar visual items that are clustered closely together, I propose that the difficulty is one related to cognition and not the effort of the visual system. Thus, items dispersed too far apart in a visual scene result in increased VWL, but items that are clustered very close together offload VWL at the expense of increasing MWL. The farther the eye and head must move to acquire information, the greater the VWL. Wickens (2002) alludes to this point when describing overlapping audio and visual channels. Two overlapping visual information sources, if they are far apart, require an added cost to the operator when scanning between them is required. For example, in the UH-60M upgrade helicopter, there are two Multi-functional Displays (MFDs) per pilot that allow them to gather pertinent information not limited to navigation route, above ground altitude, and heading. Visual scanning between the two displays is required; thus, moving the displays far apart in the cockpit would result in higher VWL because the effort required to sample the information between the two displays would increase. Conversely, if the two displays are close together, the distance between sources of visual information is reduced and the required amount of effort used to fixate between them decreases, thus decreasing VWL. If saccadic extent increases during across Task Difficulty (MWL manipulation) and Task Differences (VWL manipulation), then saccadic extent is an ocular activity measure that converges with both constructs. However, if saccadic extent varies across only one of the independent variables, then it can be implemented as a discriminator of the two constructs.

In addition to the effort required to sample visual information that is dispersed across a display and that increases VWL, saccadic suppression must also be considered in terms of the duration in which visual information is occluded. Dodge (1900) originally noted that an observer is incapable of seeing their eye movement during a saccade while looking in the mirror. This phenomenon became known as saccadic suppression (Zuber, Crider, & Stark, 1964). During this event, the visual perception system suppresses visual processing such that the observer is essentially "blind" during a saccade. The magnitude of saccadic suppression tends to be negatively correlated with the velocity of eye movements, such that during a smooth pursuit, eye movement saccadic suppression (suppression of visual contrast sensitivity) is mediated by the velocity of the eye movement (Schutz, Braun, & Gegenfurtner, 2007). This phenomenon has important implications for VWL that are primarily driven by van der Horst's (2004) data on occlusion. During a saccade, an operator has limited visual information between fixations, and typically the duration of saccades tend to last longer when they are of greater distance. Of course, the operator has the ability to increase the velocity of their saccades in order to decrease the amount of time spent between fixations, but I suggest that VWL is increased with larger saccades, even if performed faster, because the effort required to travel a larger distance is increased. However, it is unclear if saccadic suppression increases the work done by the visual system, or if the suppression reduces the load of the visual system.

1.5.2 Blink Interval

During a blink, the visual scene is briefly occluded. As previously mentioned, van der Horst (2004) conducted research in which vision was occluded using a special set of glasses that allowed the driver to request visual information. Similarly, blinks tend to be inhibited in order to not occlude the visual scene during periods where increased amounts of visual information must be acquired and performance increased (Ahlstrom, & Friedman-Berg, 2006; Brookings, Wilson, & Swain, 1996; Ryu & Myung, 2005; Van Orden, Jung, & Makeig, 2000; Van Orden, Limbert, Makeig, & Jung, 2001; Veltman & Gaillard, 1998). Thus, it is suggested that blink rate is negatively correlated with VWL, since fewer blinks indicate that more effort is being applied to the visual system.

To date, there have been no blink data collected where the visual scene is held constant (a specific fixation point) and MWL manipulated. Typically, in multi-task environments, the visual demands of a system are correlated with overall system demands. Without these data, researchers are unsure whether the inhibition of blinks is purely related to a reduction of occlusion, indicating a strong need to request visual information, or if blink inhibition is related to concentration and thus a measure of MWL. The current research will examine blink interval data across both Task Difficulty (MWL) and Task Differences (VWL) in order to draw conclusions regarding the nature of blink interval measurements. If blink intervals are reduced across Task Difficulty and Task Differences, then blink intervals may be capable of measuring both constructs, or aspects of each that are enmeshed, and would require further experimentation to discern. If blink intervals reduce across one or the other independent variables, then the data would support the notion that blink interval is diagnostic of MWL or VWL. Although I do not have the capability to hold the visual scene constant across Task Difficulty, comparisons across Task Differences with differing visual demands are capable of showing if blink interval can determine visual load while difficulty is averaged.

1.5.3 Pupil Size Variability

I suggest that the overall control system required to produce pupil changes must work more during periods where the size of the pupil varies considerably across time, compared to periods when it stays relatively constant. This means that VWL is increased when the pupil has to work harder to adapt to the state of the operator and state of the environment. Under conditions where luminosity changes frequently, the pupil response would indicate high VWL and add to fatigue, subtracting from overall visual resources; however, this response would likely have a smaller effect on MWL than perhaps a difference in actual task load. Research conducted by Beatty (1982) has provided evidence for task-evoked pupillary responses, which suggests that VWL can be affected by MWL and luminosity. However, Beatty outlines research conducted in only task evoked pupillary responses in which luminosity is controlled for, and suggests that pupil dilation is a valid indicator of processing load, a construct within the nomological network of MWL. Data collected on within-task, between-task, and between-individual variations of pupil activity were examined. Resource and general capacity models are addressed in explaining why pupil activity is capable of measuring processing load, but the general conclusion is that pupil activity is capable of indicating a general amount of resources being used, yet is incapable of discriminating specific processes. An analogy from Kahneman and Beatty (1967) is used, in which an amperage meter on a house is likened to pupil activity, but that specific metric does not inform the observer about the particular devices in the house that are using the resources. That analogy relates to VWL in that pupil variability indicates an adaptive process taking place, where higher variability suggests that the visual system is expending more effort to acclimatize to the state and situation of the operator. Depending on the purpose of the measurement, researchers must account for these environmental and operator state variables in order to make sound conclusions concerning the causes of pupillary responses. For example, Marshall (2002) suggests using the Index of Cognitive Activity (ICA), which is a pupillary response measure that accounts for variations of luminosity, and essentially discriminates dilation and contraction due to cognitive processes with respect to non-task loading events (changes in luminosity) related to pupil changes.

Each of these measures varies slightly with regard to the six previously mentioned measurement criteria. Blink interval may be sensitive to VWL and diagnostic of different types of workload, namely, MWL and VWL. Saccadic extent is diagnostic of VWL, but the sensitivity of it during low workload periods may be in question because of the efficiency of the visual system in terms of VWL as a meaningful measurement—i.e., its correlation with performance. Pupil size variability is capable of being sensitive to VWL, but its diagnostic properties are lacking because the variability could occur for many unknown reasons, including luminosity and operator state variables. Depending on the ocular-meter used to gather eye tracker data, the implementation requirements can vary markedly, but typically they are much more difficult to implement than basic primary task analysis techniques and subjective measures. Modern eye trackers have also become less obtrusive and, thus, have increased operator acceptance. Reliability about the specific dependent variables used in this study can only be generalized from Wilson's (2002) data indicating that psycho-physiological variables are reliable across similar flight scenarios flown by the same pilot two weeks apart.

1.6 Ocular Activity Variables as a Measure of Mental Workload

Previous research has linked fixations to information processing (Salthouse, Ellis, Diener, & Somberg, 1981), but few researchers have correlated fixation duration variability with MWL. The only data that have been reported on fixation duration variability in the context of MWL showed significant differences between novice and expert drivers (Wikman, Nieminen, & Summala, 1998), and novice versus expert pilots (Tote, Stephens, Vivaudou, Ephrath, & Young, 1983). Wikman and colleagues showed that the variability of fixation durations on in-car secondary task instruments can discriminate between novice and expert drivers, but they only reported differences across specific instruments and not for all fixations, which the current study will examine. Specifically, novice drivers had significantly larger fixation duration variability than expert drivers. The correlation of fixation duration variability with drivers' experience suggests that an aspect of MWL related to operator history can be detected by measuring this variable. Similarly, Tote and colleagues found that novice pilots tend to have increased fixation durations compared to expert pilots, but variability was not examined. Data from Wikman's study suggest that fixation duration variability provides information about task familiarity and may possibly generalize to training effectiveness. Pupil size has also been linked to MWL (Ahlstom & Friedman-Berg, 2006; Beatty, 1982) along with other ocular activity variables. The following review will provide a detailed account of research conducted on ocular measures correlated with MWL.

1.7 Research Correlating Ocular Activity Variables with Mental Workload

Previous research has found that measures of ocular activity are capable of predicting MWL. However, none of this research has held task difficulty constant while comparing ocular activity measures across tasks that are rated similarly, but require the operator to perform different tasks. The following literature review outlines past research in terms of real and simulated flights, air traffic controllers, and more general dual task scenarios. These data sets largely overlap with some of the previous data reported on VWL, but the following presentation outlines the research in greater detail. The organization of this review will inform the hypotheses related to ocular activity predictions across Task Difficulty and Task Differences.

1.7.1 Real and Simulated Flights

Psychophysiological and subjective MWL data were collected during a non-simulated flight scenario, as 10 pilots each flew a Piper Arrow (fixed wing aircraft) during separate 90 min scenarios under both visual (VFC) and instrument flight conditions (IFC) (Wilson, 2002). Measures of heart rate, heart rate variability, eye blinks, electrodermal activity, topographically recorded electrical brain activity, and subjective measures of workload were recorded. In order to ensure reliability of the physiological measures, each pilot flew the same scenario twice under VFC and IFC conditions. Results indicate that there were no significant differences in blink activity for the replication of flight scenario, indicating that blink activity is a reliable measure of workload. In Wilson's study, each flight was broken down into 22 segments, which included a preflight baseline, preflight checklist, engine start, and a number of other aviation segments under both instrument flight rules (IFR), and visual flight rules (VFR). IFRs consist of a set of regulations in which a pilot aviates and navigates by using aircraft instrumentation only. This occurs under instrument meteorological conditions (IMC), which is a term that describes deteriorating weather conditions where the outside visibility of the aircraft is such that pilots must fly primarily by reference to their instrumentation. Additionally, VFR occurs under visual meteorological conditions (VMC), which are a set of aviation regulations that allows the pilot to fly by visual reference to the environment outside the window. In real world settings, both IMC and VMC are environmental influences on workload that manipulate the task demands of the visual processing structure. Consistent with previous literature, blink rates were found to decrease across time during both IFR conditions and high speed IFR conditions compared to VFR flight segments, indicating a higher level of MWL.

Multiple significant comparisons were found in Wilson's blink rate data, but an overall pattern emerged. Wilson's blink rate data suggest that blink rates were less consistent during the middle segments of the task (before and after mid-flight landing, "touch and go"). Blink rate also tended to be significantly higher during VFR conditions compared to IFR conditions. Subjective data measuring MWL indicate lower MWL during the VFR segments and higher MWL during the IFR segments. Subjective ratings were recorded after each of the 22 segments on a scale from 0–100. A rating of zero indicated the lowest MWL and 100 indicated the highest MWL rating.

In their flight simulator task, Veltman and Gaillard (1998) had 12 pilots fly a fixed wing aircraft through a tunnel while performing an auditory based memory task. Both the flight task and memory task varied at four levels of difficulty, and were paired such that the easiest tunnel task and easiest memory task were matched, along with a matching of the difficult tunnel task and difficult memory task. MWL was measured with both subjective assessments and several physiological measures, including eye blinks. During the tunneling task, pilots were asked to pursue a jet at a variable distance (manipulation of tunnel task difficulty) through a tunnel and keep their aircraft in the center of the tunnel. The memory task was a continuous memory task (CMT) that was presented via headphones. Subjects listened for target letters and indicated each letter target or non-letter target by pressing a match or non-match button. Blink data suggested that as more visual information had to be processed, the interval between blinks increased (fewer blinks). However, during an increase in the memory task difficulty, blink interval decreased (more blinks), suggesting that verbal rehearsal tends to decrease blink interval. These data are similar to data reported by De Jong and Merckelback (1990), who found that blinks tend to decrease during auditory "information uptake", while they increase during silent verbal rehearsal.

A study conducted by Tole and colleagues (1983), examined visual scanning patterns, dwell duration, and fixation sequences of both novice and expert pilots. During the flight simulation task, pilots were asked to manually fly (no auto-pilot) the simulator on an instrument landing system (ILS) course. Workload was manipulated by a decision rule algorithm task, during which pilots were to respond to three number sequences. If the first number was largest and the second number smallest, then the pilot was to respond verbally by saying, "plus." If the first number was largest and the last number the smallest, then the pilot response would be, "minus," and so forth. MWL was increased, via temporal loading, by shortening the duration between the audio presentations of the numbers. The participants were, therefore, required to apply the algorithm more quickly, which implies that more effort is required for a task if the task is not well-learned. The purpose of this workload manipulation was to load task demand through a channel unrelated to visual activity in order to monitor scanning patterns affected by a secondary task. Results suggest that fixation durations, defined as the length of time a fixation occurs, for the novice pilots became increasingly long as the period between auditory presentation and responses shortened (increase in MWL). Similar results were reported for the expert pilots, but fixation duration or "staring" was less heavily affected by increases in MWL. This data suggests that fixation durations cannot only detect MWL levels but also operator efficiency, or more automatic processing by the expert pilots of the primary task compared to novice pilots. A summary of ocular activity research related to real and simulated flying is presented in table 1.

Table 1. Ocular activity variables correlated with MWL during real and simulated flight.

	Positively Correlated with MWL	**Negatively Correlated with MWL**	
Authors	**Fixation Duration**	**Blink Rates**	**Blink Duration**
Wilson (2002)		*	
Veltman and Gaillard (1998)		*	*
Tole et al. (1983)	*		

1.7.2 Air Traffic Controllers

A number of other studies show that ocular activity is correlated with MWL during tasks meant to simulate those of an ATC. Ahlstrom and Friedman-Berg (2006) manipulated weather display strategies at three levels (display presented on top of controller workstation, weather presented directly on control display, and no weather display) and two different weather scenarios (higher frequency storm pop-ups versus lower frequency storm pop-ups). MWL was measured via the air traffic control workload input technique (ATWIT; Stein, 1985) that takes a subjective MWL rating of the ATC during the task at set intervals (every five minutes in this case), with a scale from 1 (low workload) to 10 (high workload). Several ocular activity measures were recorded, including point of gaze (POG); blink frequency and duration; saccade frequency and distance; and pupil diameter. Overall subjective workload ratings indicated no significant difference as a function of storm condition or weather display condition, and demonstrated relatively low overall workload levels. However, subjective workload ratings did increase significantly with an increase in the number of aircrafts in the scenario. Ocular activity results indicated that blink frequency did not decrease with aircraft density; however, blink duration consistently decreased across aircraft density. Blink frequency refers to the number of blinks, and blink duration refers to how long the blink lasts. Additionally, there was no relationship between saccade frequency and aircraft density, but saccadic extent reduced significantly with the number of aircraft in the display. It is unlikely that these results are due to tightly packed aircraft because the mean distance between aircraft actually increases significantly on the ATC's monitor screen with respect to aircraft density. This suggests that saccadic extent is related to workload and not just visual target density. Pupil diameter was also found to increase linearly with aircraft density. Although subjective measures were not found to be significantly different across weather display conditions, blink duration was the quickest during the control condition (thought to be the highest MWL condition because of reduced external support) and increased linearly across the workstation and auxiliary conditions. It is possible that these results were found because subjective measures were insensitive to subtle increases in workload. It is also likely that, as

Huey and Wickens (1993) suggest, different measures of workload actually measure different aspects of workload. This would mean that subjective workload measures are possibly less sensitive than physiological or continuous measures. This research demonstrates that some measures of ocular activity may not only provide continuous data about workload levels, but also inform the designer/evaluator what the subject is viewing when MWL transitions occurs.

In a study directed at understanding how ATCs scan and use information displays, Willems, Allen, and Stein (1999) had volunteers participate in a TRACON air traffic control simulation in which workload was manipulated by task load (6 aircraft per 15 min and 12 per 15 min), and type of display screen was manipulated with visual noise (with and without overflying aircraft). This type of manipulation was meant to examine how MWL affects ocular activity while controlling for certain types of visual presentation. Additionally, subjective workload ratings were assessed via the Air Traffic Workload Input Technique (ATWIT) at a rate of nine responses per 45 min scenarios, as well as the TLX to measure post-scenario MWL levels. For the ocular data, measures relating to fixations, saccades, and blinks were recorded. Other dependent variables were measured, such as performance and over-the-shoulder ratings, but the previously mentioned variables are of most importance here. Results indicate that the only independent variable that had a consistent effect on general eye movement was task load. Increasing the task load decreased the fixation duration. Multiple univariate ANOVAs were also conducted on individual scene planes (e.g., radar and keyboard). The overarching results indicate that task load was the major predictor of ocular activity in the scene planes where significant findings were reported. The fixation data suggest that more cognitive processing occurred when subjects directed their attention to the ATWIT device (a subjective MWL measurement device that required input at set intervals) and radarscope, as opposed to the keyboard area and flight strip bay. This suggests that the ATWIT and radarscope require the most cognitively demanding processing. Additionally, both of the subjective workload measure survey scores (TLX and ATWIT) increased as a result of increased task load, verifying that task load was successfully manipulated.

A similar study examined blink activity during an air traffic control simulation (TRACON for windows) by Brookings, Wilson, and Swain (1996). Researchers were interested in determining the differential sensitivity of various workload measures by manipulating air traffic volume and air traffic complexity. Eight subjects completed three ATC scenarios. In the first, MWL was manipulated via traffic volume where 6, 12, and 18 aircraft would enter air space within 15 min. The total scenario lasted 45 min. In the second scenario, air traffic volume was held constant (12 aircraft per 15 min segment) and traffic complexity was manipulated by the ratio of arriving flights to departed flights and over flights. This manipulation was also achieved by having the pilots either not hear or fail to perform their commands properly, as well as by the heterogeneity of the aircraft type. The third scenario, an overload scenario, occurred last when ATCs had to orchestrate 15 aircraft in 5 min. The NASA TLX was used as a subjective measure of MWL and data were recorded for these measures at the end of each segment per scenario. Ocular activity was also collected as one of the physiological measures, which included eye blink rate, saccade rate, and saccade amplitude. Results indicate that TLX scores showed significant increase in task difficulty (low, medium, and high), but no main effect for traffic volume versus complexity. Blink rate data suggest a similar pattern, in that a significant decrease was observed across task difficulty, but no difference was observed between complexity and volume. Saccade measures did not reveal any significant differences between difficulty and traffic manipulation. Blink data were largely explained by the suggestion that operators under heavier task load did not want to occlude objects of important relevance and, thus, they blinked fewer times under higher MWL levels. These data are summarized in table 2.

Table 2. Ocular activity variables correlated with MWL during simulated ATC scenarios.

	Positively Correlated with MWL	**Negatively Correlated with MWL**		
Authors	**Pupil Diameter**	**Blink Rates**	**Blink Duration**	**Fixation Duration**
Ahlstom and Friedman-Berg (2006)	*		*	
Brookings, Wilson, and Swain (1996)		*		
Willems, Allen, and Stein (1999)				*

1.7.3 Dual-Task Scenario

In a dual task of tracking and mental arithmetic, Ryu and Myung (2005) attempted to combine multiple physiological indices in order to determine the mental effort required for each task. Electroencephalogram (EEG), electrooculogram (EOG), and electrocardiogram (ECG) tests recorded alpha rhythm, eye blink intervals, and heart rate variability, respectively. During the tracking task (primary task), participants were required to monitor an attitude indicator programmed in C++, which simulates that of an aircraft using instrument landing on a final approach course in which pilots control the aircraft pitch for glide slope correction. Participants were instructed to ensure that an aircraft icon stayed on a moving horizon bar by manipulating the altitude with a mouse. Difficulty was manipulated at three levels in which velocity of the horizontal bar was selected, averaging 100, 120, and 140 pixels per sec. The arithmetic task (secondary task), manipulated at two levels and presented via audio, consisting of addition problems that were displayed on the screen next to the tracking task. For the low MWL condition, 16 pairs of double-digit numbers were to be mentally computed and responded to via the number pad on a keyboard. None of the numbers had to be “carried over” during the addition process in the low MWL condition. In the high MWL condition 16 pairs of three-digit numbers had to be added, all of which had one number that needed to be carried. The TLX was also given to each subject after all six 7-min task runs, in order to assess the task loading. Tasks were also counter-balanced in order to deter learning effects. Results indicated that blink interval increased with the difficulty of the tracking task, but there was no significant main effect for the arithmetic task and no significant interaction. Ryu and Myung (2005) suggest that these blink data are only an indicator of the load on the visual perception system, since there were no significant differences caused by the arithmetic task which was presented via audio. Table 3 provides a summary of Ryu and Myung's data, along with other ocular activity data collected during dual-task scenarios.

Table 3. Ocular activity variables correlated with MWL during dual-task scenario.

	Positively Correlated with MWL		**Negatively Correlated with MWL**		
Authors	**Blink Interval**	**Fixation Frequency**	**Blink Frequency**	**Blink Duration**	**Saccadic Extent**
Van Orden, Limbert, Makeig, and Jung (2001)		*	*	*	*
Van Orden, Jung, and Makeig (2000)			*		
Ryu and Myung (2005)	*				

1.7.4 Summary of the Mental Workload Literature

Several patterns have emerged from research previously reviewed. First, six of the studies demonstrate correlations between blink activity and MWL. Although the data were measured in terms of intervals (positive correlation), frequency, and rate (both negative correlation), the general finding is that blinks are inhibited under conditions of increased MWL. Studies indicate that these findings can be generalized across both auditory loading tasks (Veltman & Gaillard, 1998) and visual loading tasks (Van Orden, Jung, & Makeig, 1999). Fixation duration data tend to show that under conditions of increased MWL, fixations tend to be longer in duration. Studies from Willems, Allen, and Stein (1999), and Tole et al. (1983) have supported this notion, while Wikman, Nieminen, and Summala, (1998) demonstrate that fixation duration variability increases as a result of MWL manipulations in terms of operator histories (novice and expert drivers). Additionally, pupil size tends to increase under heavily task-loaded situations (Ahlstom & Friedman-Berg, 2006), in terms of how external support mediates MWL. Specifically, the use of a static storm forecast tool compared to a dynamic storm forecast during ATC operations indicated an increase in MWL. Saccadic extent tends to decrease under higher MWL conditions (Van Orden, Limbert, Makeig, & Jung, 2001).

1.7.5 Summary of the Visual Workload Literature

Considerably less research has been conducted on VWL aside from blink interval, which has also been shown as a measure of MWL. Verway and Veltman (1996) found that during a driving task with a visually loaded secondary task, blink interval increased, suggesting that it is related to increased visual demand, and, thus, VWL. Van der Horst (2004) suggested that the rate of visual information requests accurately measures VWL, but no specific eye movement activity was recorded. Saccadic extent and pupil size variability as a measure of VWL are largely related to the assumptions previously mentioned about VWL. Although saccadic extent has been shown as a correlate of MWL (Van Orden, Limbert, Makeig, & Jung, 2001), the visual demands of the task were correlated with the mental demands. Thus, the data are inconclusive regarding their implications for construct clarification in terms of discriminating MWL and VWL. To date, no research has been conducted in which pupil size variability has been measured in terms of visual task loading. However, greater variability indicates a more effortful response.

The goal of this research is to identify possible ocular activity variables that correlate with mental and visual workload independently. Based on the previous research and the assumptions made about VWL, the current research suggests that these ocular activity variables will diverge (convergence will also be tested) in such a way that can be described as a disjoint union set, which can be seen in figure 3.

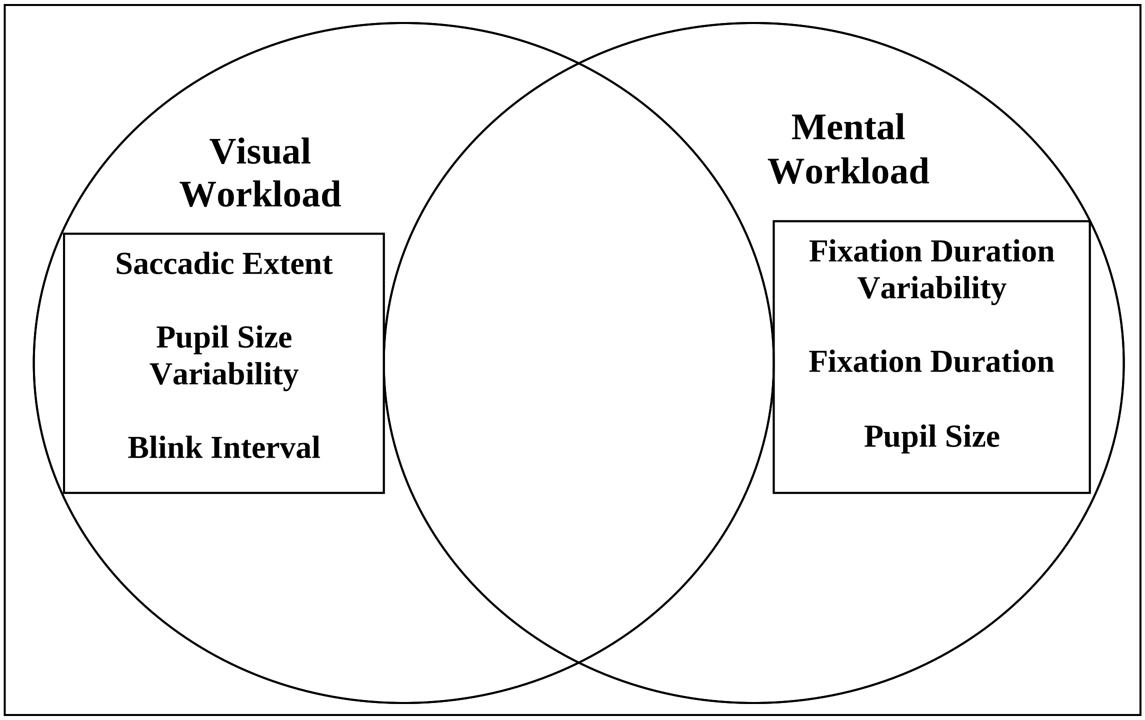


Figure 3. Disjoint union of VWL and MWL ocular activity measure.

1.8 Hypotheses

The following hypotheses are primarily driven by correlations that have previously been reported in the literature and the assumptions that I previously made about VWL. The current study will examine six previously mentioned ocular activity measures as they occur during Task Difficulty and Task Differences during simulated UH-60M upgrade flight scenarios. The general scientific hypothesis is that different ocular activity variables diverge with respect to Visual and Mental Workload. All possible comparisons will also be conducted to examine the convergence of the previously mentioned ocular activity variables, and the alpha inflation will be controlled using Dunn’s test. The experimental hypotheses are as follows.

1.8.1 Hypotheses of Ocular Activity Measures as a Function of Task Difficulty

(1) Fixation duration variability will be greater during the Action on Contact task (high MWL) than during the Hover task (lower MWL). This hypothesis is primarily supported by Wikman, Nieminen, and Summala’s (1998) data, which suggest that novice versus expert drivers differ in their fixation duration strategy that is captured by variability. MWL is mediated by operator histories; thus, this hypothesis will evaluate whether fixation duration variability is capable of capturing MWL that is induced as a result of task loading, but operator history will be held constant by using only expert pilots.

(2) Fixation duration will be greater during the Action on Contact task than during the Hover task. This suggestion is based on previous research demonstrating that as Task Difficulty

increases, the amount of time required to extract visual information is increased (Salthouse & Ellis, 1980).

(3) Average pupil size will be greater during Action on Contact tasks than during the Hover task. Similar to data collected by Ahlstom and Friedman-Berg (2006), pupil size tends to be significantly larger during periods of higher MWL.

1.8.2 Hypotheses of Ocular Activity Measures between Task Differences

Although there is no standardized manipulation check, it is suggested that VWL will be higher for the pilot on-controls compared to the pilot off-controls. The primary responsibility of pilot on-controls is to aviate. Thus, the visual information relevant to his goal is dispersed throughout the environment, both inside and outside the cockpit. This implies that VWL will be higher because more effort is required to obtain visual information that is farther apart compared to visual information that is closely packed. The pilot off-controls, however, tends to be focused in a more localized area, the instrument cluster. I predict the following hypotheses, which are based on the assumption that VWL is higher for the pilot on-controls compared to the pilot off-controls.

(1) Saccadic extent will be significantly greater for the pilot on-controls compared to the pilot off-controls because the pilot on-controls is required to scan a larger viewing field (i.e., outside the window).

(2) Pupil size variability will be greater for the pilot on-controls compared to the pilot off-controls. Although previous research has suggested that pupil size is a valid indicator of MWL, I suggest that variability will capture the effort exerted by the visual system, which is an important component of VWL, because larger deviations from the mean suggest that the visual system is having to adapt to changing state and environmental conditions. The pilot on-controls must switch his visual attention from inside the cockpit to outside the cockpit much more frequently than the pilot off-controls. The luminosity of the inside the window versus outside the window optical array is different (brighter outside the window and darker inside the window), thus, more transitions from inside to outside the window would require the pupil to adapt to changing luminosities more frequently.

(3) Blink interval will be greater for the pilot on-controls compared to the pilot off-controls because the task of the pilot on-controls is much more visually demanding, suggesting that the visual occlusion that occurs during a blink will be inhibited compared to the pilot off-controls. The pilot on-controls task can be likened to a driver versus a passenger task. The primary responsibility of the driver is to maintain eye contact with the road and safely maneuver the vehicle. If he loses eye contact with the road for a relatively short period of time, his safety could be compromised. The passenger, however, is not responsible for constant visual monitoring. Similarly, in a helicopter, the pilot on-control’s primary responsibility is visio-spatial (i.e., aviating) whereas the pilot off-controls tends to perform

secondary tasks related to communication, which are more verbal-linguistic tasks. Thus, the VWL for the pilot on-controls is greater than the VWL for the pilot off-controls.

2. Methods and Procedures

2.1 Participants

Participants (*N*=6) included U.S. Army helicopter aviation pilots flight-testing an upgrade of the UH-60M Black Hawk helicopter. Unfortunately, a larger participant sample was unable to be recruited because of increased demand for highly experienced pilots during the current time of heightened international conflict. Research was conducted on Redstone Arsenal in Huntsville, AL, at the Software Engineering Directorate (SED) in a newly built Systems Integration Laboratory (SIL). All pilots were male (one African American and five Caucasians) with an age range of 28 to 37 (M = 32.16). They were selected as part of a Limited Usability Test (LUT) conducted during the acquisition cycle of the UH-60M upgrade. Data from one of the pilots had to be dropped because the image capturing device was bumped during the flight scenario. All pilots were trained for the UH-60M upgrade during a three-week period prior to data collection, but had no experience with the specific vignettes flown during the experimental period. Pilots were compensated via monetary reward bundled within a yearly salary. The flight hours in a UH-60 A, L, or M model prior to the study for each of the pilots are presented in table 4. The hours are acquired based on pilot duty and are additive. Rows represent each pilot participant.

Table 4. Flight hours per pilot per pilot role.

Instructor Pilot	Pilot	Pilot in Command	Total Hours
0	600	200	800
1200	400	600	2200
0	600	1400	2000
0	740	60	800
42	510	1122	1674
0	750	0	750

2.2 Apparatus

Ocular activity data were collected using an Applied Science Laboratory (ASL) Eye-Trac 6 head-mounted eye tracker and a laser-guided head tracker (LaserBird II) from Ascension Corporation. The laser-guided head tracker is a device used to locate the position of the head in order to obtain point of gaze data with regard to head position. The head-mounted eye tracker contains an infrared camera that records and illuminates the cornea and pupil of the eye. The camera then passes the image to an ASL control box that uses pupil and cornea reflection discrimination techniques based on bright field object recognition, which are implemented in ASL's eye tracker software (Eye-trac 6 .net User Interface Version 1.25.0.1). The distance between the center of the pupil and cornea reflection is then measured during eye rotation, and a corresponding environmental coordinate is associated with each recorded distance. The laser-guided head tracker allows head position to be subtracted from point of gaze data, ensuring that the participant is capable of moving his head freely.

During the installation and set-up procedure, researchers were required to measure the horizontal and vertical distance of areas within the participant station that are of primary importance, also known as scene planes. If using eye-head integration, which was implemented for the current study, the user must specify scene planes in three-dimensional space. This is done by specifying the distance between three points of each scene plane to the laser-guided head tracker (LaserBird II) in three-dimensional space. For the purposes of this study, three scene planes were implemented, which can be viewed on the image in figure 4. These include scene plane zero, the calibration scene plane, and both MFD units, which corresponded to scene plane one and two for each pilot. The outside the window (OTW) scene corresponded to scene plane zero, while the Control Display Unit (CDU) was represented by scene plane three. Scene plane zero is an infinite extension, so it was used to capture data related to observations of the pilot's notepad and controls. A complete eye-head integration system was employed for the pilot seat and the co-pilot's seat. Thus, each system had three scene planes that were essentially mirrored images of each other.

In order to effectively operate these devices, three primary stations must be used—an experimental station, a control station, and a participant station. The experimental station consisted of a computer, an eye monitor, and a scene monitor. This allows the experimenter the ability to run the ASL software, control the eye tracker, and monitor the fixation of the camera on the eye, as well as the visual scene projected at the experimental station. The control station consists of a video and data processing unit, which calculates the point of gaze and pupil diameter. It also sends the video signal of a crosshair representing the point of gaze to the stationary scene camera image, as well as a similarly displayed crosshair outlining the pupil and cornea reflection on the eye image.

Figure 4. Scene plane locations for pilot and co-pilot eye tracking systems.

NOTE: Scene plane zero represents the outside the window view as well as the knee pad and controller. Scene plane one represents the left MFD for both pilots. Scene plane two represents the right MFD for both pilots. Scene plane three represents the CDU for both pilots.

The participant station consists of a 160° flight scenario display panel and cockpit of an actual shell of a downed UH-60M. The cockpit controls contain a high fidelity simulation of a UH-60M upgraded aircraft built to replicate a production quality unit. The two laser-guided head trackers were placed roughly 36 in above the pilot and co-pilot seats, and 48 in apart in order to not interfere with each other. The image from figure 5 provides a visual representation of how the eye tracking system was connected. Between the control station and experimental station, an RS-232 cable connects to the control unit and the experimenter’s serial port of the computer. Additionally, “eye” and “scene” cables connect the control unit to the eye and scene monitors. The control unit and participant station were connected by a bundle of video cables, which carry the eye image to the control unit, and a separate cable that carries the eye camera commands from the control unit to the eye camera. The stationary scene camera, located over each pilot’s shoulder, connects to the control unit and is sent to the scene monitor via the scene cable. A GPS master clock from ESE (model ES-185) was used to attach a time stamp to the video recording in order to sync the video data with ASL's numerical data so that data is accurately sorted during the post hoc analysis. Data were collected using the ASL user interface software and then analyzed using ASL's Eyenal (Version 1.25.0.1).

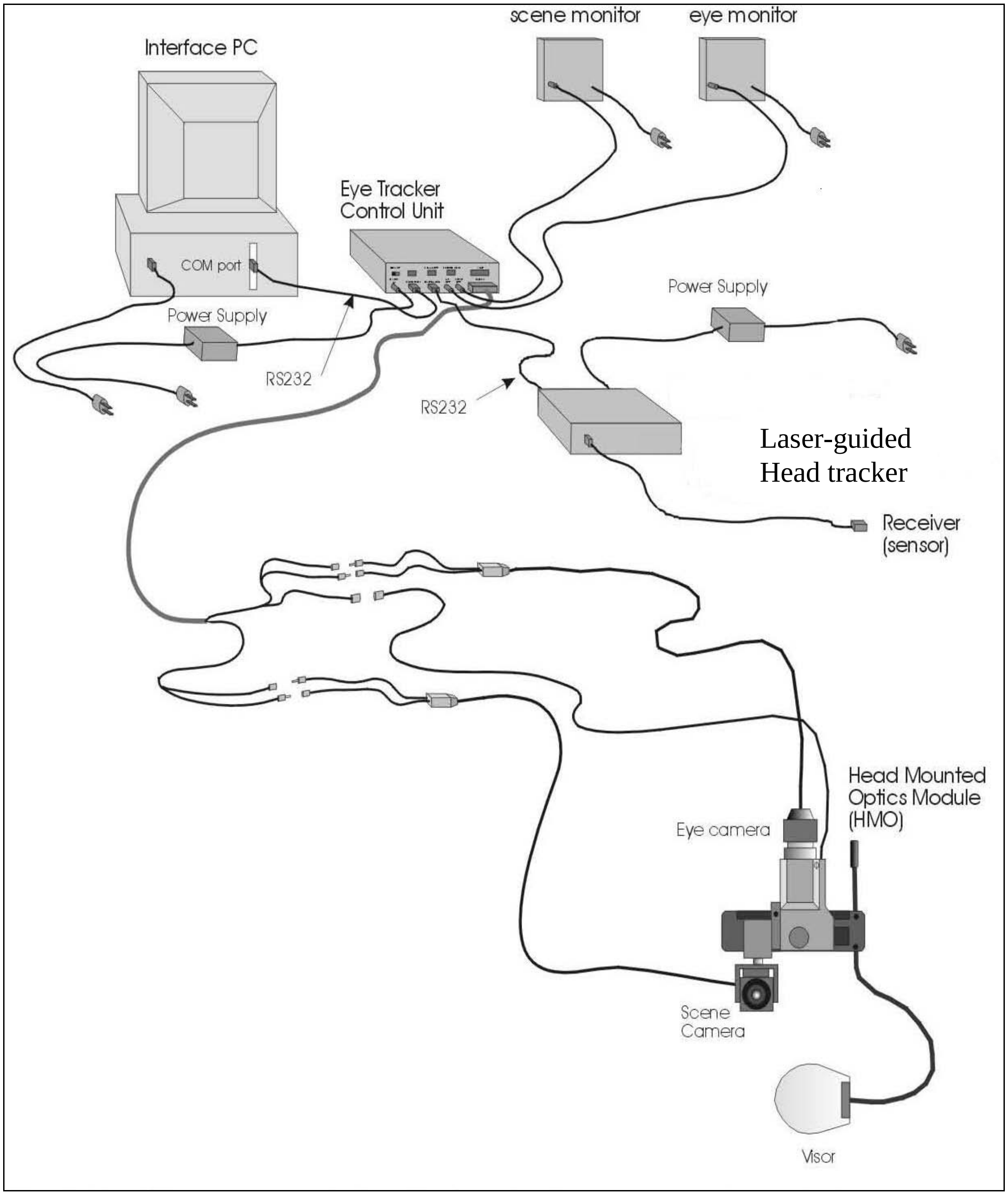


Figure 5. Setup of eye tracker with eye-head integration diagram.

The BWL was also administered at the end of each flight scenario for each ATM task. As previously mentioned, the BWL is a unidimensional measure of MWL, employing a decision tree-based 10-item response scale that assesses operators' spare capacity to perform extra tasks (Roscoe 1984). An image of the scale can be viewed in appendix A. The primary reason that this instrument was selected is because, given the number of tasks that must be rated, it is relatively quick to administer. Additionally, data from Tsang and Johnson (1987), as well as the previously mentioned data, suggest that the BWL is both a reliable and valid measure of operator workload.

2.3 Procedure

After all required consent forms were completed (appendices B and C), participants began a standard training procedure during the month of September 2008 in order familiarize themselves with the new components of the UH-60M upgrade. Once pilot training was completed and all hardware was validated in an airworthiness review conducted by System Simulation and Development Directorate (SSDD), data collection began. During a three-week data collection period (October 6–24), three crews (pilot and co-pilot) flew six simulated flight scenarios in a one week period. Each pilot crew participated in weeklong blocks to ensure that other pilot crews not wait while completing the experiment, thus reducing participant attrition. The flight scenarios were designed by an experienced Black Hawk pilot with 2875 hr as an instructor pilot, 975 hr as a pilot, and 3550 hr as a pilot in command. They were developed in order to mimic flight scenarios that were typical for Black Hawk helicopters (air cavalry and aeromedical evacuations). Before entering the SIL, pilots went through a pre-flight briefing during which the scenario was explained. Once pilots entered the SIL, they were fitted with ASL's eye tracker, calibrated, and began pre-flight mission planning. For each flight scenario, the current investigator was able to collect ocular activity data on three of the six scenarios flown. As an example, a detailed scenario description of flight scenario one is provided in appendix D. Depending on the flight, each scenario lasted roughly 2 hr. During this extended period, pilots were fully occupied and concentrating in a multi-tasked environment. Typically, this type of environment reduces the Hawthorne effect, indicating that the data is reliable even though participants were under constant observation. Pilots were asked to complete mission objectives using the equipment in any way they saw fit and were instructed to fly as if it were a real-life situation. The purpose of this strategy was to increase ecological validity such that data can be generalized to real-world situations. After each vignette, pilots were instructed to complete the Bedford Workload Scale for each ATM task. ATM tasks were chosen by the Army Test and Evaluation Command (ATEC) to evaluate system modifications resulting in the UH-60M upgrade model, as well as their relevance to general performance and specific pilot vehicle interfaces. Once pilots completed the vignettes and ratings, they were debriefed and returned the next day for a subsequent flight scenario.

2.4 Identification of MWL Differences across ATM Tasks and Task Differences

In order to manipulate Task Difficulty, 24 ATM tasks were examined in terms of data sorting logistics, and the nature of each task event that was rated by the pilots on the BWL scale. ATM task 1038 (perform Hovering flight) and ATM task 2042 (perform Actions on Contact) were selected as a planned comparison for two primary reasons. First, they were event specific; second, the nature of those event initiation and termination points were such that non-subject matter experts could consistently identify the start and stop points without prior knowledge about very specific pilot task activity. For example, ocular activity related to hovering consists of all eye movement data that were collected while pilots were hovering, which could be viewed based on a video recording. Although they performed other activities, such as radio communication and navigation, hovering had a distinct beginning and ending period, and ocular activity, therefore, associated with this period was capable of being assigned to this difficulty rating. Similarly, the Actions on Contact task was event-driven and had distinct initiation and ending periods. Other tasks with associated ocular activity, which can be seen in table 5, could not be categorized reliably or without confounding broader situational events.

Table 5. List of all ATM tasks.

1026-Maintain Airspace Surveillance	1176-Perform Non-precision Approach:
1032-Perform Radio Communication Procedures	1178-Perform Precision Approach
1038-Perform Hovering Flight	1184-Respond to Inadvertent Instrument Meteorological Condition (IMC)
1040-Perform visual meteorological condition (VMC) Takeoff	1188-Operate Aircraft Survivability Equipment (ASE)
1046-Perform Electronically Aided Navigation	1253-Operate Central Display Unit (CDU)
1048-Perform Fuel Management Procedures	1254-Operate Multifunction Display (MFD)
1052-Perform VMC Flight Maneuvers	2010-Perform multi-aircraft operations
1058-Perform VMC Approach	2022-Transmit Tactical Reports
1070-Perform Emergency Procedures	2024 Perform Terrain Flight Navigation
1166-Perform Instrument Maneuvers	2026-Perform Terrain Flight
1170-Perform Instrument Takeoff	2042-Perform Actions on Contact
1174-Perform Holding Procedures	2127-Perform Combat Maneuvering Flight

*Only comparison was made between Task 2042 and Task 1038

Sorting ocular activity for other ATM tasks posed several problems. First, tasks such as “maintain airspace surveillance” (ATM task 1026) were too general for specific ocular activity data points to be assigned to that specific task category. During airspace surveillance many different levels of MWL can be experienced because of various broader situational influences. Second, tasks that did have a specific start and stop period, such that ocular activity data could be assigned to that task, did not occur consistently enough to produce reliable data with enough repeated data samples to produce enough power given the low sample size. Third, given the quality of video researchers used to sort ocular activity data, it was impossible to categorize the operation of specific instrument panels into task specific behaviors. For example, the multifunctional displays were used for navigation as well as system monitoring. Thus, it was not possible to categorize specific ocular activity data points as either ATM task 1046 (perform electronically aided navigation) or ATM task 1188 (operate aircraft survival equipment). Thus, the best fit for a Task Difficulty comparison was between ATM task 1038 (M = 1.58) and ATM task 2042 (M = 4.25) because specific ocular activity data points could accurately be assigned to those task difficulties. Results of an *a priori* comparison indicate that scores were significantly different, $t(11) = 8.6$, $p > .05$. The family wise error rate was not adjusted since only one planned comparison was used (alpha set at .05). Additionally, there were an adequate number of task events so that the means of multiple ocular activity variables for each Hover and Action on Contact occurrence provided adequate statistical power given the low sample size. BWL data comparing MWL levels between Task Differences were also collected *post hoc* by experienced pilots to investigate possible differences in task difficulty across the Task Difference variable. Results indicate that there were no significant differences—$t(5) = 2.24$, $p > .05$.

2.5 Aircrew Training Manual Task Descriptions for Hovering Flight and Action on Contact

The ATM for the UH-60 Series Helicopter describes the crew actions for the pilot on-controls and pilot off-controls during specific flight tasks. During a Hovering flight (Task 1038), the pilot on-controls is to announce his/her intent to perform the Hovering maneuver and focus their visual attention primarily outside the window in order to monitor altitude and avoid obstacles. The pilot off-controls is to assist in clearing the aircraft, report unannounced drift, and warn the pilot on-controls of altitude changes and obstacles. Actions on Contact (Task 2042) are required when the helicopter's threat detection system announces, via auditory alert, that the vehicle is within tracking and/or firing range. This is generally considered a period of high MWL because the operators must quickly and safely maneuver the helicopter to a concealed position. During this task, the pilot on-controls is to remain primarily focused outside the window to avoid obstacle collision and generally reposition the aircraft as necessary in order to break radar or visual contact to avoid the threat. The pilot off-controls assists in reporting obstacle avoidance and transmits tactical reports. Thus, across both tasks, the pilot on-controls must remain primarily focused outside the window in order to safely and effectively maneuver the aircraft, while the pilot off-controls tends to be focused more on the instrument panel. This suggests that the previous data collected on pilots during a flight simulation likely generalizes in the current study, while previous data collected on display driven primary tasks, such as ATCs and dual task scenarios, generalizes better to the pilot off-controls, since their attention tends to be more focused on multiple displays.

2.6 Data Sorting and Reduction Strategy

Ocular activity data were collected throughout the entirety of the flight scenarios. However, only the ocular activity that occurred concurrently with the Hover and Action on Contact tasks was used in the data reduction process. In order to sort these data, a recording of the flight scenarios was observed so that the beginning and ending time of each of the tasks was recorded. A Global Positioning System (GPS) master clock was used to place a time stamp on each of the video images that could be matched within 1 s to the ASL eye tracking data, which also had a corresponding time stamp per data record. For the Hover task, researchers recorded the moment that each crew for each flight vignette took off into a hover maneuver and landed. During this period, pilots typically performed a hover check in which they simply lifted off the ground to verify that the hover controls were operating successfully. All the ocular activity data that occurred while the pilots were flying under hover mode was categorized as ATM task 1038, which had the lowest average BWL score. In order to account for order effects across flight scenarios, the pilots were to provide responses to the BWL on a task-by-task basis, regardless of flight scenario, such that they were rating the task and not necessarily the variety of situations that the task could occur. To identify ocular activity data associated with ATM task 2042, researchers observed the recorded videos and listened for an automated verbal alert that indicated the aircraft had been spotted by an enemy threat. Once a lock of the helicopter was obtained by the threat and the auditory threat alert system notified the pilots, they were to aviate to a location beyond the detection radius of the threat. An auditory verbal alert then indicated to the pilots that the lock had been broken. Similar to the previous data sorting strategy, researchers recorded the beginning and ending time of each Action on Contact occurrence. The target acquisition, indicated by the auditory alert, was used to signal the beginning of the Action on Contact response, while the "target broken" auditory alert was used to signal the end of the response. The corresponding segment of ocular activity data was then categorized as ATM task 2042 data.

After all the raw ocular activity data were sorted into their corresponding task difficulty cell, a fixation file was then produced for each block of data using ASL's default fixation algorithm parameters (Applied Science Laboratory Eyenal Manual). The fixation algorithm is a *post hoc* data analysis strategy that uses a moving window technique to calculate the standard deviation of a set of raw point of gaze data records. Once the standard deviation falls within 0.5° of visual angle, a fixation is begun and a temporary fixation point is calculated. After that, all records that fall within 1° of the mean point of gaze position of the original records that started the fixation are included in the final fixation calculation position. If less than three consecutive data records fall outside of one degree of visual angle, but within 1.5°, then those records are also used in the final fixation calculation. If three consecutive records fall outside of 1.5° of the temporary fixation location, the fixation is terminated. However, if less than three consecutive records fall outside of 1.5° of the temporary fixation location, they are discarded, but the current fixation continues. In order to maintain a proper perspective of resolution, it is important to note that each raw data record represents 0.016 s, or roughly 60 records per second. Thus, if any records are discarded, it almost always represents a very small portion of data typically due to very brief periods of inaccurate pupil and cornea reflection discrimination that would indicate an impossible eye rotation.

Once the fixation file for each segment of data corresponding to its respective Task Difficulty and Task Difference cell was created, the mean saccadic extent were calculated for each segment of data. The average fixation duration and pupil size were also calculated. Additionally, fixation duration variability was calculated via the standard deviation of each data segment, along with pupil size variability. The blink interval was calculated by dividing the total number of seconds by the frequency of blinks per data segment.

3. Results

A series of paired samples, *t*-tests, and paired samples, *t*-tests of variance, was conducted to evaluate how six ocular activity measures (average fixation duration, average pupil size, fixation duration variance, pupil size variance, average saccadic extent, and average blink interval) relate to Task Difficulty and Task Differences of pilots while operating a UH-60M simulator. A paired samples *t*-test of variance was conducted on fixation duration variance and pupil size variance. The *t*-test of variance is used to test if two variance parameters are equal (Glass & Hopkins, 1995). Essentially, it factors out of the standard error term, a non-zero correlation coefficient, since paired scores were expected to be correlated. Across Task Difficulty, it was expected that fixation duration, pupil size, and fixation duration variability would increase for greater task loaded events. Results indicate that the mean fixation duration variability was significantly higher for ATM task 1038, perform Hovering flight ($M = 0.25$, $SD = 0.11$), than for ATM task 2042, perform Action on Contact ($M = 0.15$ $SD = 0.05$), $t(13) = 1.86$, $p < .05$, which was the opposite of what was previously hypothesized. All other results indicate that there were no statistically significant differences for all other ocular activity comparisons across Task Difficulty. Average fixation duration comparisons yielded $t(13) = 0.68$, $p > .05$; saccadic extent comparison yielded $t(13) = 1.75$, $p > .05$; overall pupil size variability comparison yielded $t(13) = 0.16$ $p > .05$; average pupil size data yielded $t(27) = 0.75$, $p > .05$; and blink interval comparison yielded $t(13) = 0.74$, $p > .05$.

In order to evaluate which ocular activity measurements varied significantly as a function of Task Differences (pilot on-controls versus pilot off-controls), another series of paired samples *t*-tests was conducted and the family wise error rate was controlled using the Bonferroni *t*, which adjusts the alpha level to the equivalent .05 probability. As previously hypothesized, blink interval, pupil size variability (paired samples *t*-test of variance conducted), and saccadic extent were predicted to be greater for Task Differences related to higher visual demands (pilot on-controls). A significant difference showed that the saccadic extent was significantly lower for the pilot on-controls ($M = 8.29$, $SD = 1.73$) versus the pilot off-controls ($M = 11.21$, $SD = 3.88$), $t(13) = 2.31$, $p = .038$ (see table 6). Blink interval was also found to significantly increase for the pilot on-controls ($M = 8.11$, $SD = 5.10$) compared to the pilot off-controls ($M = 2.84$, $SD = 1.24$), $t(13) = 4.09$, $p = .001$. As is shown in table 7, all other comparisons of ocular activity measures between Task Difficulty and Task Differences were not statistically significant. Additionally, no ocular activity variables were significant across both independent variables.

Table 6. Means for significantly different ocular activity values across task difficulty and task differences.

	Task Difficulty		Task Difference	
Ocular Variable	**Hover**	**Action on Contact**	**On-Controls**	**Off-Controls**
Fixation Variability	0.25 *s*	0.15 *s*	—	—
Inter-Fixation Deg.	—	—	8.29 *deg*	11.21 *deg*
Blink Interval	—	—	8.11 *s*	2.84 *s*

* *s* = seconds and *deg* = degrees of visual angle.

Table 7. Summary of ocular activity *T*-test statistics.

	Task Difficulty			Task Difference		
Ocular Variable	*df*	*t*	*p*	*df*	*t*	*p*
Fixation Variability	13*s*	1.86*s*	*< .05	13	1.07	> .05
Pupil Variability	13	0.16	> .05	13	1.43	> .05
Inter-Fixation Deg.	13	1.75	> .05	13	2.31	*< .05
Blink Interval	13	.74	> .05	13	4.09	**< .001
Fixation Duration	13	.68	> .05	13	.26	> .05
Pupil Size	27	.75	> .05	23	.75	> .05

* = significance at the .05 level and ** = significance at the .001 level

4. Discussion

4.1 Summary of Results

One of the primary motivating factors of this research was to attempt to discriminate VWL and MWL by examining how ocular activity variables change across Task Difficulty while holding Task Differences constant, and separately, by holding Task Difficulty constant across Task Differences. Data indicate that fixation duration variability is sensitive to Task Difficulty, while saccadic extent and blink interval are sensitive to differences in visual demands, as indicated by variations in Task Differences. Specifically, fixation duration variability was greater during the Hover task compared to the Action on Contact task; saccadic extent was greater for the pilot off-controls compared to the pilot on-controls; and blink interval was greater for the pilot on-controls compared to the pilot off-controls. Thus, only one of the previously suggested hypotheses was confirmed—namely, that blink interval was significantly greater for the pilot on-controls compared to the pilot off-controls. All other hypotheses were statistically unsupported, although fixation duration variability and saccadic extent demonstrated statistical relationships to Task Difficulty and Task Differences, respectively, but in the direction opposite than expected. The following discussion will present the Visual and Mental Workload constructs as a disjoint union, with further analysis of the relationship between each ocular activity variable and its respective correlate.

The current data set suggests that Visual and Mental Workload can be represented as converging constructs, as illustrated in figure 6. Although none of the variables converged in the present study, blink data has been demonstrated as sensitive to visual and linguistic information processing structures, indicating that it can be used as a measure of Visual and Mental Workload. Saccadic extent varied across Task Differences—a variable indicating difference in visual demands—while showing no significant differences across Task Difficulty. Given that there was no Task Difficulty difference across Task Differences, as indicated by the BWL data collected on Task Differences, it is suggested that saccadic extent varied as a result of differences in VWL. Fixation duration variability varied as a function of Task Difficulty and remained relatively constant across Task Differences, indicating that it is a sensitive measure of MWL. Blink interval data are slightly more complex to interpret, however, and the relationship between each variable and its respective construct will now be discussed.

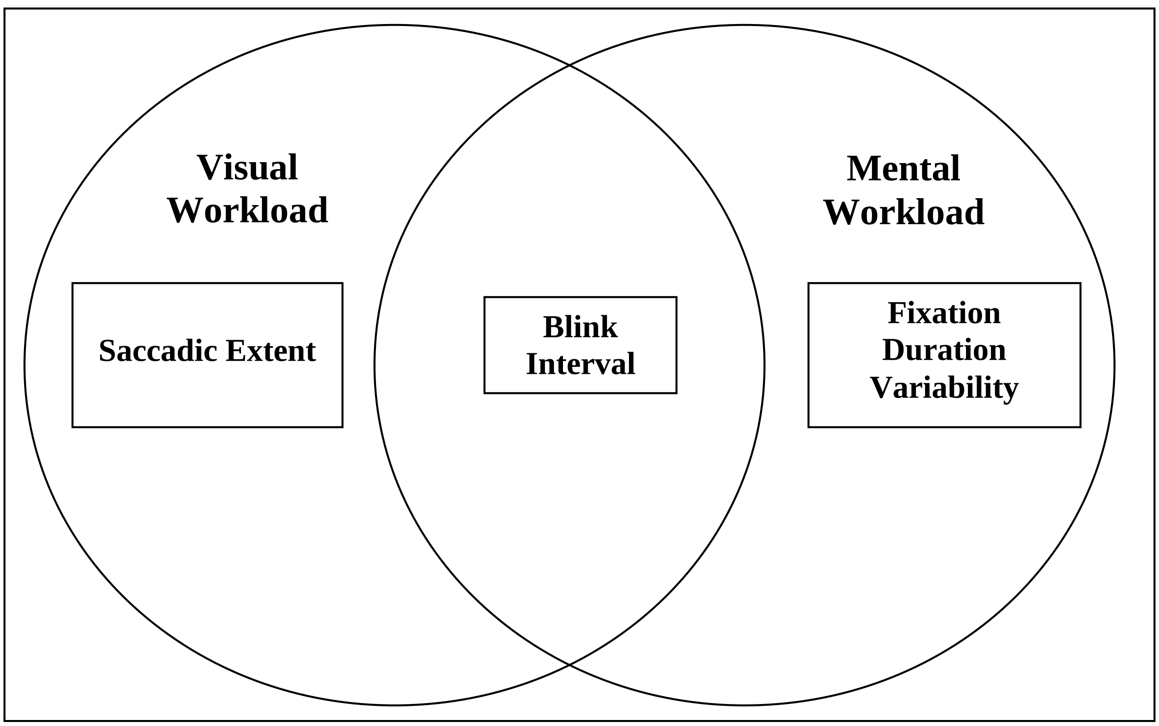


Figure 6. Disjoint union of VWL and MWL ocular activity measures.

As mentioned in section 1, VWL is primarily tied to the amount of effort the visual system exerts in order to access visual information, while MWL refers to the amount of "cognitive effort" used for general task completion. The ocular activity metrics identified to predict VWL are primarily meant to measure the effort component of the visual system. The variables identified as a measure of MWL represent the metrics that reflect visual information processing and are driven by resource demands of the task, as well as the availability of resources within the individual (operator history) and resources that are external to the individual (external support). Thus, VWL is effort used during the *acquisition* of visual information, and MWL, as it is measured via ocular activity, represents the *servicing* of visual information. How these measures relate to different perceptual and cognitive processing demands is a particularly challenging problem, considering that some ocular activity metrics may be sensitive to auditory and linguistic tasks as well as visio-spatial tasks. For example, De Jong and Merckelback (1990) report that blinks increase during verbal rehearsal tasks (auditory-linguistic task), while other researchers that have been mentioned in the earlier review generally find that blinks are inhibited during visually demanding tasks. To date, no researchers have systematically examined several ocular activity measures as they relate to task loading on different perceptual and cognitive processing structures (e.g., visio-spatial versus audio-linguistic), but the current research attempted to identify eye metrics that are sensitive to perceptual effort (VWL) and general cognitive processing (MWL). However, in a multi-tasked environment, the operator is constantly time-sharing between auditory and visual perception. Thus, future research should attempt to further control the perceptual and cognitive processing structures associated with various eye metrics.

Data from the current study support the illustration shown in figure 6, but are based on a very similar but more abstract model of general perceptual and processing workload. As is shown in figure 7, the left hemisphere represents the effort required to acquire perceptual information and can relate to any perceptual structure, while the right hemisphere represents the resources used to service the information. Based on this model, VWL is the limiting factor in terms of MWL that results from servicing only visual information. For example, if one is to perform a digit span test in which the numbers are quickly displayed and dispersed across a large visual array, and the participant cannot acquire the visual information, they will be unable to rehearse or service the information. Thus, saccadic extent provides a facet of information about VWL that can serve as a limiting operator constraint if the availability of spare visual capacity reaches a point of overload.

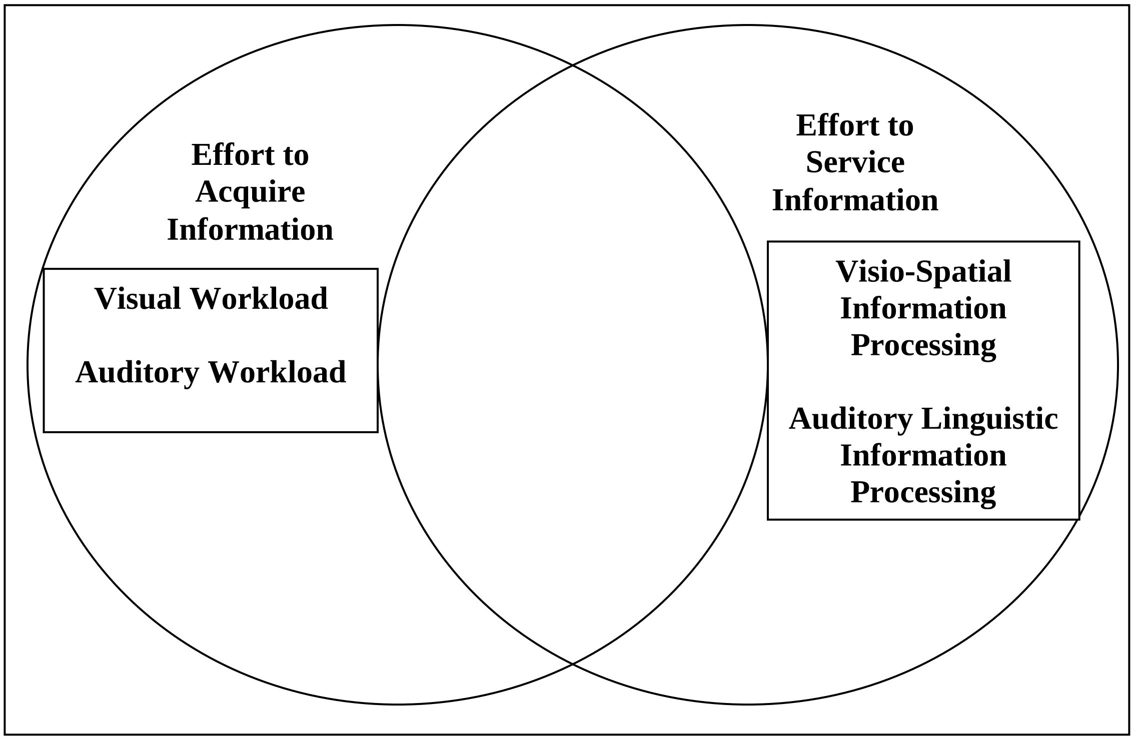


Figure 7. Workloads associated with information acquisition and information servicing.

4.2 Saccadic Extent

The relationship between saccadic extent and pilot Task Differences suggests that the visual information relevant to aircrew performance is dispersed across a larger array for the pilot off-controls. This primarily has implications for the evaluation of spare visual capacity. In the research by Wickens (1979), MWL was defined in terms of secondary task performance, suggesting that processing resources, such as visual processing, are demanded by a primary task to the extent that the performance of a secondary unrelated task deteriorates performance. Typically, the eye will need to exert more effort to perform a saccade between two points that are farther apart in distance compared to a movement between two points that are closer together. If an operator must share tasks that both require visual input, then they must change the velocity of their eye movement in order to acquire visual information at a quicker rate. Future studies should investigate saccadic velocity as another indicator of VWL, but the current study suggests that as visual information is dispersed, indicated by larger saccadic extent, the spare visual capacity of the operator is hindered because they are required to move more quickly to access the information at the same rate compared to visual information that is closer together.

For a task such as driving, it has been suggested that vision is the primary source of information-processing input (Sabey & Staughton, 1975). Given the importance of this processing structure, the current data on saccadic extent provide preliminary evidence that the pilot off-controls exerts more visual effort than the pilot on-controls. The pilot off-controls performed consistently longer saccades, which could indicate two things in regards to VWL. First, the VWL experience by the pilot off-controls may have been significantly greater than the pilot on-controls. Taken at face value, many would likely argue that the VWL demands of the pilot performing the primary aviation tasks (pilot on-controls) are more visually demanding than the pilot off-controls, given that the secondary task of the pilot off-controls is to monitor the outside the window view as support to the pilot on-controls. However, one of the primary features of the UH-60M upgrade is a flying system referred to as “fly-by-wire,” which allows pilots to automate much of the mechanical aviation process under conditions such as hovering. Although data were collapsed across both Task Differences conditions (Hover with automated flight controls and Actions on Contact with manual flight controls), this may have greatly reduced the VWL typical of the pilot on-controls, which would explain the reduced VWL experienced by the pilot on-controls compared to the pilot off-controls. This would be similar to suggesting that a driver’s passenger who is in charge of navigation experiences a higher level of VWL because their saccades tend to transition from deep within the interior—perhaps a map on their lap—to outside the window. This type of visual transition would tend to be a much greater distance than drivers’ saccadic extent between the outside the window view and the rear view mirror, as an example. However, a metric such as saccadic velocity paired with head movements would provide further insight in terms of the rate of information acquisition, as opposed to just the distance the operator travels between bits of information. Additionally, it is estimated that over 90% of the information required for driving is visually oriented (Sabey & Staughton, 1975). Thus, previous research suggests that the visual demands of the driver are much higher than a passenger performing secondary tasks, similar to that of an aircrew flying a helicopter. This observation leads to the second possible explanation of the reported saccadic extent data.

The greater saccadic extent observed from the pilot off-controls may be a result of excess visual resources. The driver in the previous example must stay fixated on important visual information nearly all the time in order to successfully operate the vehicle. The distance of a saccade is highly influenced by the location of important visual information, which suggests that VWL, like MWL, is also a multi-faceted construct that is influenced by more than just the distance the eye moves, but also by the location and quality of important visual information. Further research should be conducted in order to investigate the expected value of visual information in order to quantify the importance of a saccade. This would provide designers information about how detrimental it is to place certain types of information outside the scope of an operator’s close visual range.

This explanation indicates a possible limitation of the study resulting from a limited number of levels of the Task Difficulty variable. The saccadic extent data set may suggest that the pilot off-controls had enough spare visual capacity to monitor a larger visual area. If VWL converges with MWL, as measured by the BWL, then relatively depressed scores on the BWL, which were obtained, may indicate that spare visual capacity was also available to the operator. This suggests that the pilot off-controls performed longer saccades that were outside of their immediate instrument cluster, indicating that they had the spare visual capacity to roam from their primary source of visual information. In order to confirm this suggestion, further studies must be conducted in which VWL is examined during known under-load periods, medium load periods, and over-load periods. This would provide a more complete profile of saccadic behavior as it relates to a larger visually loaded range of situations.

Uncontrolled and systematic changes in the coupling of anthropometry and cockpit design may also serve as a confounding variable to the saccadic extent because it is measured in terms of visual angle with unconstrained head rotation. It is possible that the visual information relevant to the pilot off-controls is only slightly further apart, resulting in the 2.92° mean difference because of slight but systematic differences in seating position. For example, if the location of the head is slightly closer to two fixated points, then the saccadic extent will be calculated as greater compared to when the head is farther away.

In terms of the six measurement criteria previously mentioned, the current data suggest that saccadic extent is diagnostic of VWL because it varied as a function of Task Differences and not Task Difficulty. If saccadic extent were statistically significant across both groups, the current data set would not suggest diagnosticity as a measurement property of saccadic extent. The sensitivity of saccadic extent is only partially determinable based on the current data set. Results suggest that a small (2.92°) but statistically significant difference between pilot on and off the controls indicate that the measure is capable of being sensitive. In order to assess the sensitivity of a measurement, there must be known, but subtle differences that the measure is capable of discriminating. Further evaluation, where the distance between visual information is systematically controlled, would allow researchers to assess the sensitivity of saccadic extent as a measure of VWL. Conclusions about the reliability of saccadic extent cannot be gathered since each flight scenario was only flown once by each pilot. The implementation requirements, intrusiveness, and operator acceptance are all comparable for saccadic extent, fixation duration variability, and blink interval. The implementation requirements are much more time consuming than subjective scales, but eye tracker data are still capable of being collected with only one dedicated technician. The time required for accurate calibration of a single pilot lasts roughly 5 min, depending on individual differences regarding ocular anatomy. Current eye tracking systems tend to be relatively unobtrusive since point of gaze data can be calculated with regards to head position, also known as eye-head integration, which allows for free head movement. Operator acceptance is often related to the intrusiveness of a measurement, but no specific subjective scales were used to assess operator acceptance.

4.3 Fixation Duration Variability

Fixation duration has been shown to relate to information processing (Salthouse & Ellis, 1980), but the only data collected on fixation duration variability have varied by level of experience. The current study held level of experience constant but still found variability across Task Difficulty. This indicates that fixation duration variability may also be a sensitive measure of MWL, as it is manipulated by task demands and not only operator history, as reported by (Wikman, Nieminen, & Summala, 1998). The diagnostic value of fixation duration variability is indicated similarly to saccadic extent. Fixation duration variability was statistically different across the Task Difficulty variable, but not the Task Difference variable, indicating that it only changes as a function of MWL and not VWL. The other four measurement criteria implications are similar to the conclusions draw in regards to saccadic extent.

Kundel and Nodine (1978) made the distinction between *survey* fixations and *examination* fixations. Survey fixations tend to be quicker in duration and bring the observer into the region of interest that contains the target, while examination fixations tend to be much longer and provide the observer more detail about the fixation point. Kundel and Nodine suggest that this is greatly affected by the difficulty of extracting information. A lower fixation duration variability during the Action on Contact event demonstrated that the operator altered their scanning strategy to fixation durations that tended to be more similar to one another during the high workload event. Given the nature of the Action on Contact event, in which the operator must quickly search the visual field for the threat location, it is likely that quick survey fixations were used most often. During the Hover task, however, survey and examination fixations were used. These results may also be related to relatively low overall BWL scores.

The highest MWL score in the current study was relatively low compared to the highest possible score, which indicated on the BWL that there is “Insufficient workload capacity for easy attention to additional tasks.” Although this score is significantly greater than the scores recorded for the Hover task, they are still low compared to higher possible scores on the BWL that indicate an inability to perform primary or secondary tasks. Similar to an inverted U function of MWL and performance, MWL can actually improve performance relative to very low task demand situations. Thus, one possible explanation for the fixation duration variability is that as MWL reaches an optimal level for performance, fixation duration variability becomes smaller because the operator reaches an optimal level of capacity to extract visual information strategically. In order to test this, further data would need to be collected where a broader range of MWL levels were evaluated, similar to the previous suggestions made regarding VWL.

4.4 Blink Interval

The data on blink interval have several implications. First, as reported by Verway and Veltman (1996), blink interval may be sensitive to subtle changes in MWL, but incapable of capturing larger magnitudes of Task Difficulty changes. Similar results were found in the present study because of the nature of the different tasks for which MWL data were collected. This would suggest that there are subtle changes in MWL between the pilot on-controls compared to the pilot off-controls that blink interval data is sensitive to, but were not captured by the BWL between Task Differences. Additionally, Veltman and Gaillard (1998) found that blink interval increased as more visual information had to be processed, but during a memory task, blink intervals decreased. This suggests that blink interval is diagnostic with respect to operator processing resources (visual versus cognitive), and the primary reason blink interval is placed in the center of the disjoint union in the previous figure. Although the current data set only provides evidence that blink interval varied across Task Differences, other researchers have found that blink interval is sensitive to MWL as well as VWL (Casali & Wierwille, 1984; Wierwille & Eggemeier, 1993). Future studies may look to examine the relationship between multiple blink variables, such as interval and duration, and tasks with more controlled visual and mental components.

The reason blink interval is a sensitive measure of VWL may be because of the visual occlusion that occurs during a blink. Although the average blink duration is only 120 ms. (Tecce, 1992), the visual scene is briefly occluded during this period. As just mentioned, researchers have noted that the duration of blinks decreases with increasing visual demands, which suggest that blink activity is highly related to the visual connection between the operator and their visual tasks. Another reason that blink activity captures aspects of Visual and Mental Workload is that both tend to be highly correlated. During tasks such as aviating or driving, in which successful operation is closely tied to the visual system, primary task difficulty increases that are simultaneous with MWL and VWL increase are almost always correlated in some way. For example, a driver on a crowded highway experiences increased MWL because he has to concentrate on multiple sources of variability, while at the same time VWL is increased because he must attend to the visual stimuli more rigorously.

4.5 Discussion of Non-Significant Results

Results demonstrate that no ocular activity variables that were measured in this study converged with both Visual and Mental Workload. As previously mentioned, blink interval was suggested as a convergent measurement primarily on the basis that other studies have shown it to be capable of measuring cognitive processing, while the current data set suggest it is capable of measuring VWL. Average fixation duration data were not statistically significant across both independent variables. Although other researchers have demonstrated a statistically significant relationship (Salthouse & Ellis, 1980) in experimental circumstances, the current results suggest that they do not generalize to a more naturalistic multi-tasked field test environment. This is likely due to the constantly changing nature of the visual information and its value. Pupil size was also not statistically significant, yet previous research has demonstrated a relationship between pupil dilation and MWL (Ahlstom & Friedman-Berg, 2006). It is possible that a Type II error may have emerged for three reasons. First, the luminosity of the environment was not controlled, which may have caused pupil size changes across Task Difficulty events that hid the relationship between MWL and pupil size. Second, pupil size may not be a sensitive measure of MWL during relatively low task loading periods, which were indicated by the BWL data. Third, pupillary responses can occur for many different reasons not directly related to MWL. For example, Lownstein and Loewenfeld (1964) noted that pupil size decreases with fatigue, while Tryon (1975) surveyed sources of pupil variation and discussed that commonly used substances, such as caffeine, can cause changes in the pupillary response. The current study did not control these sources of variability, which may have produced a Type II error.

5. Conclusions

MWL is a multifaceted construct that requires several measures in order to gain a complete understanding of the effortful processes involved in complex human-systems operations. The purpose of measuring VWL is three-fold. First, it allows developers and evaluators the ability to gain information about the specific perceptual system that is being overloaded or has spare capacity under certain task constraints, while concurrently providing information about more general aspects of workload (MWL), since certain measures of VWL converge with MWL. Second, it can inform the designer of specific aspects of the system that may need to be altered. For example, if a driver uses a vehicle telematics system with a visual display, evaluators may find that it draws the operator’s visual attention away from the road and reduces safety. In order to account for this, the designer may implement an auditory system that alerts the driver of important navigation events that greatly reduce the need to focus his/her visual attention on the display, thus offloading VWL onto the auditory system. Essentially, this means that measuring VWL provides information to designers and evaluators about the available capacity of the visual system, which can be used to inform design decisions and not purely diagnostic. Third, I maintain that measuring VWL can provide construct clarification within the broader workload domain. The term MWL, or even more generally, workload, is often used to characterize task demands (difficulty, number, rate, or complexity of demands), the level of performance an operator is capable of achieving, an operator’s response to task demands—as opposed to the demands directly—or an operators perceptions of task demands (Huey & Wickens, 1993). Although arguably non-parsimonious, measuring the VWL construct separately from the broader workload domain would provide more specific and clear data regarding the nature of systems under evaluation, as well as design implications.

6. References

Ahlstrom, U.; Friedman-Berg, F. J. Using Eye Movement Activity as a Correlate of Cognitive Workload. *International Journal of Industrial Ergonomics* **2006**, *36*, 623–636.

Backs, R. W.; Walrath, C. L. Eye Movement and Pupillary Response Indices of Mental Workload During Visual Search of Symbolic Displays. *Applied Ergonomics* **1992**, *23* (*4*), 243–254.

Baddeley, A. D. *Working Memory*; Oxford: Clarendon Press, 1986.

Beatty, J. Task-evoked Pupillary Responses, Processing Load, and the Structure of Processing Resources. *Psychological Bulletin* **1982**, *91* (*2*), 276–292.

Brookings, J. B.; Wilson, G. F.; Swain, C. R. Psychophysiological Responses to Changes in Workload During Simulated Air Rraffic Control. *Biological Psychology* **1996**, *42*, 361–377.

Casali, J. G.; Wierwille, W. W. On the Measurement of Pilot Perceptual Workload: A Comparison of Assessment Techniques Addressing Sensitivity and Intrusion Issues. *Ergonomics* **1984**, *27*, 1033–1050.

Cook, R. I.; Woods, D. D. Operating at the Sharp End: The Complexity of Human Error. In M. S. Bogner (Ed.), *Human error in medicine* (pp. 255–301). Hillsdale, NJ: Erlbaum (1994).

Cooper, G. E.; Harper, R. P. The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities (AGARD Report 567). London: Technical Editing and Reproduction Ltd, 1969.

Corwin, W. H.; Sandry-Garza, D. L.; Biferno, M. H.; Boucek, G. P.; Logan, A. L.; Jonsson, J. E.; Metalis, S. A. Assessment of Crew Workload Measurement Methods, Techniques and Procedures. *Process, Methods, and Results* (WRDC-TR-86-7006). Wright Patterson Air Force Base, Ohio (1989).

De Jong, P. J.; Merckelbach, H. Eyeblink Frequency, Rehearsal Activity, and Sympathetic Arousal. *International Journal of Neuroscience* **1990**, *51*, 89–94.

Dodge, R. Visual Perception During Eye Movement. *Psychological Review* **1900**, *7*, 454–465.

Edland, A.; Svenson, O. Judgment and Decision Making Under Time Pressure. In O. Svenson & J. Maule (Eds.). *Time pressure and stress in human judgment and decision making.* (pp. 27–40). New York: Plenum, 1993.

Glass, G. V.; Hopkins, K. D. *Statistical Methods in Educational Psychology*; Needlham Heights, MA: Simon & Schuster, 1995.

Hart , S. G.; Staveland, L. E. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload*, (pp. 139–183). Amsterdam: North Holland, 1988.

Hancock, P. A.; Desmond, P. A. *Stress, Workload and Fatigue*; Mahwah, NJ: Lawrence Erlbaum, 2001.

Hendy, K. C., Liao, J.; Milgram, P. Combining Time and Intensity Effects in Assessing Operator Information-processing Load. *Human Factors* **1997**, *39* (1), 30–47.

Huey, B. M.; Wickens, C. D. *Workload Transition: Implications for Individual and Team Performance*; Washington, DC: National Academy Press, 1993.

Jex, H. R. Measuring Mental Workload: Problems, Progress, and Promises. In P. A. Hancock & N.Meshkati (Eds.), *Human Mental Workload*, (pp. 5–39). Amsterdam: North-Holland, 1988.

Kahneman, D.; Beatty, J. Pupillary Response in a Pitch Discrimination Task. *Perception and Psychophysics* **1967**, *2*, 101–105.

Knust, S. R.; Marshal, S. P.; Ishazaka, K. Using Eye Activity to Study Cognitive Processes Underlying Individual Differences. *Proceedings of the IEA 2000/HFES 2000 Congress*, 2000.

Kramer, A. Physiological Metrics of Mental Workload: A Review of Recent Progress. In D. Damos (Ed.), *Multiple task performance,* (pp. 279–328). London: Taylor & Francis, 1991.

Kramer, A. F.; Coyne, J. T.; Strayer, D. L. Cognitive Function at High Altitude. *Human Factors* **1993**, *35* (*2*), 329–344.

Land, M. F.; Lee, D. N. Where We Look When We Steer. *Nature* **1994**, *369*, 742–744.

Lowenstein, O.; Loewenfield, I. E. The Sleep-waking Cycle and Pupillary Activity. *Annals of the New York Academy of Sciences* **1964**, *117*, 142–156.

Marshall, S. The Index of Cognitive Activity: Measuring Cognitive Workload. *Proceedings of the IEEE Conference on Human Factors and Power Plants* (*7-5—7-9*) 2002.

O'Donnell R. D.; Eggemeier F. T. Workload Assessment Methodology, In K. Boff, L. Kaufman and J. P. Thomas (Eds.), *Handbook of Perception and Human Performance,* (pp. 42.1–42.49). New York: Wiley, 1986.

Orlady, H. W.; Orlady, L. M. *Human Factors in Multi-crew Flight Operations*; Aldershot, UK: Ashgate, 1999.

Raby, M.; Wickens, C. D. Strategic Workload Management and Decision Biases in Aviation. *The International Journal of Aviation Psychology* **1994**, *4* (3), 211–240.

Reid, G. B.; Nygren, T. E. The Subjective Workload Assessment Technique: A Scaling Procedure for Measuring Mental Workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload,* (pp. 185–218). Amsterdam: Elsevier, 1987.

Roscoe, A. H. Assessing Pilot Workload in Flight: Flight Test Techniques. *Proceedings of NATO Advisory Group for Aerospace Research and Development (AGARD) (AGARD-CP-373)*. Neuilly-sur-Seine, France: AGARD, 1984.

Ryu, K.; Myung, R. Evaluation of Mental Workload with a Combined Measure Based on Physiological Indices During a Dual Task of Tracking and Mental Arithmetic. *International Journal of Industrial Ergonomics* **2005**, *35*, 991–1009.

Sabey, B. E.; Staughton, G. C. Interacting Roles of Road Environment, Vehicle and Road User in Accidents. In *Proceedings of the Fifth International Conference of the Association for Accident and Traffic Nedicine, London*¸ (pp. 1–17), 1975.

Salthouse, T. A.; Ellis, C. L. Determinants of Eye-Fixation Duration. *American Journal of Psychology* **1980**, *93* (*2*), 207–234.

Salthouse, T. A.; Ellis, C. L.; Diener, E. M.; Somberg, B. L. Stimulus Processing During Eye Fixations. *Journal of Experimental Psychology: Human Perception and Performance* **1981**, *7,* 698–712.

Schlegel, R. E.; Gilliland, K.; Schlegel, B. Development of the Criterion Task Set Performance Data Base. *Proceedings of the 30th Annual Meeting of the Human Factors Society* (pp. 58–62). Santa Monica, CA: Human Factors Society, 1986.

Schutz, A. C.; Braun, D. I.; Gegenfurtner, K. R. Contrast Sensitivity During the Initiation of Smooth Pursuit Eye Movements. *Vision Research* **2007**, *47,* 2767–2777.

Stein, E. S. *Air Traffic Controller Workload: An Examination of Workload Probe* (DOT/FAA/CT-TN84/24). Atlantic City International Airport, NJ: U. S. Department of Transportation, Federal Aviation Administration Technical Center, 1985.

Svensson, E.; Angelborg-Thanderz, M.; Sjoberg, L.; Olsson, S. Information Complexity: Mental Workload and Performance in Combat Aircraft. *Ergonomics* **1997**, *40* (3), 362–380.

Tole, J. R.; Stephens, A. T.; Vivaudou, M.; Ephrath, A. R.; Young, L. R. *Visual Sscanning Behavior and Pilot Workload*; (NASA Contractor Rep. No. 3717); Hampton, VA: Langley Research Center, 1983.

Tryon, W. W. Pupillometry: A Survey of Sources of Variation. *Psychophysiology* **1975**, *12* (1), 90–93.

Tsang, P. S.; Wilson, G. Mental Workload. In G. Salvendy (Ed.), *Handbook of Human Factors and Ergonomics* (2nd ed.), (pp. 243–268). New York: Wiley, 1997.

Van der Horst, R. Occlusion as a Measure for Visual Workload: An Overview of TNO Occlusion Research in Car Driving. *Applied Ergonomics* **2004**, *35*, 189–196.

Van Orden, K. F.; Jung, T. P.; Makeig, S. Combined Eye Activity Measures Accurately Estimate Changes in Sustained VIsual Task Performance. *Biological Psychology* **2000**, *52*, 221–240.

Van Orden, K. F.; Limbert, W.; Makeig, S.; Jung, T. Eye Activity Correlates of Workload During a Visuospatial Memory Task. *Human Factors* **2001**, *43* (*1*), 111–121.

Veltman, J. A.; Gaillard A.W.K. Physiological Workload Reactions to Increasing Levels of Task Difficulty. *Ergonomics* **1998**, *41* (5), 656–669.

Verwey, W. B. How Can We Prevent Overload of the Driver? In A. M. Parkes & S. Franzen (Eds.), *Driving Future Vehicles*, (pp. 235–244). London: Taylor & Francis, 1993.

Verwey, W. B.; Veltman, H. A. Detecting Short Peaks of Elevated Workload: A Comparison of Nine Workload Assessment Techniques. *Journal of Experimental Psychology: Applied* **1996**, *2*, 270–285.

Wickens, C. D. The Structure of Attentional Resources. In R. Nickerson (Ed.), *Attention and Performance* VIII, (pp. 239–257). Hillsdale, NJ: Erlbaum, 1980.

Wickens, C. D.; Helleberg, J.; Xu, X. Pilot Maneuver Choice and Workload in Free Flight. *Human Factors* **2002**, *44*, 171–188.

Wickens, C. D.; Hollands, J. G. *Engineering Psychology and Human Performance* (3rd ed.); New Jersey: Prentice Hall, 2000.

Wickens, C. D.; Lee J. D.; Gordon-Becher; S. E.; Liu, Y. *An Introduction to Human Factors Engineering* (2nd ed.); New Jersey: Prentice Hall, 2003.

Wickens, C. D.; Xu, X.; Helleberg, J.; Marsh, R. Pilot Visual Workload and Task Management in Freeflight: A Model of Visual Scanning. *Proceedings of the 11th International Symposium on Aviation Psychology.* Department of Aerospace Engineering, Applied Mechanics, and Aviation. Ohio State University. Columbus, OH, 2001.

Wierwille, W. W. Physiological Measures of Aircrew Mental Workload. *Human Factors* **1979**, *21*, 575–593.

Wierwill, W. W.; Eggemeier, F. T. Recommendations for Mental Workload Measurement in a Gest and Evaluation Environment. *Human Factors* **1993**, *35*, 263–281.

Wikman, A. S.; Nieminen, T.; Summala, H. Driving Experience and Time-sharing During In-car Tasks on Roads of Different Width. *Ergonomics* **1998**, *41*, 358–372.

Willems, B.; Allen, R. C.; Stein, E. S. *Air Traffic Control Specialist Visual Scanning II: Task Load, Visual Noise, and Intrusions Into Controlled Airspace* (DOT/FAA/CT-TN99/23). Federal Aviation Administration, William J. Hughes Technical Center. Atlantic City International Airport, NJ, 1999.

Wilson, G. F. An Analysis of Mental Workload in Pilots During Flight Using Multiple Psychophysiological Measures. *International Journal of Aviation Psychology* **2002**, *12* (*1*), 3–18.

Yerkes, R. M.; Dodson, J. D. The Relation of Strength of Stimulus to Rapidity of Babit-formation. *Journal of Comparative Neurology and Psychology* **1908**, *18*, 459–482

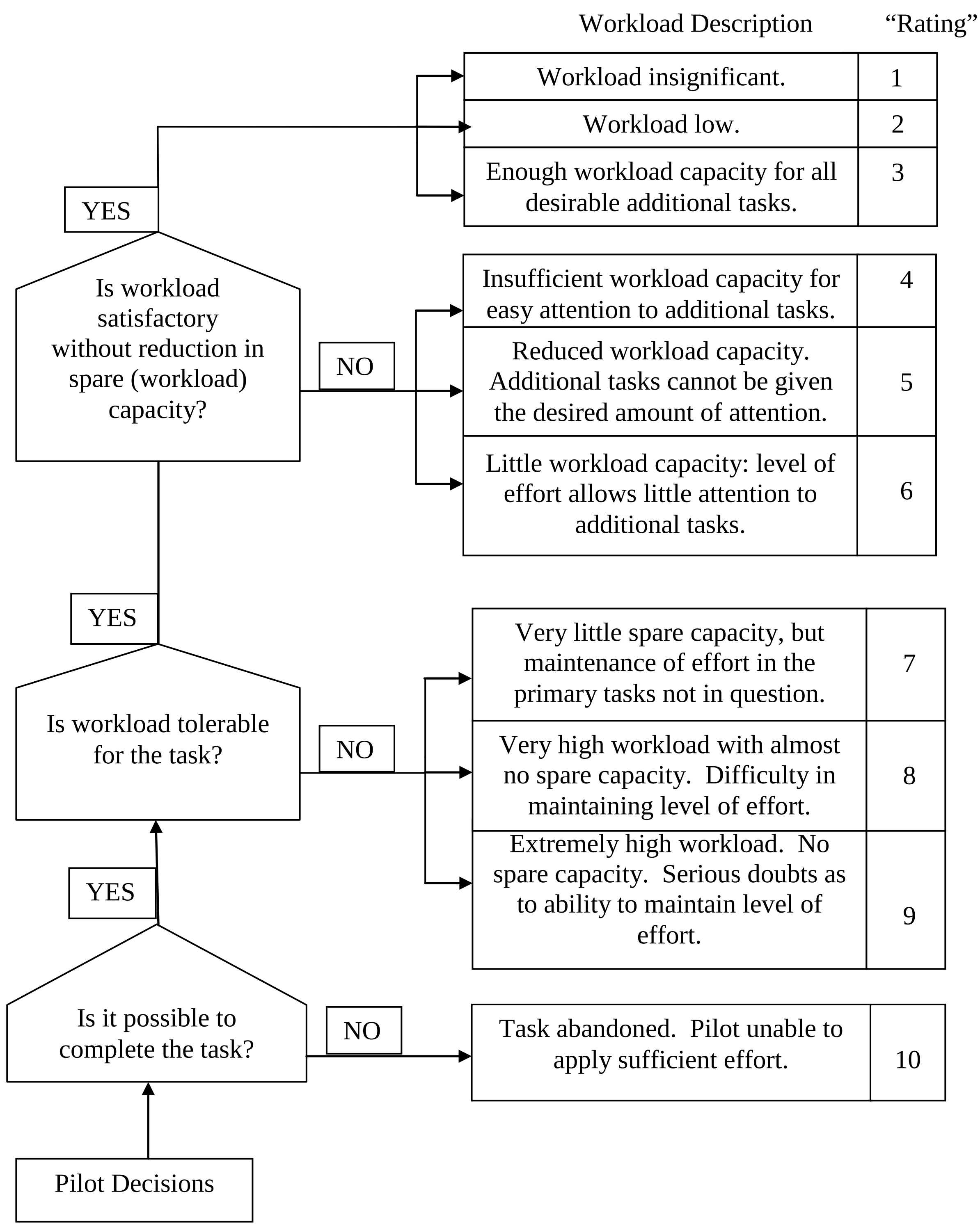
Young, M. S Stanton, N. A. Mental Workload: Theory, Measurement, and Application. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors: volume 1*, (pp. 507–509). London: Taylor & Francis, 2001.

Young, M. S.; Stanton, N. A. Mental Workload. In Stanton, N.; Hedge, A.; K. Brookhuis, E. Salas; Hendrick, H. (Eds). *Handbook of Human Factors and Ergonomics Methods*, (pp. 39.1–39.9). USA: Routledge, 2005.

Zuber, B. L.; Crider, A.; Stark, L. Saccadic Suppression Associated with Microsaccades. *Quarterly Progress Reports, 74,* 224–249, 1964.

INTENTIONALLY LEFT BLANK.

Appendix A. Bedford Workload Rating Scale



INTENTIONALLY LEFT BLANK.

Appendix B. Signed Independent Review Board Form from the College of Liberal Arts

UAH
The University of Alabama in Huntsville

Department of Philosophy
College of Liberal Arts

Huntsville, Alabama 35899
Phone: (256) 824 6555

William S. Wilkerson
wilkerw@email.uah.edu

Sage Jessee
c/o Dr. Anthony Morris
Department of Psychology
Huntsville, AL 35899
January 28, 2008

Dear Mr. Jessee,

As chair of the IRB Human Subjects Committee, I have reviewed your proposal, *Pilot Visual Scanning of Areas of Interest during UH-60M Flight Simulations* and have found it meets the necessary criteria for **expedited review** according to 45 CFR 46. I have approved this proposal, and you may commence your research. Please note that this approval is good for one calendar year from the date above. If data collection lasts beyond this period, a renewal application must be filed.

Contact me if you have any questions.

Sincerely,

Dr. William Wilkerson,
Chair, UHSC

A Space Grant College
An Affirmative Action/Equal Opportunity Institution

INTENTIONALLY LEFT BLANK.

Appendix C. Waiver of Informed Consent

UAH

The University of Alabama in Huntsville

Department of Philosophy
College of Liberal Arts

Huntsville, Alabama 35899
Phone: (256) 890-6555

Sage Jessee
c/o Anthony Morris
Department of Psychology
UAHuntsville
Huntsville, AL 35899

December 2, 2008

Dear Mr. Jessee,

As chair of the IRB Human Subjects Committee, I have reviewed your request to waive informed consent for the project *Pilot Visual Scanning of Areas of Interest During UH-60M Flight Simulations* and have found it meets the necessary criteria for approval according to 45 CFR 46. You are approved to waive documentation of informed consent. Prior IRB approval for your research remains in effect and expires 28 January 2009. If data collection continues past this period, a renewal application must be filed with the IRB.

Please contact me if you have any questions.

Sincerely,

Dr. Nicholaos Jones
Chair, UHSC

INTENTIONALLY LEFT BLANK.

Appendix D. Detailed Description of Flight Scenario 1

Event	**Comments**	**Action**	**PVI Task**	**Flight Segment**
Aircraft Pre Flight and cockpit set up conducted.	Follows check list to AUX PWR ON. Conducts system checks and set-up.		Loads AMPS. Set radios runs ground checks, set FDDCP.	Preflight Segment
Aircraft run-up HIT Check conducted.			Performs HIT check.	
Commo check - Baseline lead requests radio check with flight.	Pilot responds with, "Chalk 2 commo check 1, 2, 3"	Chalks 3, 4, 5 would follow with commo checks.	Sets radios makes commo check with lead.	
Aircraft Hover taxi to runway 34.	Pilot contacts ground for taxi clearance.	Lead positions on runway. M lines up on Lead at 45 dg angle and 3 rotor disc separation.	Ground taxi for T/O	
Kiowa Warrior departs for Air route recon.	Notional radio call from KW that he is departing for recon.		Observes	Baseline
Longbow depart to provide overwatch.	Notional radio call from Apache that he is departing.		Observes	
Flight departs Bike for PZ Erwin.	After liftoff, notional radio call by trail A/C that flight is off and formed.		Follows lead	A/C lift off Segment
Flight arrives PZ, uploads troops.			Follows lead	A/C lands to pick up troops
Flight departs PZ, proceeds to SP2.		After liftoff, notional radio call by trail A/C that flight is off and formed.	Follows lead	A/C lift off
Flight arrives SP 1.			Follows lead	A/C cruises to WPs
Flight arrives ACP1.			Follows lead	
Flight arrives ACP2.		Need threat locale.		

SIL receives observation report of zsu-23 spotted vicinity of Baker Airfield, N 35 16.93 W116 04.62.	Threat is a armored vehicle with detection range of 12.5 mi and engagement range of 2 mi. Threat range overlaps the ACP1 to ACP2 route.	Crew notes report alters course to avoid detections.	Crew uses plot on map function to determine threat position. Crew uses TIV mode.	
Flight arrives ACP3.				
KW clears the LZ, sends SITREP.				
Flight arrives RP1.				
Flight lands LZ Saratoga off loads troops, picks up supplies .				A/C lands
Flight departs LZ.	SIL sends pre-formatted Free Text message that LZ is clear.			A/C lift off
Flight arrives SP2.				A/C cruises to WPs
Flight arrives ACP4.				
Flight arrives ACP5.				
Flight arrives ACP6.				
Flight arrives RP2.				
Flight land LZ Goldstone.				A/C lands
Flight departs LZ Goldstone.				A/C lift off
Aircraft has Stab failure.			Crew reacts to emergency. Flies to Bike Lake and lands.	A/C cruises to WPs with Stab failure
Flight arrives SP3.				
Flight land Bike Lake.				A/C lands

List of Symbols, Abbreviations, and Acronyms

ANS	autonomic nervous system
ASL	Applied Science Laboratory
ATCs	Air Traffic Controllers
ATC	air traffic controller
ATEC	Army Test and Evaluation Command
ATM	Aircrew Training Manual
ATWIT	Air Traffic Workload Input Technique
BWL	Bedford Workload Scale
CDTI	cockpit display of traffic information
CDU	Control Display Unit
CNS	central nervous system
CMT	continuous memory task
CWI	cognitive workload index
ECG	electrocardiogram
EEG	electroencephalic activity
EEG	Electroencephalogram
EOG	electrooculogram
ERPs	event related brain potentials
GPS	Global Positioning System
ICA	Index of Cognitive Activity
IFC	instrument flight conditions
IFR	instrument flight rules
IMC	instrument meteorological conditions
ISL	instrument landing system

LUT	Limited Usability Test
MFDs	Multi-functional Displays
MWL	mental workload
OTW	outside the window
POG	point of gaze
SED	Software Engineering Directorate
SIL	Systems Integration Laboratory
SSDD	System Simulation and Development Directorate
SWAT	Subjective Workload Assessment Technique
TA	time available
TLX	Task Load Index
TR	time required
TRACON	Terminal Radar Approach Control
VFC	visual flight conditions
VFR	visual flight rules
VMC	visual meteorological conditions
VWL	visual workload

NO. OF COPIES	ORGANIZATION
1 ELECT	ADMNSTR DEFNS TECHL INFO CTR ATTN DTIC OCP 8725 JOHN J KINGMAN RD STE 0944 FT BELVOIR VA 22060-6218
1	ARL FIRES CENTER OF EXCELLENCE FIELD ELEMENT ATTN RDRL HRM AF C HERNANDEZ 3040 NW AUSTIN RD STE 221 FT SILL OK 73503-9043
1	ARL HRED AMEDD FLD ELMT ATTN RDRL HRS EA V RICE-BERG BLDG 4011 RM 217 1750 GREELEY RD FT SAM HOUSTON TX 78234-5094
1 CD	ARMY G1 ATTN DAPE MR B KNAPP 300 ARMY PENTAGON RM 2C489 WASHINGTON DC 20310-0300
1	ARMY RSRCH LAB-HRED ATTN RDRL HRM CU 6501 E 11 MILE RD MS 284 BLDG 200A 2ND FL RM 2104 WARREN MI 48397-5000
1	ARMY RSRCH LAB/HRED ATTN RDRL HR T LETOWSKI BLDG 520 RM 39 ABERDEEN PROVING GROUND MD 21005
1	ARMY RSRCH LABORATORY-HRED ATTN RDRL HRM A J MARTIN MYER CENTER RM 2D311 FT MONMOUTH NJ 07703-5601
1	ARMY RSRCH LABORATORY-HRED ATTN RDRL HRM DW E REDDEN BLDG 4 RM 332 FT BENNING GA 31905-5400
1	US ARMY RSCH LABORATORY – HRED AVNC FIELD ELEMENT ATTN RDRL HRM DJ D DURBIN BLDG 4506 (DCD) RM 107 FT RUCKER AL 36362-5000
1	US ARMY RSCH LAB-HRED JFCOM JOINT EXPERIMENTSATION J9 JOINT FUTURES LAB ATTN RDRL HRM AJ J HANSBERGER 115 LAKEVIEW PARKWAY STE B SUFFOLK VA 23435
1	US ARMY RSCH LABORATORY ATTN RDRL CIM P S FOPPIANO BLDG 459 ABERDEEN PROVING GROUND MD 21005
1	US ARMY RSRCH LAB ATTN RDRL HRM C A D DAVISON 320 MANSCEN LOOP STE 115 FT LEONARD WOOD MO 65473
5	US ARMY RSRCH LAB-HRED FIELD ELEMENT ATTN RDRL HRM DI M S JESSEE (5 COPIES) REDSTONE ARSENAL AL 35898
1	US ARMY RSRCH LABORATORY ATTN RDRL HRD G R SPINE BLDG 333 PICATINNY ARSENAL NJ 07806-5000
1	US ARMY RSRCH LABORATORY ATTN RDRL HRM DI T DAVIS BLDG 5400 RM C242 REDSTONE ARSENAL AL 35898-7290
1	US ARMY RSRCH LABORATORY- HRED ATTN RDRL HRM DQ M R FLETCHER AMSRD-NSC-WS-E BLDG 3 RM 341 NATICK MA 01760-5020
1	US ARMY RSRCH LABORATORY- HRED ATTN RDRL HRM CK J REINHART 10125 KINGMAN RD FT BELVOIR VA 22060-5828

NO. OF COPIES	ORGANIZATION
1	US ARMY RSRCH LABORATORY- HRED ATTN RDRL HRM CN R SPENCER DCSFDI HF HQ USASOC BLDG E2929 FT BRAGG NC 28310-5000
1	US ARMY RSRCH LABORATORY- HRED ATTN RDRL HRM AS C MANASCO SIGNAL TOWER BLDG 29808A RM 303 FT GORDON GA 30905-5233
2	US ARMY RSRCH LABORATORY- HRED ATTN RDRL HRM AT J CHEN ATTN RDRL HRM AT C KORTENHAUS 12350 RESEARCH PARKWAY ORLANDO FL 32826-3276
1	US ARMY RSRCH LABORATORY- HRED ATTN RDRL HRM AP D UNGVARSKY BATTLE CMD BATTLE LAB POPE HALL BLDG 4709 BCBL 806 HARRISON DR FT LEAVENWORTH KS 66027-2302
1	US ARMY RSRCH LABORATORY- HRED ATTN RDRL HRM AV S MIDDLEBROOKS 91012 STATION AVE RM 348 FT HOOD TX 76544-5073
1	US ARMY RSRCH LABORATORY- HRED ATTN RDRL HRM YA M BARNES 2520 HEALY AVE STE 1172 BLDG 51005 FT HUACHUCA AZ 85613-7069

NO. OF COPIES	ORGANIZATION
1	US ARMY RSRCH LABORATORY- HRED ARMC FIELD ELEMENT ATTN RDRL HRM CH C BURNS BLDG 1467B RM 336 THIRD AVE FT KNOX KY 40121
1	US ARMY RSRCH LAB ATTN RDRL HR L ALLENDER ABERDEEN PROVING GROUND MD 21005
1	US ARMY RSRCH LAB ATTN RDRL CIM G T LANDFRIED BLDG 4600 ABERDEEN PROVING GROUND MD 21005-5066
1	US ARMY RSRCH LABORATORY- HRED ATTN RDRL HRM D D HARRAH ABERDEEN PROVING GROUND MD 21005
1	US ARMY RSRCH LAB ATTN RDRL HRM B J LOCKETT BLDG 459 ABERDEEN PROVING GROUND MD 21005
3	US ARMY RSRCH LAB ATTN IMNE ALC HRR MAIL & RECORDS MGMT ATTN RDRL CIM L TECHL LIB ATTN RDRL CIM P TECHL PUB ADELPHI MD 20783-1197