WARNING COMPLIANCE: EFFECTS OF STRESS AND WORKING MEMORY

by

JESSICA HELMICK-RICH
M.S. University of Central Florida, 2004

A dissertation submitted in partial fulfillment of the requirements
For the degree of Doctor of Philosophy
In the Department of Psychology
In the College of Arts and Sciences
At the University of Central Florida
Orlando, Florida

Fall Term
2005

Major Professor: Peter A. Hancock

ABSTRACT

This study investigated the effects of cross-modality warning presentation and retention in a dual-task paradigm in a simulated military environment under various task-induced stress levels. It was also intended to determine what role working memory played in the mode of warning presentation that resulted in the highest retention and subsequent compliance. An all within participant design was created in order to determine if scores on working memory span tasks predicted performance across the varying forms of warning presentation. Furthermore, task-induced stress levels were varied over the course of the experiment to identify if workload transitions affected performance. Results revealed that when the presentation format and the response format matched (e.g., verbal-verbal), behavioral compliance was greater then when presentation and response format were mismatched (e.g., verbal-pictorial). Thus, it is not necessarily the presentation type that affects compliance, but the combination of presentation and response mode. Analysis also revealed that the pictorial-pictorial warning combination resulted in greater behavioral compliance compared to verbal-verbal or written-written combinations. The format of warning presentation did not affect performance on the operational tasks as predicted. Thus, the visual/spatial operational task, regardless of its complexity was not interrupted in time-sharing with intra-modal warning presentations or cross-modal time-sharing. As predicted, task based stress affected the WCCOM task in all experimental procedures. Results further revealed that as task demand increased, performance on the WCCOM task decreased. Task demand did affect the operational tasks, the shooting and the navigation tasks. The shooting task, which was less complex than the navigation task was not affected by lower levels of task demand, but at the greatest level of demand (eight warnings) performance in the operational task, degraded. Degradations in performance on the more complex task, the navigation task, materialized at a

moderate level of task demand (four warnings). For subjective ratings, task demand did affect workload ratings. As the task demand increased, the subjective workload ratings also increased, revealing a true association between workload and subjective ratings. The working memory separability hypothesis was supported by the working memory span tasks, but consequently they were not predictive of the warning presentation format.

I dedicate this work to my support pyramid: To My Mother, Father, and Husband.

ACKNOWLEDGEMENTS

First, I would like to thank my committee. My chairman and advisor, Dr. Peter Hancock, I would like to thank you for giving me the freedom to conduct research in the area of Human Factors that I find the most captivating. I would also like to thank you for making me commit 100% to my project and to myself. The two intra-department committee members, Dr. Mouloua and Dr. Rinalducci, have been two of my most influential professors for a number of years. Thank you for being by my side from day one. A special thank you to Dr. Rinalducci and his wife who have welcomed me into their home time and time again as if I were part of their family. My outside committee member, Dr. Oron-Gilad, no words could describe how thankful I am that you stood by me and guided me through this dissertation process. You have been my champion from the minute you took me on as your student.

I would also like to thank Kelly Burke, my colleague for helping many late nights and early mornings on the programming aspects of this project. You are a true friend. To my assistant and friend Jennifer Moore, I am so glad I selected you the first day of this project. Young grasshopper, you have proven to have the skills and abilities to be a great scientist yourself. Thanks to the rest of the support staff, for your assistance on this project.

On a more personal level, I would like to thank my family and friends whom without, I would not have accomplished so much. My mom and dad, thank you for the years of encouragement and support. You have both been an inspiration to me in so many ways. I love you and hope I have made you proud. To my sister Adrian…. can you believe it! Thank you for giving me many conversations filled with laughter, support, and the advice to never give up. To my beloved husband, Noah, thank you for showering me with love. Together we can accomplish all of our dreams.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

ANOVA	Analysis of Variance
ANSI	American National Safety Institute
C-HIP	Communications-Human Information Processing model
GLM	General Linear Model
ISO	Organization for International Standardization
LSD	Fisher's Least Significant Difference
NASA-TLX	NASA-Task Load Index
RSME	Rating Scale Mental Effort
WCCOM	Warning-Color Compliance Task
PCP	Proximity Compatibility Principle
VP	Verbal presentation, Pictorial response
PP	Pictorial presentation, Pictorial Response
WP	Written presentation, Pictorial response
VV	Verbal presentation, Verbal response
PV	Pictorial presentation, Verbal response
WV	Written presentation, Verbal response
VW	Verbal presentation, Written response
PW	Pictorial presentation, Written response
WW	Written presentation, Written response

CHAPTER ONE: INTRODUCTION

Warnings are a central part of all work environments. Warnings serve three functions: firstly, they improve safety; secondly, they influence people's behavior; and thirdly, warnings provide information to the user about the hazard, compliance behavior to follow consequences of non-compliance and therefore enable the user to make informed decisions (Laughery & Hammond, 1999). Not all accidents can be avoided (Reason, 1990), but often compliance with the warning message eliminates the occurrence of bodily damage or death. Warnings have proven to be valuable, thus resulting in an increase in safety behavior and the donning of protective equipment. Empirical findings suggest that when a warning was presented, safety behavior, increased up to forty percent (Wogalter, Godfrey, Fontelle, Desaulniers, Rothstein, & Laugherty, 1987; Otubso, 1988; Wogalter, Fontenelle, & Laughery, 1985). Although warnings are an effective way to communicate hazards, they may also be a source of stress. Frequently, when warning messages are presented to individuals it is in a work environment while performing other job-related tasks. Thus, while individuals are performing their primary job responsibilities they may need to heed single or multiple warning messages. While performing dual tasks (the operational task and the warning compliance task) the effects of stress from multi-tasking may be crucial to performance outcomes (Driskell & Olmstead, 1989; Kanki, 1996). Not only is multitasking an issue that may influence the individual's stress level, the sheer number of warnings presented may also have an effect on performance. The amount of warnings presented could also increase the level of task demand imposed on the operator. Yet, a gap in the warnings domain exists in identifying the effects of stress on compliance behavior when multiple warnings are presented in a dual task environment.

In an attempt to narrow down the focus to one of the various stressors that affect performance, the present research investigated the effects of task demand in a military setting. Military tasks are environmentally demanding and dynamic, thus they are the most relevant setting to observe the effects that task demand has on behavior. Furthermore, warnings are commonplace in military tasks; they communicate hazardous information that is intended to persuade behavioral compliance. Military personnel are required to perform their job tasks while simultaneously presented with such warning messages. Since some military tasks are inherently risky and cannot be avoided, developing the most effective warning that would improve safety, influence behavior, and provide information to make informed decisions is imperative to the life of military personnel. Therefore, a military simulated task is an appropriate experimental setting to test presentation format and task demand in a real-world dual-task paradigm.

Although stress has not been an area that has received a lot of attention in the warnings domain, a timeless question that has perplexed researchers is how to develop the most effective warning that will improve response so that individuals will comply with the warning every time they encounter one. The quest to answer this question has identified many characteristics of a warning that has enhanced their effectiveness. The research in the safety arena on warnings has concentrated on several different aspects that not only include the warning design, but the individual(s) that perceived the warnings. Historically, the focal point of warning research has been on intrinsic factors pertaining to the warning message. Intrinsic factors are variables that are inherent to the warning itself such as signal word (Leonard, Matthews, & Karnes, 1986; Silver & Wogalter, 1989; Wogalter & Silver, 1990), font (Braun, Silver, & Stock, 1992), and placement (Wogalter, Godfrey, Fontenelle, Desaulniers, Rothstein, & Laughery, 1987). The manipulation of these intrinsic warning factors has been beneficial in increasing compliance, but has not proven to be entirely effective. These results may be due to the isolating effects of only investigating the intrinsic warning factors and not the symbiotic relationship of intrinsic and extrinsic non-warning factors.

Until recently, researchers have limited their focus to intrinsic warning factors and little has been done to understand the effects of extrinsic non-warning factors on performance. Extrinsic non-warning factors have been described as variables that are not embedded in the warning itself. Rogers, Lamson, and Rousseau (2000, p. 103) defined extrinsic non-warning factors as "those factors that are specific to individual(s) who interact with the warning." Previously, variables such as age, gender, cultural background, cognitive factors, and personality variables have been identified as factors that affected behavioral compliance (Rogers, Rousseau, & Lamson, 1999). Subsequently, stress as an extrinsic non-warning factor has also been recognized as one that influences behavior, yet a modest amount of research has been conducted to support this claim (Magurno & Wogalter, 1994). Stress is an extrinsic non-warning factor of interest in this study because it is inherent to most real-world tasks, specifically environments that involved multiple tasks. Understanding how robust warnings were to environmental stressors, such as task demand, is an important area that needs further development. An overarching objective of the present research is to determine if stress affected behavioral compliance with warnings in a realistic setting where multi-tasking is necessary.

An individual's ability to comply with multiple warning presentations may be affected by working memory capacity. Previous literature in the warnings domain suggests that not only the number, but also the mode of information presentation affects compliance rate (Lehto & Miller, 1986; Rousseau, Lamson, & Rogers, 1998; Wogalter, & Usher, 1999). Yet to date, working memory, as an extrinsic non-warnings factor, has generated little empirical interest in the warnings literature. Therefore, working memory may play a role in the modality of warnings that result in the highest behavioral compliance. Thus, understanding the mode of presentation that results in the highest rate of compliance for individuals who differ in working memory ability would serve as a way to develop better warnings. Yet, these differences in working memory have yet been accounted for in the literature, and as a result, warnings have not been designed with the individual's ability to store, process, and later recall the information in mind. One intent of the present work to determine if working memory can account for the warning format that yields the greatest behavioral compliance.

In light of the aforementioned gaps that still exist in the warnings domain; the focus of the current research sought to determine if: a) task-induced stress affected compliance behavior; b) if the format of presentation/or response affected compliance; and c) if individual differences in working memory capacity played a role in warning compliance.

CHAPTER TWO: LITERATURE REVIEW

Warnings

The Purpose they Serve and how they Differ from other Forms of Communication

Warnings are used as tools to communicate hazards (Braun & Silver, 1999). Hazards are defined as those conditions that may be detrimental to the individual receiving the message. Military tasks are good examples of how warning implementation provide the necessary means for survival. Hazardous equipment, hazardous tasks, and hazardous environments are inherent to military tasks that are often "dull, dirty, and dangerous" missions. Thus, the implementation of warning messages has proven to be crucial to military safety and effectiveness and has resulted in the protection of military personnel. Furthermore, warnings presented in high-risk situations are intended to influence or change military personnel's behavior by informing them of possible risks associated with a particular technical system or environment, thereby enabling more informed decisions and better judgments. Conveying warnings prior to their interaction with a tool or a system shifts the responsibility to the individual operator. The military workforce has a right to be well-informed about risks that are involved with using a product or system before they encounter them. Thus, warnings are employed in military settings to protect military personnel from harm that may be avoided if warnings compliance is successful.

It is imperative to begin with a warning definition and describe the type of warnings that will be used in the current study. The focus of this study is only on safety warnings as they are the most relevant type of warning necessary for military operators. Safety warnings are those that inform the individual about circumstances that pose bodily harm or death. Other types of warnings are intended to persuade individuals to comply due to negative consequences that may occur. These consequences do not necessarily include bodily injury or death. Since the present experimental task was a military task, warnings needed to coincide with the severity of the situation. For the present purpose, only safety warnings were considered and, unless otherwise specified, safety warnings are subsequently referred to as "warnings".

Warnings are often used interchangeably with other terms that are intended to communicate information to the user. Distinctions between safety warnings and other warnings, instructions, and rules deserve further clarification. Firstly, safety warnings differ from other types of warnings in that safety warnings are messages communicating the possibility of injury or death. Other warnings, which do not fall under the umbrella of safety warnings, are messages about other negative consequences such as property damage, social disapproval, loss of time or money, or penalties imposed by the police, employers, parents, or other authorities (Ayers, Gross, Wood, Horst, Beyer, & Robinson, 1989).

Another construct that is often confused with warnings are instructions. Instructions are directions that also call for compliance. Initially instructions do not seem to differ significantly from warnings. Warnings, like instructions, call for compliance, but in addition, warnings notify the operator of the hazardous situation before the danger is imminent. Unlike instructions, warnings alert the user of consequences that may occur. The message content alone, whether it is a warning or instructions, does not act as a warning, but is person and setting specific. If an individual does not feel threatened, he or she may not comply with a warning. Yet, in a military environment, military personnel are trained on the importance of warning information and the varying levels of threat conveyed by a warning message. While non-military personnel have a larger range of options as to whether they will comply, military personnel are more restricted as to what actions they may take. Therefore, it can be assumed that military operators will comply with all warning messages that are presented, unless the message is not received or cannot be recalled.

Thirdly, warnings also differ from rules, yet the distinction between the two can be indistinct. Rules are a prescribed guide for conduct or action. Rules can include warnings, but like instructions, rules do not include consequences that may occur. Rules warn the user and imply that if the rules are disobeyed then an authority figure (such as police or owner) may be displeased and take action. Contrary to rules, warnings are not imposed from an authoritative standpoint, but are implemented to warn about a potential hazard.

Fourthly, warnings and alarms are often times used interchangeably. Alarms have many functions (see McDonald, 2001). Stanton defines alarms as an “unexpected change in system state, a means of signaling state changes, a means of attracting attention, a means of arousing the operator and a change in the operators mental state (Stanton, 1994, p. 6)." Alarms alert individuals when a problem occurs and attracts attention at the time of onset of change. Warnings differ from alarms in that warnings inform individuals of the potentially hazardous nature of the product or environment prior to their interaction rather then alert them of immediate danger. Moreover, warnings also inform individuals of the compliance behaviors that they should follow in order to help protect them from possible bodily injury. In addition, in this line of research the warnings that were presented in the experimental task were not an inherent part of the system state. Unlike alarms, which notify the user when the system state changed, warnings used in this experiment were presented a priori and it was the individual who was responsible for recognizing the warning factors once they were present.

In summary, warnings are intended to inform individuals of hazards before they are in imminent danger. Furthermore, they are a form of communication that, not only notifies individuals of the hazard, but also relay the consequences of non-compliance. These factors set warnings apart from other constructs of communication and should not be used interchangeably with other forms of communication. In this line of research, warnings were presented to individuals before their interaction with the military simulated task communicating the hazards that may have been present in the environment. In addition, when the individuals were simultaneously performing the military task, warnings were presented and the individual was instructed to comply with the warning message. Thus, the experimental setting mimicked real world military tasks, which are inherently dangerous and utilize warnings to communicate hazards while performing their required job responsibilities.

Warnings from a Theoretical Perspective

Warnings can be viewed from a number of different theoretical frameworks. The frameworks that are often cited in the warnings literature are the communication theory (Lasswell, 1948) and the information-processing model (Shannon & Weaver, 1949). Recently Wogalter, Dejoy, and Laugherty (1999) combined the basic components of each of the two models into one hybrid model, the Communications-Human Information Processing model (C-HIP). The hybrid model integrated the three components of the communication model and extended the last component, the receiver, by incorporating the information-processing model. This model includes the three conceptual stages of the communication model, in addition to the information-processing component of the receiver (See Figure 1). A warning message has to be processed at each stage of this model in order to produce the desired compliance behavior.

Bottlenecks may occur at any given stage in this model, thus leading to the possibility of non-compliance.

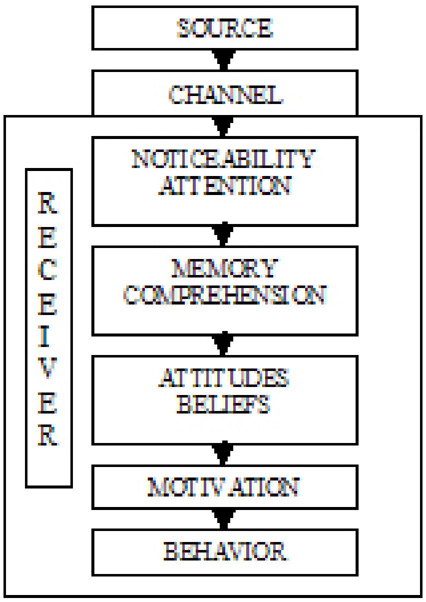


Figure 1: Communication Human Information-Processing (C-HIP) model.

The C-HIP model is critical to the understanding of warnings and compliance behavior since it identifies the stages of processing that individuals experience when presented with a warning. In the context of C-HIP model, when a receiver is presented with one or multiple warnings it is possible to identify a bottleneck. Often, in empirical testing the receiver and warning are isolated in order to determine where the bottleneck occurred. Yet, it is the interaction between the receiver and the warning that is critical to the understanding of the information processing experience.

The C-HIP model incorporates both the internal warning factors and external non-warning factors. Although the C-HIP model is the most inclusive model that exists at the time, it is lacking many crucial aspects that affect warning compliance behavior. For instance, the effects that stress may have on the many stages of information processing are not taken into account. Furthermore, individual differences in working memory are also not included in the model. In order to have a comprehensive view of the factors that effect compliance behavior, this line of research set out to determine if stress affected compliance behavior. In addition, it was of interest in the current study to determine if bottlenecks in processing were due to individual difference variables (receiver) and/or the format (channel) in which the warning information was presented.

Modality

It is commonplace in military tasks for personnel to receive warning information via different sensory modalities. These different modalities are most typically in the form of auditory and/or visual information. Although there are other types of sensory modalities, including taste and kinesthesis, the most common type of communication is still in the form of visual and auditory modalities. In order to keep the testing environment analogous to the real world military task, visual and auditory warnings were investigated. This work also sought to determine the format of communication that results in the highest rate of behavioral compliance in a military environment. The following section on modality therefore describes the empirical support for the various modes of presentation in the memory domain, followed by the literature on mode of presentation in the warnings domain.

Researchers in the warnings domain were not the first to look at the effects that modality had on performance. In the human memory arena, the literature on modality comparisons dated back more then three decades, opinions still differ on which medium best-communicated information (Clark, 1983; Mayer, 1997; Penney, 1975; Penney & Butt, 1986). Albeit there was still conflicting evidence supporting the superior modality, auditory modes of communication were generally found to be superior to written information in working memory tasks (Penney, 1975; Murdock, 1968; Watkins & Watkins, 1980). Conflict arose when the experimental process differed and long retention intervals in the experimentation process or scoring methods differed from experiments that resulted in the auditory superiority finding (Turvey, 1969).

It has been found in memory research that in free recall tasks, auditory superiority remained consistent with previous findings (Penney, 1975; Craik, 1969). If the interval between presentation and recall were silent or if non-verbal distractions were present, auditory information was recalled at a greater rate than visual information. Recall was reduced for verbal recall when an auditory distracter was presented in the time interval between presentation and response more so then when non-verbal distracters were present. Both distracters, verbal or visual reduced the rate of recall on visually presented words (Broadbent, Vines, & Broadbent, 1978; Gardiner, Thompson, & Maskarinec, 1974).

The task that fills the interval of time between presentation and recall is important to the development of the best medium of communication of warnings information. In a previous section, warnings were described as a form of communication that warned individuals of danger before the hazard was eminent. Thus, what happened in interval between the warning and the hazard onset was critical to recall. If the distracter task interfered with the warning in the interval between presentation and recall, individuals did not only have degradations in recalling the warning information, but also lacked the ability to comply. The span of time between warning onset and recall is an important factor to consider in the design of warnings since warnings are presented before the individual encounters a hazard. Thus, the task that fills the time between warning presentation and the hazard onset must be considered. In the current line of research, individuals are performing a visual/spatial military task during and between warning stimuli, thus the auditory format of warning presentation should result in superior behavioral compliance.

The Effects of Modality on Warnings Compliance

In the warnings domain, researchers are also interested in the best mode of communication that results in the highest rate of compliance behavior. Wogalter and Young (1991) investigated the voice-print warning difference in a mock chemistry task. In a series of experiments, two lab experiments and one field, the benefits of voice warnings over print warnings were demonstrated. The series of studies compared the single modality of print to verbal; they also included a combination of warning modes of print plus verbal. In single modality comparisons, verbal warnings resulted in greater behavioral compliance than printed warnings. These results were consistent with previous findings in the memory literature (Penney, 1975; Craik, 1969).

The concentration in the warnings literature moved away from the voice-print warning comparisons to investigate the benefits that pictorials may have on behavioral compliance. Voice-print warning comparisons were not abandoned, but experimentation included pictorials to the respective warnings. An increase in pictorial warnings occurred for multiple reasons including pictorials utility for conveying messages (Young & Wogalter, 1990); specifically, conveying messages to individuals who could not read a printed warning due to visual impairments, inadequate reading ability, or a language barrier (Boersema & Zwaga, 1989; Collins, 1983; Zwaga & Easterby, 1984; Wogalter & Silver, 1995).

Consequently, in the multitude of empirical data that existed on warning modality the independent variables were consistent; for the most part, written-pictorial-auditory differences were compared (or some combination of those formats). Yet, variations in the dependent variables existed. The dependent variables in the warning literature ranged from if a warning was noticed, read, recalled, complied with, understood, altered perceptions, etc. (Friedmann, 1988; Wilkinson, Cary, Barrs, & Reynolds, 1997; Wogalter & Young, 1991). Thus, it is crucial when reading the literature on warning format that the variations in dependent variables are kept in mind.

Jaynes and Boles (1990) examined the differences in noticability, readability, recall, and compliance behavior on verbal, pictorial, and pictorial plus written warnings. The highest rate of compliance was found in the verbal plus pictorial condition, yet was not significantly different from verbal alone. Results of this study also yielded verbal warnings as the warning that was noticed and read most frequently. Yet, recall for the verbal plus pictorial condition was highest compared to the other four conditions. Jaynes and Boles' (1990) went beyond just looking at verbal written differences and investigated the effects of adding a pictorial would have on performance. Subsequently, the results Jaynes and Boles (1990) found were consistent with Wogalter and Young's (1991) results, also concluding verbal warnings as the superior format of communication.

Consequently, warnings are often not presented in isolation, yet, does the voice superiority finding hold up when the environment presents multiple stimuli? Wogalter, Rashid, Clarke, and Kalsher (1991) investigated the effects that multi-modal warnings had on behavioral compliance in a cluttered environment. Uncluttered environments yielded significantly greater compliance behavior compared to cluttered environments. Furthermore, when voice warnings were present, compliance scores were significantly greater than when they were absent. In addition, when strobe or pictorials were present, no differences were found. Unfortunately, the results of this study suggests that environments where warnings may not be salient from their background, auditory warnings result in better compliance. Therefore, when designing warnings for busy areas or when the visual area is cluttered, auditory warnings may result in being the superior mode of communication.

In an empirical study, Friedmann (1988) wanted to determine the effects of adding a pictorial to a written warning. The three levels of warnings used in the study were, written warning, written warning plus a proactive pictorial, and written warning plus a reactive pictorial. Friedmann (1988) found no significant differences in compliance between warnings with a pictorial and warnings absent of a pictorial. Otsubo's (1989) study yielded similar non-significant results. The four levels of warning information used were, written, pictorial, written plus pictorial, and no warning condition. The dependent variables included noticability, readability, recall, and compliance. No significant differences were found between warning presentation formats on any of the dependent variables.

The pictorials in Friedmann (1988) and Otsubo's (1989) studies may have yielded insignificant differences in format due to inadequacies in conveying the intended message. Collins, Lerner, and Perman (1982) among other researchers, have found that the many pictorials currently being used were not well-understood (Laux, et al., 1989; Wolff & Wogalter, 1993). Pictorials that were more abstract were not as well understood as concrete pictorials (Wolff & Wogalter, 1993). Subsequently, the American National Safety Institute (ANSI, 1991) and the Organization for International Standardization's ISO 3864 (ISO, 1984) recommend that all symbols must be at least 85% or 67% correct (respectively) to be considered acceptable. In this study, all symbols presented in the study were tested and yielded within acceptable standards for the more stringent ANSI (1991) criterion.

The effects that cross modal warnings have on warning effectiveness have yielded equivocal results. Many researchers have supported the literature yielding pictorials as the superior modality in which to present warning information; they have argued that pictorials were “instant reminders” of the hazard (Peters, 1984), more recognizable then words, (Paivio, Rogers & Smythe, 1968; Standing, Conezio, & Haber, 1970), and conveyed information more rapidly and effectively then verbal messages (Dorris & Perswell, 1978). Pictorials have been found to enhance warning recall in some circumstances (Young & Wogalter, 1990). Yet, in other cases, no significant effects of pictorials have been found (Friedmann, 1988; Otsubo, 1988; Ursic, 1984). Furthermore, when pictorials were compared to written or auditory warnings, auditory, not pictorials, resulted in the superior form of warnings communication (Jaynes & Boles, 1990; Wogalter & Young, 1991). Pictorials may have been an effective form of communicating hazards, yet the varying results of pictorials warning effectiveness may have been the result of inadequate pictorials used in the experimental design.

Although most of the empirical data points to verbal presentation of information as the format which results in the best performance, some mixed results are still prevalent. It is difficult to determine the most optimal warning presentation from the research that currently exists in the warning domain. The effects of warning modality on recall and behavioral compliance have been equivocal (Friedmann, 1988; Jaynes & Boles, 1990; Otsubo, 1989; Ursic, 1984; Wogalter & Young, 1991). Some researchers suggest that the mixed results on modality presentation may have been due to inadequate pictorials used in the experimental design, yet the problem may be deeper than that. It was the intent of this line of research to determine if the format of warning information alone results in compliance or if individual differences in memory are the root to these inconsistent findings.

Since this study focused not only on presentation format, but also on memory, it is crucial to look at the research on warnings and memory. Warnings are presented before the hazard is eminent, thus the warning information may need to be processed and temporarily stored in memory. In the interest of this study, the factors that were most relevant when deciding on a format were minimized to remembering, processing, and recollection of the warning; which lead to behavioral compliance. These factors were important particularly in cases where warnings were not continually present. For instance, auditory warnings were presented and often times not repeated, unlike visual warnings that could have been presented continuously. However, there still remains a gap in the literature on variables that affected warning recall (Lehto & Miller, 1986). Moreover, the literature that does exist yielded little or no effect on warning manipulations on memory (Desaulniers, 1987; Strawbridge, 1986). Thus, it was of interest in this line of research to determine the mode of warning presentation that was best remembered, processed, and recalled. This study extended the literature in the aforementioned domains by examining what may be a predictor variable, memory, on warning format in a multi-task military environment.

Comparing the different formats of modality presentation was not a novel concept, but having investigated the cognitive underpinnings entailed in processing the varying types of modality was innovative. Therefore, two approaches were embarked on in order to investigate the cognitive components that were involved in processing warnings of varying modality. First, in order to determine if attentional resources were utilized better when divided across modalities (auditory and visual stimuli, for example) rather than displayed via two auditory or two visual channels, warnings were presented in all formats while individuals simultaneously conducted the operational task (Wickens, 1984; Wickens, Sandry, & Vidulich, 1983). This approach determined if the cognitive resources were available to perform two tasks without interference. Secondly, individuals may have been predispositioned to certain modality preferences, which may have been a factor that determined which modality was best suited at an individual level. Thus, this study investigated if working memory capacity played a role in modality preferences. These topics are discussed in detail in another section of this review.

Warnings and Stress: Fundamental Concepts in Stress and Performance

A wealth of literature exists on the effects of internal warning variables such as signal word, placement, font, color, etc. on warning effectiveness (Braun, Silver, & Stock, 1992; Leonard, Matthews, & Karnes, 1986; Silver & Wogalter, 1989; Wogalter, Godfrey, Fontenelle, Desaulniers, Rothstein, & Laughery, 1987; Wogalter & Silver, 1990). Intrinsic warning factors have proven to augment warning salience thereby increasing compliance behavior. Although warning design has increased warning compliance, it has not proven to be one hundred percent effective in influencing behavioral compliance. Thus, a need to investigate other variables, such as extrinsic non-warning factors, that may effect warning compliance is necessary.

There is not a complete void in the warnings literature on the effects of extrinsic non-warning factors on behavioral compliance. Recently, a thrust toward investigating the effects of variables such as social influence, cost of compliance, sensation seeking, and stress on warning compliance has emerged (Magurno & Wogalter, 1994; Weaver, Gerber, Hancock, & Ganey, 2003; Wogalter, Allison, & McKenna, 1989; Wogalter, Godfrey, Fontenelle, Desaulniers, Rothstein, & Laughery, 1987; Wogalter, Magurno, Rashid, & Klein, 1998). The commencement of external non-warning factors research has opened a door to the consideration of other factors that are outside the warning itself. Magurno and Wogalter (1994) identified stress as a variable that effects warning compliance and is in need of further investigation. Consequently, less than a handful of articles have been written on the effects of stress on warning compliance.

The current line of research was designed to test the effects of stress on warning compliance, specifically task demand. Yet, before delving too far into the objectives of the research an overview of the theoretical underpinnings of the stress concepts that were used to identify the effects that stress had on performance is described.

Stress

Stress Theories

There are many theories of how stress affects performance, yet there are very few unified theories that encompass the effects of various forms of stressors on performance. Early unitary theories on stress and human performance attributed emotional arousal as the source of performance decrements (Cannon, 1915; Selye, 1956). The arousal theory was supported by many studies using the narrow band approach, testing the effects of various stressors on a single task (Hockey & Hamilton, 1983). The effects that different stressors had on performance varied, yet they fit into a recognizable pattern. Incentives improved performance and stressors such as noise, thermal stress, and fatigue degraded performance. Later, combining two stressors resulted in the canceling of the decremented effects of a single stressor (Broadbent, 1963; Wilkinson, 1963). The results of these studies, and many others, patterned the inverted-U function and supported the arousal theory (Yerkes & Dodson, 1908).

Although the Yerkes-Dodson Law of arousal has been used to explain the effects of stress on performance, evidence suggested that this theory was too simplistic and flawed (Hancock &

Ganey, 2003; Hancock, Ganey, & Szalma, 2002; Hockey & Hamilton, 1983). Others have developed frameworks of stress, but none have been successful in explaining the inconsistent effects of stress on performance (Hockey & Hamilton, 1983).

Hancock, Ward, Szalma, Stafford, and Ganey (2002), recognized that creating a descriptive framework that entailed the effects of stress on performance as having been difficult for two reasons. The effects depended upon features of the environment and the individual operator, and second, the effects of various sources of stress were not uniform across all forms of information processing. In an attempt to circumvent the limitations of the inverted-U, Hancock and Warm (1989) developed a unified theory that enabled the prediction of the effects of stress both psychologically and physiologically, which they termed a *dynamic model of stress and performance.*

Hancock and Warm's (1989) model was based upon three approaches which were known as the "trinity of stress." The three approaches consist of input features (environmental stressors), adaptation features (coping mechanisms), and output features (changes in bodily functions and performance efficiency). This model provided a general architecture that explained the effects that various stressors had on individual capabilities at each level of the aforementioned approaches. In accordance with the model of stress and attention, performance was affected by stress when it increased to the point that it was outside of the comfort zone (see Figure 2). "Input level stress increases through change and intensity, prolongation of exposure time, or both in combination; output is eventually affected (Hancock & Warm, 1989, p.526)."

This model was also a bipolar representation of the effects of stress, taking into account the underload and overload of stress on the individual. If individuals were in a state of "hypostress" then he or she was not receiving enough stress to perform at an optimal level and were out side of the comfort zone. Subsequently, when stress became overwhelming, individuals reached the point of discontinuity, and were again outside of the comfort zone, in "hyperstress."

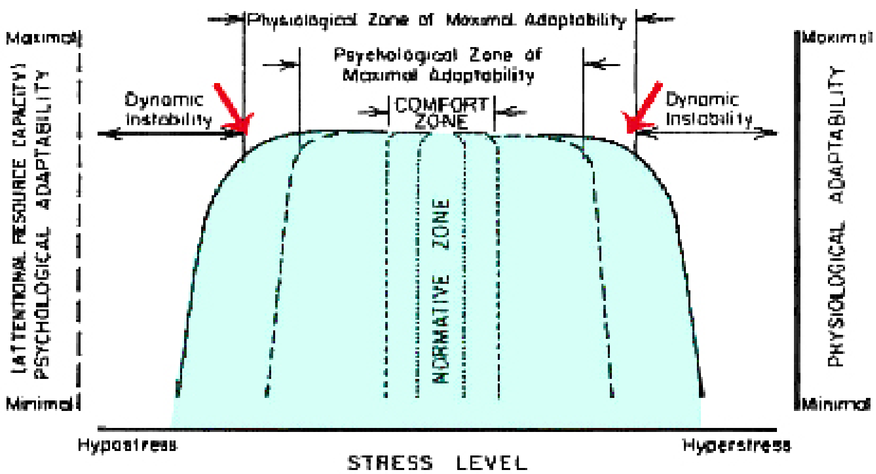


Figure 2: Hancock and Warm's (1989) model of stress and attention.

Although there were many theoretical models of stress that could have been used to predict warning compliance behavior under stress, this study applied the Hancock and Warm Model (1989) because it was an overarching theory that was useful in predicting the effects of varying levels of stress on behavioral compliance.

Warnings and Stress: Integrating Theory and Application

Stress has received little attention in the warnings domain, yet has been identified as an external non-warning factor that may well affect compliance behavior (Wogalter, 1994). Magurno and Wogalter (1994) introduced stress to the warnings domain in their empirical investigation on the effects of time pressure, social evaluation, and warning placement on compliance behavior. Results indicated that under low stress (no time pressure and evaluated from afar) participants complied more often than did participants under high stress conditions (under time pressure and evaluated at a close range). Furthermore, participants in the high stress condition also felt more stress, were less likely to see protective gear, indicated that the experimenter bothered them, and felt more worried. In addition, participants complied less often when the warning was posted on a sign compared with the warning written in the instructional page. No interaction was found between stress and warning placement.

Magurno and Wogalter (1994) pointed out that in their experiment they combined two stressors, social evaluation and time pressure, in order to produce a level of stress that was strong enough to identify any effects that it may have had on compliance behavior. Why the authors decided to combine the stressors was still unclear because previous research on time pressure indicated that time pressure alone affected performance (Goodie & Crooks, 2004; Hockey, 1978; Weinstein, 1977). Consequently, the effects that each individual stressor had on performance in this experiment could not be parsed apart. What we did know was that combining the two stressors caused enough stress to push participants outside of the comfort zone into the area of "hypostress" (Hancock & Warm, 1989). Furthermore, Magurno and Wogalter (1994) predicted that an interaction between stress and warning placement would have been found. The experimenters thought that the effects of stress would cause a narrowing of attention, task instructions would be noticed, and the posted warning would be ignored. Unfortunately, no interaction between stress and placement was found in this experiment.

In a follow up study conducted by Wogalter, Magurno, Rashid, and Klein (1998) the effects of the stressors, time pressure and social influence, on behavioral compliance were isolated. Results of the study yielded time pressure as a significant factor affecting compliance behavior and a non-significant effect of social influence. Thus, the presence of time pressure negatively effected compliance to warnings. The author suggested additional research that manipulates stress using other tasks, participant samples, and situations in order to support the current research findings.

The effects of task demand on warning compliance have also been investigated in two experiments. Task demand is described as the number of activities that an individual performs at one time. The various amount of information to be processed at once could be overwhelming and cause interference in one or more tasks. Duffy, Kalsher, and Wogalter (1995) concluded that performing two tasks simultaneously effected behavioral compliance with a written warning. In this study, participants were instructed to attach several pieces of video equipment. A warning was placed to one piece of equipment, an extension cord. In the task load condition, participants were to perform another task simultaneously while attaching the video equipment. No differences were found in compliance behavior between the control and the task load condition. Subsequently, the authors noticed that participants in the task load condition were not performing the two tasks simultaneously, yet in a serial order. Thus, it could not be determined from this experiment if stress had any effect on performance in a dual task paradigm. In order to test the effects of task load in a dual task paradigm, an experimental task where participants were required to perform tasks simultaneously is in need.

To test the effects of behavioral compliance in a genuine dual task paradigm, Wogalter and Usher (1999), examined the effects of concurrent task loadings on warning compliance behavior. In this task, participants were directed to read the task instructions and install a disk drive. This experiment consisted of three task demand conditions: control, low, and high. While installing the disk drive, participants in the stress condition were to solve single (low task demand) or double (high task demand) math problems while simultaneously installing the disk drive. Results of this experiment yielded differences in the high task demand condition and control condition on behavioral compliance. High task load participants complied significantly less than did participants in the control condition. Math performance scores were significantly better in the low task load condition than in the high task load. Finally, the time it took to complete the experimental task was longer in the high task load than in all other conditions.

Wogalter and Usher (1999) demonstrated that stress degraded behavioral compliance with warnings, increased the time it took to complete a task, and resulted in decrements in performance accuracy. Performing the installation of the computer disk drive alone did not affect compliance behavior, thus indicating that performing one of the experimental tasks did not exceed the resources that were available to the individual performing the task. Yet, when a second task, mental arithmetic, was added to the procedure, performance degraded. The two tasks together may have exceeded the resources that were available to the individuals participating in the study, thus, resulting in performance decrements. Alternatively, in accordance with Wickens' resource model (1992), the simultaneous processing of two tasks that tapped the same resources could have caused interference, also yielding decrements in performance on both tasks.

The compilation of literature that examines the effects of stress on compliance behavior with warnings results in less than a handful of data. Yet, from the small amount of data that does exist, it is beginning to appear that stress may be an external non-warning factor that may influence warning compliance. What should be noted is that in the aforementioned studies stress had a negative affect on performance, yet previous literature on the effects of stress concluded that stress did not always produce degradations in performance. It is obvious that a solid foundation cannot be built on this small amount of experimentation, and more research in this area is needed.

As mentioned previously, warnings presented in real-world tasks are not presented in isolation. In tasks, such as flying, driving, or working with hazardous equipment, individuals are concurrently performing many tasks when presented with single or multiple warnings. The demand that the task imposes on individuals can be measured by task performance and workload measures. Therefore, it is imperative to look, not only at behavioral compliance to warnings in a dual task paradigm, but the effects that multiple warnings produce. Thus, this line of research was interested in investigating the effects that increasing task demand had on warning compliance while simultaneously performing a simulated military task.

Task Demand

The current line of research focused solely on cognitive stress; specifically the effects that task demand had on performance. Some confusion in the literature existed between the construct of workload and task demand. Thus, it is important to differentiate between the two. Hiburn and Jorna (2001) suggested that task load was the demand imposed by the task itself and workload was the subjective experience of task demand. Similarly, Parasuraman and Hancock (2001) also suggested that there was a distinction between the two constructs. "Workload may be driven by the task load imposed on the human operators from external environment sources but not deterministically so, because workload is also mediated by the individual response of the human operators to the load and their skill levels, task management strategies, and other personal characteristics" (p. 306). Thus, task load is the demand placed on the individual while performing a task, while workload is experience that the individual has while attempting to adapt to the demand. In the current study, the demand that modality and the size of the memory set was investigated and the effects they separately had on subjective workload.

Task demand can vary within and between laboratory and real world settings. Many tasks have a steady, uniform demand on the individual. Other tasks fluctuate in task demand ranging from low, to medium, to high task demand. Historically, vigilance tasks performed in the laboratory were of low task demand, an abundance of this type of empirical data exists (Davies & Parasuraman, 1982; Warm, 1984). Results from vigilance studies have shown a consistent association between task demand and workload such that as task demand increases performance declines and subjective workload increases (Warm, Dember, Gluckman, & Hancock, 1991; Warm, Dember, & Hancock, 1996, Szalma, et al., 2004). Subsequently, the results of multi-task conditions dissociations emerge between workload and performance (Yeh & Wickens, 1988). Consequently, very little empirical data looked at task demand transitions, such that the demand shifted from low to moderate or high demand. Subsequently, such cases of demand transition were prevalent in real world tasks. For instance, soldiers wait in a low demand environment until they are jolted into combat (bombing, weapon firing, etc.). In this scenario, task demand began at a low level and then spiked to a very high task demand level. Therefore, what we can gleam from the data that does exist from research on vigilance and multi-tasking is that the relationship between task demand and workload is not always directly associated and more research in this area is in need.

Furthermore, O'Donell and Eggemeir (1986) suggested that performance and workload were not associated when the task demand exceeded the resources available, otherwise workload and performance were associates. Yeh and Wickens (1988) were also interested in finding the link between workload and task demand and used the attentional resource theory (Kahneman, 1973; Navon & Gopher, 1979; Wickens, 1980). They identified three circumstances in which dissociation could occur a) when more resources are invested into the task to improve resource-limited tasks (Norman & Bobrow, 1975); b) if working memory demands are increased due to multi-tasking; c) when performance is sensitive to a subtask element and subjective workload measures reflect a more global demand. This phenomenon could also be explained in the context of the Hancock and Warm (1989) model. When task demand is at low or moderate level, the individual performing the task could adapt to the task demand, and thus, performance and workload will be true associations. Furthermore, when task demand increases and goes beyond the resources available, performance would increase, or no change would occur (dissociations and insensitivities).

It was the interest of the current study to extend the literature to look at the task demand-workload relationship by investigating the effects of modality and the size of the memory set on subjective workload. Previous research on modality, (Wickens, 1992; Wickens, Sandry, & Vidulich, 1983) has indicated that in general resources were utilized better when divided across modalities (auditory and visual stimuli, for example) rather than displayed via two auditory or two visual channels. In accordance with the multiple resource theory, it was hypothesized that the visual modality in which warnings were presented (written and pictorial), would have affected performance on the operational task because it was predominantly a visual and spatial task. Thus, it was hypothesized verbal warnings would not have interfered with the secondary visual-spatial task because the codes would not have competed for resources and would therefore result in the lowest workload scores compared to pictorial and written warnings that would compete for resources.

In addition to modality, the size of the memory set (task demand) was also predicted to affect workload. Task demand was varied by implementing three levels of warning presentation which increased in the size of memory set to be stored, processed, and later recalled. Two, four, and eight warning-color combinations (which are later identified as levels 2, 4, and 8) were presented in the WCCOM task. Based on Miller's (1956) work where he presented the idea that short-term memory (working memory) could only retain and recall 5-9 bits of information, it was hypothesized that when the memory set was smaller, performance on the WCCOM and the operational task would not be affected. Yet, when the size of the memory set was larger, working memory resources would no longer be available and performance would degrade on both the WCCOM task and the operational task. Although all warning levels (2, 4, and 8) in the WCCOM task were within Miller's 7 ± 2 theory of working memory capacity, it was not the only task that was performed. The operational task was performed simultaneously with the WCCOM task, which also tapped working memory resources. Therefore, it was hypothesized that when the size of the memory set was two or four performance on both the WCCOM task and the operational task would not be affected. Yet, when the size of the memory set was level eight, performance on both the WCCOM task and the operational task would degrade.

Working Memory

Working memory is a familiar term used in psychology, specifically in the area of cognitive psychology. Although the term working memory is common, trying to figure out what the term actually means is a little more difficult. There are three main reasons why it is difficult to define what working memory is, a) short-term memory and working memory are often used interchangeably, b) there are various metaphors used to describe working memory, c) numerous structural representations of working memory exist.

The first reason that defining working memory is so difficult is due to the fuzziness in distinction between working memory and short-term memory (Brainerd & Kingma, 1985). These terms are often used interchangeably in the literature, hence, making it difficult to understand if the constructs of working memory and short-term memory are one in the same or separate memory systems. Secondly, many metaphors are used to describe working memory. For instance, the "resource metaphor", the "box" or "mental energies" are often used to describe certain aspects of the working memory system. Each metaphor highlighted a different aspect of working memory, and, depending on the theorists, different functions of the memory systems. Thirdly, unitary and non-unitary models of working memory exist. One of the most controversial topics in working memory is the notion that working memory is comprised of either a single or a unitary pool of resources.

Although there are many definitions of working memory (Cowan 1988, 1995; Engle, Kane, & Tuholski, 1999; Kieras, Meyer, Mueller, & Seymour, 1999), Baddeley's (1992) definition is one of the most widely accepted, "a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks such as language comprehension, learning, and reasoning" (p. 556). The commonality that all of these definitions have, regardless of the model that represents the construct, is that working memory is responsible for the storage and processing of information.

The Parallel between Working Memory and Attention

Many models of working memory and attention have made the distinction between the two concepts, but parsing them apart can be difficult, if not impossible. For instance, resource theories of attention were later incorporated into working memory models of attention

(Kahnemman, 1973; Navon & Gopher, 1979). Attention and working memory share many commonalities and separating the two are difficult, and perhaps impossible. It is hard to parse apart where attention begins and working memory ends because they are closely related. Many theorists consider them one in the same. Baddeley (1993) suggested that working memory and attention were so tightly knit that "working attention" better suited the integration of the two. The processing resources that are involved in both constructs have blurred the lines between them. As aforementioned, the unitary, non-unitary resource theories of attention were later integrated into working memory models. Furthermore, the introduction of the central executive (Baddeley, 1986) and the functions that it controlled were very similar to attention functions.

A consensus about the relationship between attention and working memory did not exist. Not all working memory theorists supported the notion that attention was an integrated part of the working memory system (O'Reilly, Braver, & Cohen, 1999). Subsequently, the more accepted models of working memory included attention as a key component. For instance, Baddeley had made changes to the original multiple component model of working memory (Baddeley & Hitch, 1974). Baddeley (1996) then hypothesized that the central executive controlled and regulated the two slave systems, as well as, focused and switched attention. In addition, Cowan (1988) had also taken attention into account as a function of working memory and had advocated that the allocation of attention was controlled jointly by (a) the automatic recruitment of attention to especially noticeable events and (b) voluntary, effort that demanded processes directed by the central executive. Similarly, Engle, Kane, and Tuholski (1999) supported the notion that working memory consisted of limited capacity controlled attention. They suggested that controlled attention capabilities were central to individual differences in working memory capacity. What these researchers considered controlled attention, was the same construct that Baddeley and Hitch (1974) had coined the central executive.

Although there is a need for further investigation into attention and the role it plays in working memory, the majority of the research supports the notion that attention is a component of working memory, and that these two processes are not separate. It is imperative to this research endeavor to decide which working memory contention is supported. Working memory was the focus of the current research endeavor, not attention. However, many of the stress models specifically looked at attention without mention of working memory capacity or limitations. Specifically, Hancock and Warm's (1989) model of stress and attention and Wickens' (1984) processing resources in attention model, all focus on attention; again with no mention of working memory. One explanation of why working memory was not considered in these models (and no clarification was provided) is that the stress models were developed before the link between attention and working memory was established. Thus, in this line of research the stress models were used to define stress and the effects that it had on working memory.

The many theories that encompassed the literature on working memory all have had unique perspectives on the architecture and functions of the system. Controversy still exists in the recent literature on working memory as a single or general construct of working memory and attention as component of working memory. For the purpose of this line of research, the separate pool of resources was adopted (Daneman & Tardiff, 1987; Engle, Cantor, & Carullo, 1992; Turner & Engle, 1989), as well as, attention as a component of working memory (Baddeley & Hitch, 1974; Engle, Kane, & Tuholski, 1999). Specifically, in support of the work by Shah and Miyake (1996), the notion that separate pools of resources fuel the cognitive activities of spatial and verbal working memory was assumed. Furthermore, the separate pools of resources were not in the periphery, but played a more central role in the processing and storage components of working memory.

Individual Differences in Working Memory and Warnings

In the present line of research, working memory resources were taken to represent the independent multiple resource capacity notion (Daneman & Tardiff, 1987; Engle, Cantor, & Carullo, 1992; Shah & Miyake, 1996; Turner & Engle, 1989). Moreover, the work of Shah and Miyake (1996) was of particular interest because of their focus on individual differences in the separability of working memory resources for spatial thinking and language processing. In their research, they found that independent limited-capacity pools of resources existed and were associated with different modalities. They addressed this issue by examining the individual differences in working memory capacity by determining that different types of information and their relationship to performance on different types of tasks placed high demands on working memory. Specifically, they found that individual differences in spatial working memory predicted performance on spatial working memory tasks and not language tasks (and vice versa), thus their results suggested that the separability of working memory existed (for more detail see Shah and Miyake, 1996).

In light of these findings, it was of interest in the current research to determine if individual differences in working memory played a role in determining if verbal and spatial working memory capacities predicted performance in different modalities. The literature on memory and warning compliance is lacking. Thus, it was the purpose of the current study to determine if individual differences had an effect on the compliance behavior of individuals that varied in their verbal and spatial abilities. Since there was evidence (Shah & Miyake, 1996) that independent limited-capacity pools of resources existed and were associated with different modalities, it was reasonable to apply this notion to a real world problem. Thus, the current study set out to determine if individual differences in working memory predicted the optimal format to present warning information that resulted in the highest behavioral compliance.

CHAPTER THREE: THE PURPOSE OF THE STUDY

This chapter indicates the purpose of the present study and the experimental hypotheses which are subdivided into specific predictions. It has been found that when format differences were compared in memory research, verbal information was remembered and recalled more often than written or pictorial information (Penney, 1975; Murdock, 1968; Watkins & Watkins, 1980). Consistent with the auditory superior finding in the memory literature, verbal warnings have been found to communicate hazards better than pictorials or written formats that resulted in higher behavioral compliance (Jaynes & Boles, 1990; Wogalter & Young, 1991). Consequently, the literature on format differences was isolated to environments where the individual was performing only one task. Unfortunately, these results may not have transferred to real-world environments where individuals often received warning information while performing an operational task. It was the purpose of this research to determine which format of warning presentation, verbal, written, or pictorial, was the optimal format to communicate hazards when an operator was simultaneously performing another task.

Previous research, (Wickens, 1992; Wickens, Sandry, & Vidulich, 1983) indicated that in general, attentional resources were utilized better when divided across modalities (auditory and visual stimuli, for example) rather than displayed via two auditory or two visual channels. The operational task in this study is the Ghost Recon task, which was predominantly a visual and spatial task. Ghost Recon is considered he operational task because in a real world environment, Ghost Recon would be the primary task responsibility of the operator. It was hypothesized that participants would have a significantly higher rate of compliance behavior when warnings were presented in verbal compared to written and pictorial format because the warning information would have had less interference on the operational task. Furthermore, it was predicted that compliance behavior would be significantly higher in the pictorial warning condition than in the written warning condition. In addition, it was also predicted that performance on the operational task, would have also resulted in higher performance scores in the verbal warning condition for that same reason.

Hypothesis 1. It was predicted that compliance behavior would be significantly higher in the pictorial warning condition than in the written warning condition.

Hypothesis 2: Because the operational task in this study was predominantly a visual and spatial task, it was hypothesized that participants would have a significantly higher rate of compliance behavior when warnings were presented in verbal format, compared to written and pictorial, because the warning information would have less interference on the operational task.

Historically, the majority of empirical data on warning modality has investigated the effects of one warning message either in written, auditory, or pictorial format, or in a combination of the two modes (Friedmann, 1988; Wilkinson, Cary, Barrs, & Reynolds, 1997; Wogalter &Young, 1991; Young & Wogalter, 1990). Subsequently, due to manufacturers' liability issues and the inexpensive cost of providing warnings, the increase in the sheer number of warnings has increased drastically (Twerski, Weinstein, Donaher, & Piehler, 1976). Chen (2000) found that as the number of low-criticality warnings increased, sensitivity for correctly identifying the level of threat decreased, thus suggesting that when non-critical warnings increase warnings of moderate threat are also perceived as non-critical. Thus, with an increase in task load, the demand placed on the individual while performing a task, behavioral compliance may be affected. Therefore, it was the purpose of this study to determine the optimal amount of warning information that could have been presented to an individual before the demands of the task affected performance.

It was also the interest of the current study to look at the effects of modality and the size of the memory set on subjective workload. Previous research on modality, (Wickens, 1992; Wickens, Sandry, & Vidulich, 1983) has indicated that in general resources were utilized better when divided across modalities (auditory and visual stimuli, for example) rather than displayed via two auditory or two visual channels. In accordance with the multiple resource theory, it was hypothesized that the visual modality in which warnings were presented (written and pictorial), would have affected performance on the operational task because it was predominantly a visual and spatial task. Thus, it was hypothesized verbal warnings would not have interfered with the secondary visual-spatial task because the codes would not have competed for resources. Thus, it was hypothesized that since verbal warnings should not interfere with the operational task, they would result in the lowest subjective workload rating compared to pictorial or written warnings since they share the same working memory code.

In accordance with the Hancock and Warm model of stress and attention (1989), stress would affect performance when the task demands were outside of the comfort zone. Thus, it was predicted that when the number of warnings presented was at levels two and four (2 or four warning-color combinations), performance on neither the warning compliance task nor the operational task would be affected. Yet, when eight warnings were presented behavioral compliance and performance on the operational task would degrade. Additionally, the mode of warning presentation would also affect task demand. Since the operational task in the current study is a visual/spatial task it was predicted that verbal warnings would have less interference vice pictorials or written warnings.

Hypothesis 3: It was predicted that when the number of warnings presented was two or four, performance on neither the warning compliance task nor the Operational task would be affected.

Hypothesis 4: When eight warnings were presented behavioral compliance and performance on the Operational task would degrade.

Hypothesis 5: Verbal warnings will result in a lower subjective workload ratings compared to written and pictorial because verbal warnings will have less interference on the operational task which is a visual/spatial task.

O'Donell and Eggemeir (1986) suggested that performance and workload were not associated when the task demand exceeded the resources available, otherwise workload and performance were associates. This phenomenon could also be explained in the context of the Hancock and Warm (1989) model. When task demand was at low or moderate levels, the individual performing the task could adapt to the task demand and thus performance and workload would be true associations. Yet, when task demand was at a high level, the individual performing the task could no longer adapt to the task demand.

Hypothesis 6: It was predicted that subjective workload and task demand would be correlated in conditions when the number of warning presentations was two or four.

Hypothesis 7: Subjective workload measures for conditions with two warning presentations would be significantly lower compared to conditions with four warning presentations.

Hypothesis 8: Workload measures for conditions with eight warning presentations would exceed the resources available and task load would not be associated with workload measures.

The empirical data is lacking when it comes to the variables that affect the memory of warnings (Lehto & Miller, 1986). Moreover, the literature that does exist yielded little or no effect on warning manipulations on memory (Desaulniers, 1987; Strawbridge, 1986). Yet, in domains of higher-level cognition, researchers found that working memory ability played a critical role. Furthermore, the processing and storage components of working memory tasks were found to be important factors in the prediction on spatial and verbal tasks (Baddeley, 1986; Daneman & Tardiff, 1987; Engle, Cantor, & Carullo, 1992; Shah & Miyake, 1996). Therefore, it was in the interest of this line of research to determine if individual differences in working memory ability played a role in determining the warning modality that would result in the highest recall, retention, and compliance behavior.

Hypothesis 9: Individuals low in both verbal and spatial working memory abilities would yield non-significant differences between warning presentation/format types.

Hypothesis 10: Individuals high in both verbal and spatial working memory abilities would yield non-significant differences between warning presentation/format types.

Hypothesis 11: Individuals high in verbal and low in spatial working memory abilities would perform significantly better in the auditory and written condition than in the pictorial condition.

Hypothesis 12: Individuals high in spatial and low in verbal working memory abilities would perform significantly better in the pictorial condition than in the auditory or written condition.

Table 1
Hypothesis by Topic and Number

Number	Category	Hypotheses
1	Warning Compliance/Format of Presentation	It was hypothesized that participants would have a significantly higher rate of compliance behavior when warnings were presented in verbal compared to written and pictorial format because the warning information would have less interference on the operational task.
2	Warning Compliance/Format of Presentation	It was predicted that compliance behavior would be significantly higher in the pictorial warning condition than in the written warning condition.
3	Task Demand	It was predicted that when the number of warnings presented was two or four, performance on neither the warning compliance task nor the operational task would be affected.

4	Task Demand	When eight warnings were presented behavioral compliance and performance on the operational task would degrade.
5	Workload	Verbal warnings will result in a lower subjective workload ratings compared to written and pictorial because verbal warnings will have less interference on the operational task which is a visual/spatial task.
6	Workload	Subjective workload measures for two warning presentations would be significantly lower compared to conditions with four warning presentations.
7	Workload	It was predicted that subjective workload and task demand would be correlated in conditions when the number of warning presentations was two or four.
8	Workload	Workload measures for conditions with more eight warning presentations would exceed the resources available and task load would not be associated with workload measures (dissociation or insensitivities will occur).
9	Working Memory/ Individual Differences	Individuals high or low in both verbal and spatial working memory abilities would yield non-significant differences between warning presentation/format types.
10	Working Memory/ Individual Differences	Individuals high or low in both verbal and spatial working memory abilities would yield non-significant differences between warning presentation/format types.

11	Working Memory / Individual Differences	Individuals high in verbal and low in spatial working memory abilities would perform significantly better in the auditory and written condition than in the pictorial condition.
12	Working Memory/ Individual Differences	Individuals high in spatial in and low in verbal working memory abilities would perform significantly better in the pictorial condition than in the auditory or written condition.

EXPERIMENTAL METHOD

Equipment

The experimental system consisted of two separate tasks: a) the Warning Color-Combination task and b) Tom Clancy's Ghost Recon® task that will later be described in detail. The Warning Color-Combination (WCCOM) compliance task and the Ghost Recon task were presented on two separate computers (Dell Dimension 8200 desktops), with two monitors (17" and 19" flat screens), two keyboards, and mice. The computer used for the WCCOM had two speakers (Cambridge Soundworks) which was used to present auditory information. The computer used for the Ghost Recon task did not have speakers, thus no noise was emitted during the task. The two monitors were placed on a desk side-by side in order for participants to easily view both monitors (See Figure 3).

Figure 3: This picture depicts the experimental setup with monitors, keyboards, mice, and speakers.

WCCOM Compliance Task

WCCOM was one of the two tasks in the dual task paradigm. The WCCOM was a warning compliance task where participants were required to respond/comply with warning messages that were presented and retained in one of three modalities. The WCCOM database consisted of ten different warnings that required behavioral compliance (boots, earmuffs, glasses, gloves, helmet, shield, suit, respirator, meter, or mask). For example, the boots warning was an indicator to the operator to don protective footwear before entering a restricted area. The warning messages in this experiment were all occupational warnings that were used to promote mandatory action. All of the symbols in the current program of study followed the ANSI (ANSI Z535.3, 1991) standards and were also incorporated into the Australian Standard (AS 1319, 1979). Six of these symbols were tested by Cairney and Sless (1982) to find out which pictorials were most easily recognized and learned. All six of the ten mandatory action safety symbols used in the current program of study were found to be the most easily recognized and learned.

Warnings were paired with one of ten colors (red, blue, green, orange, purple, black, white, gray, brown, or yellow; recommended by the ANSI Z535.2, 1991). These colors were tested in order to validate if the colors were easily recognized. Four participants (two males and two females, mean age = 27.5) viewed all ten colors on a computer monitor and correctly identified (100%) the colors.

The WCCOM had both storage and processing requirements. The storage requirement entailed memorizing the color associated with each warning at each level of task demand. The recall requirement involved recalling the stored WCCOMs. The WCCOM was presented in one of three modalities, pictorial (See Figure 4), written (See Figure 5), or verbal. Operational stress level was manipulated by the working memory demands of the WCCOM. The number of association cues that the operator stored and retrieved increased in demand from two to four to eight associations. These associations, the WCCOM, were stored and later processed.

The pictorial and written WCCOM were presented for five seconds in the center of the computer screen, after a brief pause the next WCCOM appeared on the screen for the same duration of time. The verbal WCCOM was presented verbally via speakers in the same manner that the written and pictorials were presented. In all modalities of the WCCCOM presentation, a short beep sounded preceding the WCCOM presentation. The beep was implemented in order to prevent startle effects. This pattern continued until all of the warning-color combinations, at each task demand level, was exhausted (2, 4, or 8). Each combination of color and warning was paired randomly and appeared only once per block.

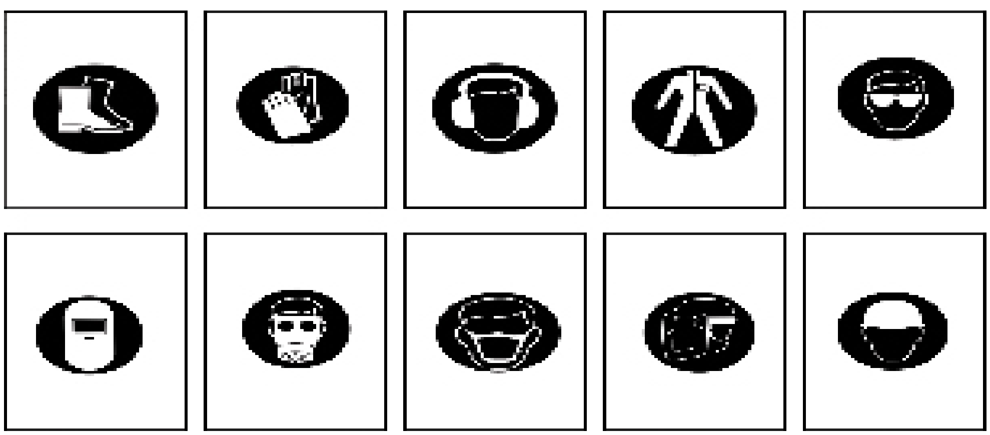


Figure 4: Pictorial Warnings for the WCCOM.

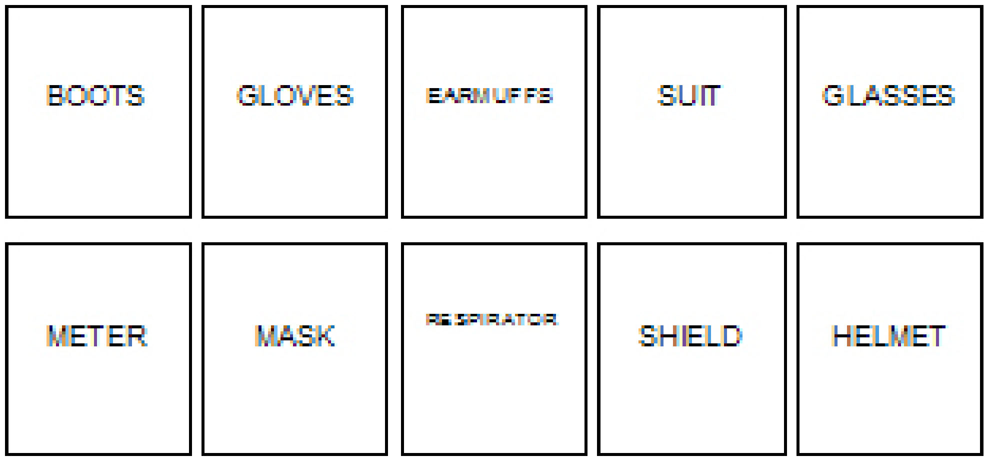


Figure 5: Written Warnings for the WCCOM.

The WCCOM pictorial was paired with a rectangle filled with one of the ten colors (See Figure 6). The written presentation of the warning was spelled out with the beginning letter capitalized in 80-point font in Arial black (with the exception of earmuffs and respirator, which were presented in 66-point font). The written warnings were paired with a written color (spelled out in the color of the pair) in the same size and font (See Figure 7). The verbal WCCOM was presented via speakers. For example, the participants heard "*boots...black*".

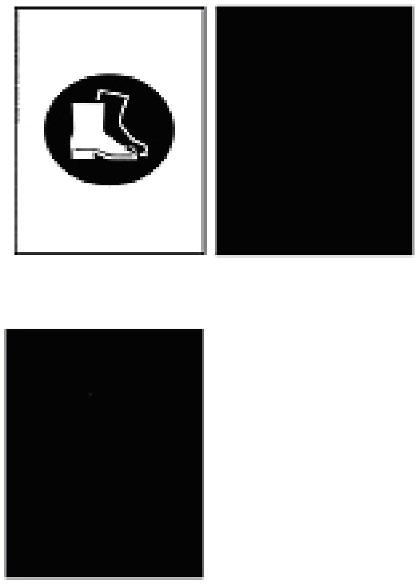


Figure 6: Example of a pictorial WCCOM (top) and the color stimulus (bottom) that elicited the key press response during the WCCOM portion of the dual task. In this particular example, the warning, boots, was combined with the color black.

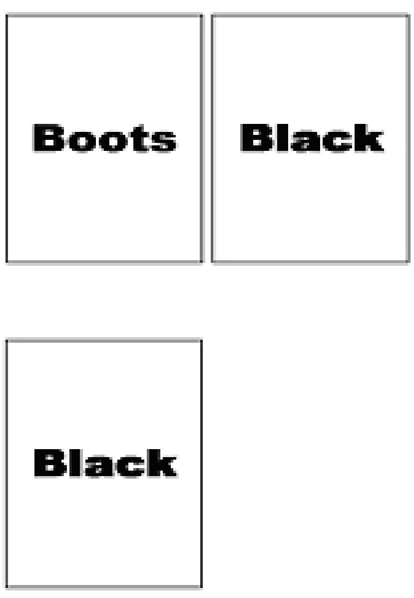


Figure 7: Example of a written WCCOM (top) and the color stimulus (bottom) that elicited the key press response during the WCCOM portion of the dual task. In this particular example, the warning, boots, was combined with the color black.

Ghost Recon Task

The second component in the dual task setting consisted of interacting with Tom Clancy's Ghost Recon® produced by Redstorm Entertainment, a commercially available first-person shooter video game. Participants were given written and verbal instructions on how to maneuver through the Ghost Recon environment using the arrow keys on the keyboard with their left hand and the mouse with their right hand. During each trial, participants completed a two-minute mission in Ghost Recon while simultaneously responding to the WCCOM.

Two different tasks were developed in Ghost Recon, a shooting task and a navigation task. One of the two environments was utilized in the experimental system. The Ghost Recon shooting task took place in an urban setting where participants fulfilled their task objective in a building (See Figure 8). The Ghost Recon objective in this experiment was to clear the building of all enemies (shoot all enemies) and participants were informed that anyone in the building was an enemy. In addition, participants were told not to leave the building for any reason. The enemies were strategically placed throughout the building and the amount of enemies in any one building ranged from five to seven. The task difficulty did not vary from building to building. This task was a visual/spatial task that involved little working memory resources. Participants were not aware of the amount of enemies in the building and did not get feedback as to how many enemies they killed compared to enemies in the building.

Performance for Experiments 1 and 2 was measured by calculating the number of enemies that the participant killed. The measure was changed for Experiments 3, 4, and 5 to a more sensitive measure, calculating the number of enemies that the participant killed by the number of enemies in that mission. Ten different missions were designed so that participants would not become too familiar with any one mission. The missions were randomly assigned to the three blocks of trials.

Figure 8: This picture depicts the Ghost Recon shooting task used in the experimental system.

The Ghost Recon navigation task took place in a rural setting where participants fulfilled their task objective in a sparsely wooded forest (See Figure 9). The objective in this experiment was to navigate sequentially from waypoint 1 to waypoint 4. A military tank marked each waypoint. The navigation task was also a visual/spatial task. This task involved navigation thus taxing spatial working memory more so than then shooting task. The navigation task was designed to be a more cognitively difficult task compared to the shooting task.

There were four waypoints in each mission. Participants were to begin at waypoint 0. Waypoint 0 was represented by a tank with a gun. The direction that the gun was facing was the direction the participant was to navigate to reach waypoint 1. No other navigational aids were provided. Once the participant was out of view of the tank (waypoint), they had to use their spatial working memory ability to navigate to the next waypoint. When participants reached the fourth waypoint, the task was successfully complete. Seven different navigational missions were designed so that participants would not become too familiar with any one mission. The tanks were strategically placed throughout the rural terrain. The task difficulty did not vary from mission to mission. The navigational missions were randomly assigned to the three blocks of trials.

Two performance measures were used. The first measure only accounted for the number of waypoints reached (in percent/ 0-100%). The second performance measure accounted for time, the amount of time it took to complete the task. Performance was measured by comparing the time it took participants to reach all four waypoints in a particular mission to the goal time for that mission, the fastest time that the participant could have navigated to all four waypoints. The equation that was used to determine performance was, (actual time-goal time)/goal time X 100. For waypoints that were not reached, a 200% was added to the equation. Thus, participants that did not reach a waypoint would not benefit by having a "0" average into their actual time (therefore giving them a time advantage).

Figure 9: This picture depicts the Ghost Recon navigation task used in the experimental system.

Table 2
Comparison of the Operational Tasks

Operational Task	Task Objectives	Performance Measure	Feedback	Spatial Complexity
Shooting Task	Kill all enemies (no friend/foe discrimination)	1)Number of enemies killed/number of enemies present	No feedback	Low
Navigation Task	Navigate from waypoint 0-4	1) Number of waypoints reached 2) Time to complete task	Feedback given at each waypoint	High

Intervening Card Sorting Task

One complete deck of playing cards was used to administer the card sorting task. In this task, the participants were asked to separate the cards into two piles, one pile for face cards and one pile for number cards. This task was used as an attempt to diminish carry over effects from the previous trial.

Individual Differences- Memory Related Tasks

Four working memory tasks were used in order to predict the processing and storage capacity of working memory. In order to test for spatial working memory capacity the spatial span was administered. Likewise, to test for verbal working memory capacity both the verification word task and the reading span were administered. Finally, to tap both verbal and spatial processing the verification arrow task was administered.

Spatial Working Memory Task

Spatial span. The spatial span task (Shah & Miyake, 1996) consisted of presenting participants with a set of English capital letters (F, J, L, P, and R) and their mirror images one at a time, each appearing in different orientations (See Figures 10-12). The objective of this task was to remember the orientation of each letter in the correct order, while deciding if the letter was normal or mirrored as quickly and accurately as possible. Each letter was presented for 2200 milliseconds in one of seven possible orientations in 45˚ increments, excluding the upright position. The participants were asked to respond aloud to indicate whether the letter was a normal or mirrored image. After the entire set of letters was presented in a trial, participants were asked to recall in serial order the orientation of the letters by clicking on the appropriate button orientation on a grid (for more details see Shah & Miyake, 1996). The span task included 20 letter sets (5 sets at each size, ranging from two to five letters), and participants were presented with increasingly longer sets of letters.

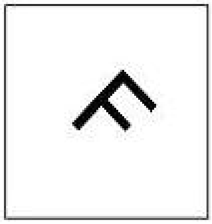


Figure 10: A “normal” F rotated at a 45˚ angle.

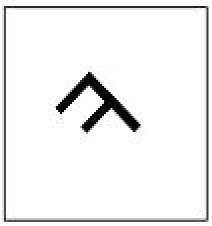


Figure 11: A “mirrored” F rotated at a 315˚ angle.

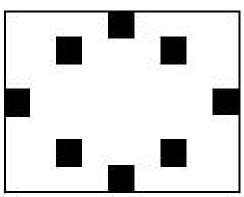


Figure 12: The recall grid.

Verbal Working Memory Tasks

Reading span. The reading span task (Daneman & Carpenter, 1980) was the analog to the spatial span task. In the reading span task participants read aloud a set of unrelated sentences one at a time and recalled the last word in each sentence. One example of a reading span sentence was “It was the movers that the couch dropped”. Participants were to recall the last word in the sentence, “dropped”. After the entire set of sentences was presented in a trial participants were asked to recall in serial order the last words in each sentence by typing them in to the “recall” box at the bottom of the computer screen. There were 20 sentence sets (5 sets at each size, ranging from two to five sentences) and participants were presented with increasingly longer sets of letters.

Verification word task. The verification word task (Shah & Miyake, 1996) was the analog to the verification arrow task. Again, participants were to decide if the sentence was a true statement or a false statement by pressing a button at the bottom of the screen labeled “True” or “False”. Following the sentence, a word appeared on the screen for 800 milliseconds. After the entire set of sentences was presented in a trial, participants were asked to recall in serial order the words by typing them in to the “recall” box at the bottom of the computer screen. There are 20 sentence sets (5 sets at each size, ranging from two to five sentences) and participants were presented with increasingly longer sets of sentences. The word in the verification word task was from a list of the most frequently used words in the English language according to Frances and Kucera (1982). Of the 275 most frequently used words, 70 two-syllables nouns (excluding proper nouns) were selected from the list and were only used once in the task.

Verbal-Spatial Working Memory Task

Verification Arrow. The verification arrow task was a combination of verbal and spatial processing. The verification arrow task (Shah & Miyake, 1996) consisted of reading short sentences (sentences ranged from three to six words in a sentence) and deciding if the sentence was a true statement or a false statement by pressing a button at the bottom of the screen labeled “True” or “False”. The sentences were the language-processing portion of the span. One example of a short sentence used for the verification arrow task was, “The world is flat”. The participant should have responded by pressing the “False” button. Following the sentence, an arrow appeared on the screen for 800 milliseconds in one of seven possible orientations in 45˚

increments, excluding the upright position (See Figure 13). The arrow portion of the span was the spatial processing portion of the task. After the entire set of sentences was presented in a trial, participants were asked to recall in serial order the orientation of the arrows by clicking on the appropriate button orientation on a grid (for more details see Shah and Miyake, 1996). The verification arrow task included 20 sentence sets (5 sets at each size, ranging from two to five sentences), and participants were presented with increasingly longer sets of sentences.

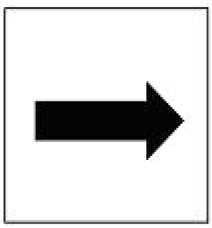


Figure 13: An arrow rotated at a 90˚ angle.

Subjective Measures

All of the questionnaires were administered via the experimental software program Inquisit Version 1.32 (Millisecond Software, 2002) on a Dell Dimension 8200 desktop computer. The questionnaires administered were the Rating Scale Mental Effort (RSME; (Zijlstra & Van Doorn, 1985, Zijlstra & Meijman, 1989, Zijlstra, 1993) and the NASA-Task Load Index (NASA-TLX; Hart & Staveland, 1988).

Rating Scale Mental Effort

The RSME is a one-dimensional scale that measures the amount of invested effort exerted during a task (see Appendix A). The scale's range is from 0-150 mm and a hash mark is placed at every 10 mm. Anchor points are identified at several locations on the scale, describing the mental effort invested, such as 'almost no effort' or 'extreme effort'. The RSME is measure by the number that is identified as the invested mental effort for a given task from 0-150. The higher the score, the more subjective mental effort was exerted. The RSME was used between trials in order to determine differences between sizes of memory set (task demand). Thus, it was administered fifteen times over the course of one block (45 times total). The RSME is not a validated study, but is a reliable measure which has been used extensively to measure workload (deWaard, 2001; Hilburn, Bakkas, Pekela, & Parasurman, 1997; Neerincx & Ruijsendaal, & Wolf, 2001).

NASA Task Load Index

Developed by Hart and Staveland (1988), the NASA-Task Load Index (NASA-TLX) is a multi-dimensional scale that has six subscales. The six subscales are mental demand, physical demand, effort, performance, frustration, and temporal demand (see Appendix B). These subscales are calculated to give an overall rating of workload or it can be individually rated for each subscale. In order to get an overall rating, the six subscales are arranged into paired comparisons and participants make 15 comparisons of each of the six subscales and rate which one of the two contributed to the participant's workload. The second part of the NASA-TLX is the weighted scales. The participant rates on 0-100 scale, 0 being low workload and 100 being high workload for each of the six subscales. Finally, the two scales are calculated for the final workload rating.

Experimental Procedure

An all within-participants design was used which consisted of three blocks (one for each modality: verbal, written, and pictorial); each block consisted of 15 trials (5 repetitions for each task demand level at 2, 4, and 8 warnings). Presentation order of the modality conditions was counterbalanced between blocks to avoid the effect of sequencing. Additionally, the presentation order of trials within each block was randomized.

Testing occurred in two sessions (approximately 2.5 hours each) on different days during one week. During Session 1, participants were asked to complete the informed consent, and a demographic questionnaire. After a five-minute break, the working memory tasks were administered followed by another five-minute break. The participants completed a practice session, which consisted of three trials of the Ghost Recon task, the WCCOM, and both tasks simultaneously (dual task setting). The participant's task was to remember the correct pairing of the warning and color combinations. The experimental task required the participants to first store the WCCOM at each level of task demand, followed by the processing requirement of the task. The processing portion of the task involved complying with the warning stimulus while simultaneously performing the operational task. When participants either saw or heard (depending on the modality of presentation in that block) the warning, they were to respond by pressing the appropriately labeled key on a second keyboard (keys 'q' through 'p' are labeled with the warning portion of each combination) with their right hand or verbally respond via a microphone.

Following the practice sessions, participants completed the first block. Session 2 consisted of the remaining two blocks of the experiment. A five-minute break was scheduled between blocks.

During both sessions, the Rating Scale Mental Effort (RSME; Zijstra, & Van Doorn, 1985) and the card-sorting task were administered following each trial and the NASA-Task Load

Index (NASA-TLX; Hart & Staveland, 1988) followed each block. Finally, participants were debriefed via a verbal and written statement.

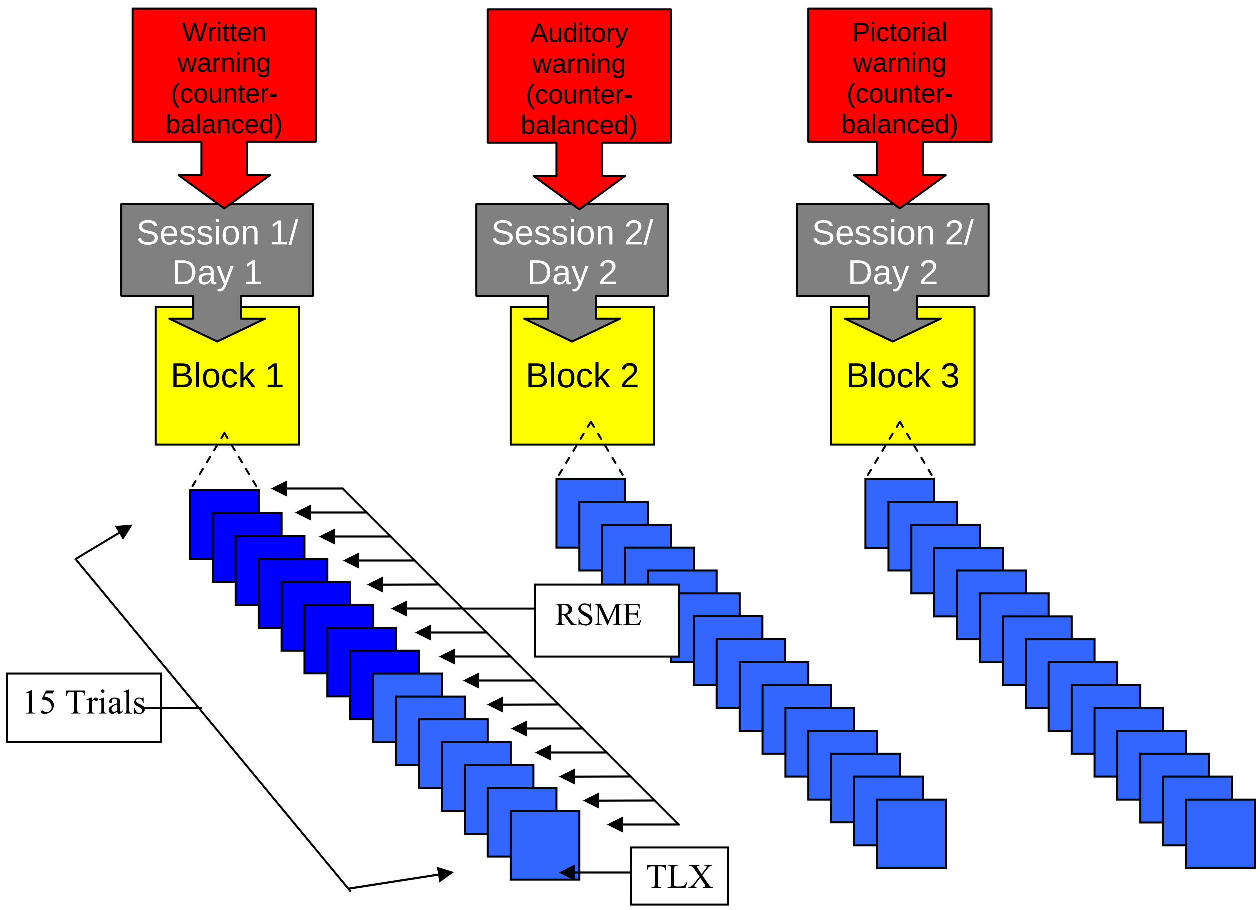


Figure 14: Visual Representation of the Experimental Procedure.

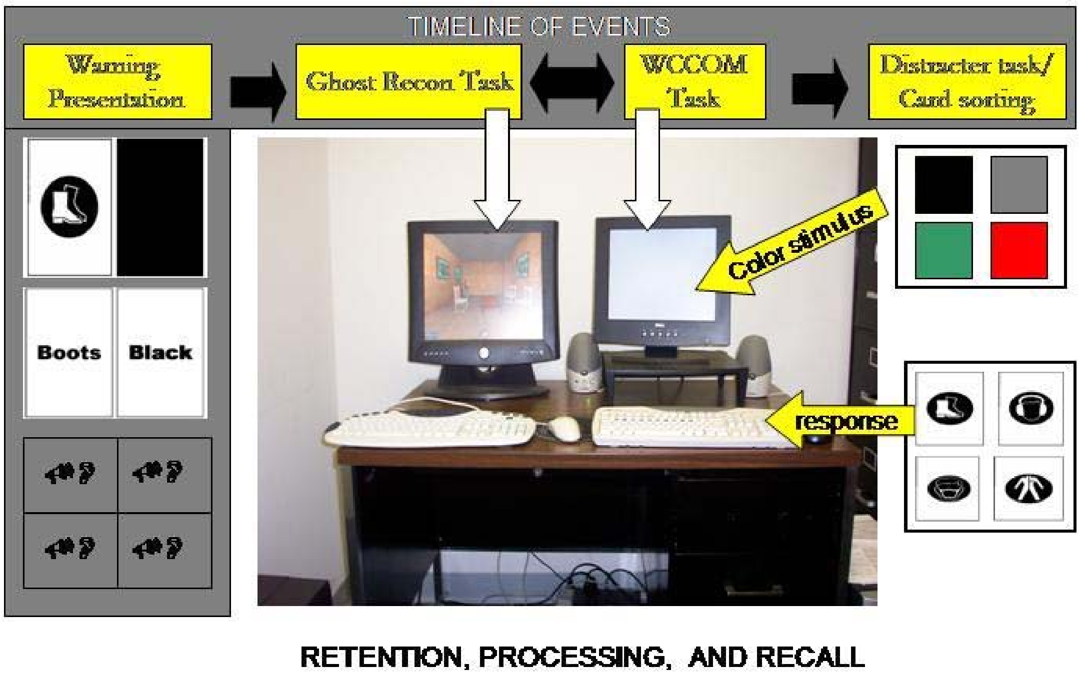


Figure 15: Visual Representation of the Experimental Task Set-up.

This line of research consisted of eight experiments. Table 2 describes the variables that were manipulated for each experiment. Experiments 1 varied from experiment two because the response mode for the WCCOM task was altered from pictorial to written response mode to identify if the format of response affected compliance. The sample size for Experiments 1 and 2 were small, therefore the sample size was increased from 6 to 11 (12 in the latter experiments). Furthermore, additional performance measures were taken on the shooting task. Experiment five introduced the third format of response mode, verbal in order to identify if differences were found between the three response formats. Finally, in Experiments 6, 7, and 8 a more cognitively complex Operational task task was implemented, a navigation task. The manipulations that occurred in each of these experiments will be further discussed at each experiment's method section.

Table 3
Overview of the Experimental Line of Research

Experiments	WCCOM Response Mode	Operational Task	Experiment Variation
1	Pictorial	Shooting	
2	Written	Shooting	Response Mode
3	Pictorial	Shooting	Sample Size
4	Written	Shooting	Response Mode
5	Verbal	Shooting	Response Mode
6	Pictorial	Navigation	Operational Task
7	Written	Navigation	Response Mode
8	Verbal	Navigation	Response Mode

EXPERIMENT 1

Experiment 1 was conducted in order to a) determine whether task-based stress (expressed by increasing working memory demand) would have systematic effects on performance across the modalities of verbal, written, and pictorial presentation in simulated operational conditions, b) identify if the modality of the warning presentation effected compliance behavior when the second task was predominantly a visual and spatial task, c) investigate the effects of modality and task load on workload, and d) begin accumulating data on individual differences in working memory capacity.

Method

Participants

Six undergraduate students from the University of Central Florida (5 females and 1male, mean age = 20.2 years) were recruited on a voluntary basis from the university's experimental recruiting website (www.experimetrix.com/ucf). Participants were paid approximately $37.50 for their participation (based on an hourly rate of $7.50). Participants could only participate in one of the eight studies. Participants were not used in any other experiments in the series of studies so that they would not become familiar with the task or become "expert users".

Materials

WCCOM Compliance Task

The WCCOM compliance task for Experiment 1 was consistent with the general materials used for the experimental system. The keyboard in this task had keys “q” through “p” labeled with a pictorial representation of the warning portion of each combination.

Ghost Recon Task

The Ghost Recon shooting task was used for Experiment 1 and was consistent with the general materials used for the experimental system.

Procedure

The procedures for Experiment 1 were consistent with the general procedure used for the experimental System.

Results for Experiment 1

Separate headings have been devised in order to separate the WCCOM task analysis from the Ghost Recon Task as well as the NASA-TLX scores and the RSME. The individual difference data will be presented in the general results section.

WCCOM Task

In order to determine whether the size of the memory set (task-based stress) had a systematic effect on performance across varying formats of warning presentation, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance

(ANOVA) was conducted on warning compliance behavior. The independent variables were format, task demand, and trial. The three levels of warning format: verbal, pictorial, and written were presented to the participant at varying levels of task demand: presentations of two, four, and eight warning-color combinations, over forty-five trials: 15 trials per block, five trials per level of task demand. The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination) at each level of task demand.

Analyses were performed using SPSS for Windows, 11.5 and an alpha level of .05 was used for all analysis. A conservative approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick & Fidell, 2001). Thus, all six participants were used in all analyses.

There was a significant main effect of format and task demand, Wilk's $\Lambda = .115$, $F(2, 10) = 18.1$, $p = .0005$, partial $\eta^2 = .78$, and Wilk's $\Lambda = .73$, $F(2, 10) = 26.4$, $p = .0005$, partial $\eta^2 = .84$, respectively. No main effect was found for trial, Wilk's $\Lambda = .54$, $F(4, 20) =.43$, $p > .05$, partial $\eta^2 = .08$. A significant interaction was found for task demand and trial, $F(8, 40) = 9.7$, $p = .03$, partial $\eta^2 = .66$.

A set of Fisher LSD post hoc tests for format type showed that participants were significantly more likely to comply when the information was presented in the written, ($M = .721$, $SD = .056$) and pictorial ($M = .733$, $SD = .061$) formats than in the verbal format ($M = .526$, $SD = .054$). Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two ($M = .789$, $SD = .035$) and four ($M = .697$, $SD = .070$) than at level eight ($M = .494$, $SD = .07$), but no significant differences in compliance were found between levels two (*M* = .789, *SD* = .035) and four (*M* = .697, *SD* = .070).

Table 4
The ANOVA Table for the WCCOM Task for Experiment 1

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	1.212	18.106	.000	.784	.997
Error(format)	10	.067				
Task Demand	2	2.043	26.428	.000	.841	1.000
Error(Task Demand)	10	.077				
Trial	4	.031	.433	.783	.080	.130
Error(Trial)	20	.072				
Format * Task Demand	4	.034	.612	.659	.109	.168
Error(Format*Task Demand)	20	.055				
Format * Trial	8	.081	1.225	.310	.197	.486
Error(Format*Trial)	40	.066				
Task Demand * Trial	8	.257	4.280	.001	.461	.984
Error(Task Demand*Trial)	40	.060				
Format * Task Demand * Trial	16	.075	.804	.677	.139	.495
Error(Format*Task Demand*Trial)	80	.094				

Computed using alpha = .05

Ghost Recon Shooting Task

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the amount of enemies killed. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15 trials. The dependent variable was the percent of enemies killed in each building (enemies in the building divided by enemies killed). There were no significant main effects of format, Wilk's Λ = .95, $F(2, 10) = .11$, $p > .05$, partial $\eta^2 = .06$, Wilk's

Λ = .51, task demand, $F(2, 10) = 1.8$, $p > .05$, partial $\eta^2 = .29$, or trial, Wilk's $\Lambda = .39$, $F(4, 20) = 3.3$, $p > .05$, partial $\eta^2 = .23$. A significant interaction was found for task demand by trial $F(8, 40) = 2.54$, $p = .024$, partial $\eta^2 = .34$.

Table 5
The ANOVA Table for the Ghost Recon Shooting Task for Experiment 1

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.002	.108	.898	.021	.062
Error(Format)	10	.019				
Task Demand	2	.039	1.771	.220	.262	.286
Error(Task Demand)	10	.022				
Trial	4	.018	.886	.490	.150	.231
Error(Trial)	20	.020				
Format * Task Demand	4	.010	.328	.856	.062	.109
Error(Format*Task Demand)	20	.030				
Format * Trial	8	.015	1.048	.418	.173	.417
Error(Format*Trial)	40	.014				
Task Demand * Trial	8	.032	2.547	.024	.337	.854
Error(Task Demand*Trial)	40	.012				
Format * Task Demand * Trial	16	.018	.940	.528	.158	.578
Error(Format * Task Demand * Trial)	80	.019				

Computed using alpha = .05

NASA-TLX

In order to determine if format of presentation affected subjective mental workload a one-way, analysis of variance (ANOVA) was conducted. The independent variable, warning format, included three levels: verbal, written and pictorial warnings. The dependent variable was the subjective mental workload effort determined by using the NASA-task load Index (NASA-TLX). The analysis yielded non-significant results, Wilk's $\Lambda = .87$, F (2, 10) = .55, $p>.05$, partial $\eta^2 = .98$. Thus, participants did not report feeling more mental demand in one warning format vice another format.

RSME

A three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the Rating Scale Mental Effort (RSME) subjective workload measure. The independent variables were format, task demand, and trial. The dependent variable of interest was the RSME. Results of this analysis yielded a main effect for warning format Wilk's Λ = .11, $F(2,10) = 32.1$, $p = .0005$, partial η^2=.87. Task demand and trial yielded non-significant results Wilk's Λ = .89, $F(2,10) = .39$, $p > .05$ partial $\eta^2 = .07$, Wilk's Λ = .23, $F(4,20) = 1.2$, $p > .05$, partial $\eta^2 = .19$ respectively. No significant interactions were found.

A set of Fisher LSD post hoc tests for format type showed that participants had significantly higher mental workload when the information was presented in the written, (*M* =59.2, *SD* = .738) and pictorial (*M* = 42.3, *SD* = 9.36) formats than in the verbal format (*M* = 25.8, *SD* = 9.78). Furthermore, analysis yielded a significant difference between written (*M* = 59.2, *SD* =7.38) and pictorial format (*M* = 42.3, *SD* = 9.36), thus participants felt more mental workload in the written condition.

Table 6
The ANOVA Table for the RSME for Experiment 1

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Task Demand	2	366.693	.385	.690	.072	.096
Error(Task Demand)	10	951.470				
Format	2	25167.670	32.116	.000	.865	1.000
Error(Format)	10	783.648				
Trial	4	263.491	1.203	.340	.194	.308
Error(Trial)	20	218.991				
Task Demand * Format	4	1126.331	2.844	.051	.363	.664
Error(Task Demand * Format)	20	396.076				
Task Demand * Trial	8	510.466	2.171	.051	.303	.780
Error(Task Demand * Trial)	40	235.132				
Task Demand * Trial	8	260.541	1.236	.304	.198	.490
Error(Format * Trial)	40	210.741				
Task Demand * Format * Trial	16	256.799	1.605	.087	.243	.863
Error(Task Demand * Format * Trial)	80	160.016				

Computed using alpha = .05

Discussion for Experiment 1

Experiment 1 was conducted to determine if a) the format of warning presentation affected behavioral compliance (Hypotheses 1-2), b) increases in task based stress (task-demand) affected compliance behavior (Hypotheses 3-4) and c) task based stress increases affected subjective workload ratings (Hypotheses 5-7).

Hypothesis 1 on warning format predicted that participants would have significantly higher rate of compliance behavior when warnings were presented in verbal compared to written and pictorial. Additionally, Hypothesis 2 also predicted verbal warning would be the superior mode of warning presentation because warnings presented verbally would have less interference on the operational task. Contrary to these hypotheses, pictorial and written warnings resulted in greater behavioral compliance than the verbal warnings. Furthermore, not only were pictorials and written warnings superior to verbal warnings, no significant effect of format was found for the operational task. Thus, these results suggest that the pictorials and written warnings did not interfere with the operational task (visual/spatial task).

For the hypotheses on task demand, Hypotheses 3 and 4, it was predicted that when two or four warnings were presented performance on neither the WCCOM task nor the operational task would be affected; yet when the warning presentation increased to eight, performance on both tasks would degrade. Consistent with these hypotheses, the WCCOM task was affected by the task based stress. When only two or four warnings were presented, no differences were found, yet a significant decrease in compliance resulted when eight warnings were presented. Contrary to what was predicted, was not affected by task demand. It was predicted in hypotheses 6 and 7 that subjective workload rating scores would be significantly lower when two warnings were presented compared to four warnings. Furthermore, it was predicted that eight warnings would exceed participants' resources and thus, eight warnings would not be associated with the workload. Inconsistent with what was hypothesized, no differences were found for task demand.

Due to the inconsistent findings with our hypotheses, an additional experiment was designed. In particular, Experiment 2 was conducted to isolate the aforementioned issues related to response format. In this experiment, pictorial response format was used across all formats of warning presentation. In Experiment 2, written response format replaced the pictorial format in order to determine if the results that were found for warning format were due solely to the response format. All other variables for Experiment 2 remained the same as Experiment 1.

EXPERIMENT 2

Experiment 2 was a replica of Experiment 1 except that it was conducted in order to determine if changing the response mode on the WCCOM task from pictorial to written words would affect performance across warning format and task load.

Method

Participants

Six undergraduate students from the University of Central Florida (4 females and 2 males, mean age = 21 years) were recruited on a voluntary basis from the university's experimental recruiting website (www.experimetrix.com/ucf). Participants were paid approximately $37.50 for their participation (based on an hourly rate of $7.50). Participants could only participate in one of the eight studies. Participants were not used in any other experiments in the series of studies so that they would not become familiar with the task or become "expert users".

Materials

WCCOM Compliance Task

The WCCOM compliance task for Experiment 2 was consistent with the general materials used for the experimental system. The keyboard in this task had keys "q" through "p" labeled with a written word representing the warning portion of each combination.

Ghost Recon Task

The Ghost Recon shooting task that was used for Experiment 2 was consistent with the general materials used for the experimental system.

Procedure

The procedures for Experiment 2 were consistent with the general procedure used for the experimental system.

Results for Experiment 2

WCCOM Task

In order to determine whether the size of the memory set (task-based stress) had a systematic effect on performance across varying formats of warning presentation, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on warning compliance behavior. The independent variables were format, task demand, and trial. The three levels of warning format: verbal, pictorial, and written were presented to the participant at varying levels of task demand: presentations of two, four, and eight warning-color combinations, over forty-five trials: 15 trials per block, five trials per level of task demand. The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination) at each level of task demand.

Analyses were performed using SPSS for Windows, 11.5 and an alpha level of .05 was used for all analysis. A conservative approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick & Fidell, 2001). Thus, all data from the 6 participants were used in the analysis.

There was a significant main effect of format, Wilk's Λ = .21, F(2, 10) = 8.5, p = .007, partial η^2 = .63, task demand, Wilk's Λ = .086, F(2, 10) = 17, p = .0005, partial η^2 = .78, and trial Wilk's Λ = .21, F(4, 20) = 4, p = .015, partial η^2 = .44, respectively. Significant interactions were found for format by task demand, F(4, 20) = 2.9, p = .05, partial η^2 = .37, format by trial, F(8, 40) = 2.4, p = .03, partial η^2 = .32, task demand by trial, F(8, 40) = 2.4, p = .03, partial η^2 = .33.

A set of Fisher LSD post hoc tests for format type showed that participants were significantly more likely to comply when the information was presented in the written, (*M* = .729, *SD* = .68) and pictorial (*M* = .719, *SD* = .073) formats than in the verbal format (*M* = .549, *SD* = .042). No significant differences were found between written, (*M* = .729, *SD* = .68) and pictorial (*M* = .719, *SD* = .073) formats

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two (*M* = .806, *SD* = .023) and four (*M* = .678, *SD* = .079) than at level eight (*M* = .514, *SD* = .069), but no significant differences in compliance were found between levels two (*M* = .806, *SD* = .023) and four (*M* = .678, *SD* = .079).

Fisher LSD post hoc tests were also conducted for trial which yielded trial one (*M* = .713, *SD* = .063) as having greater compliance scores than trial five (*M* = .581, *SD* = .055). No significant differences were found between any other trials.

In addition, Duncan post hoc analyses were conducted on the format by task demand, $F(4, 20) = 2.9$, $p = .05$, partial $\eta^2 = .37$, interaction yet the analysis revealed no significant interactions for formats and task demands at the same level of task demand or format type.

Table 7
The ANOVA Table for the WCCOM Task for Experiment 2

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.928	8.460	.007	.629	.891
Error(format)	10	.110				
Task Demand	2	1.924	17.974	.000	.782	.997
Error(Task Demand)	10	.107				
Trial	4	.151	3.994	.015	.444	.825
Error(Trial)	20	.038				
Format * Task Demand	4	.200	2.917	.047	.368	.677
Error(Format*Task Demand)	20	.069				
Format * Trial	8	.137	2.373	.034	.322	.823
Error(Format*Trial)	40	.058				
Task Demand * Trial	8	.180	2.433	.030	.327	.834
Error(Task Demand*Trial)	40	.074				
Format * Task Demand * Trial	16	.093	2.108	.016	.297	.953
Error(Format*Task Demand*Trial)	80	.044				

Computed using alpha = .05

Ghost Recon Shooting Task

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the amount of enemies killed. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15 trials. The dependent variable was the percent of enemies killed in each building (enemies in the building divided by enemies killed). There were no significant main effects of format, Wilk's $\Lambda = .92$, $F(2, 10) = .85$, $p > .05$, partial $\eta^2 = .15$, task demand, Wilk's Λ = .6, $F(2, 10) = 2.8$, $p > .05$, partial $\eta^2 = .43$, or trial, Wilk's Λ = .15, $F(4, 20) = 1.5$, $p > .05$, partial $\eta^2 = .23$. No significant interactions were observed.

NASA-TLX

In order to determine if format of presentation affected subjective mental workload ratings, a one-way, analysis of variance (ANOVA) was conducted. The independent variable, warning format, included three levels: verbal, written and pictorial warnings. The dependent variable was the subjective mental workload effort determined by using the NASA-task load Index (NASA-TLX). The analysis yielded non-significant results, Wilk's Λ = .85, F (2, 10) = .28, p>.05 partial $\eta^2 = .06$. Thus, participants did not report feeling more mental demand in one warning format vice another format.

RSME

A three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the Rating Scale Mental Effort (RSME) subjective workload measure. The independent variables were format, task demand, and trial. The dependent variable of interest was the RSME. Results of this analysis yielded a main effect for task demand, Wilk's Λ = .18, $F(2, 10) = 18.3$, $p = .0005$, partial $\eta^2 = .79$. Warning format and trial yielded non-significant results Wilk's Λ = .92, $F(2,10) = .32$, $p > .05$, partial $\eta^2 = .06$, Wilk's Λ = .45, $F(4,20) = .19$, $p > .05$, partial $\eta^2 = .04$, respectively. No significant interactions were found.

A set of Fisher LSD post hoc tests for task demand showed that participants felt the least amount of mental workload at level two (M = 12, SD = 46) compared to level four (M = 35, SD =

78) or eight (*M* = 64, *SD* = 11). Furthermore, significant differences in subjective workload ratings were found between level four (*M* = 35, *SD* = 78) and eight (*M* = 64, *SD* = 11). Therefore, participants workload increased as the task demand increased.

Table 8
The ANOVA Table for the RSME for Experiment 2

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	399.953	.319	.734	.060	.088
Error(Format)	10	1254.642				
Task Demand	2	61406.275	18.297	.000	.785	.997
Error(Format)	10	3356.031				
Trial	4	47.575	.193	.939	.037	.083
Error(Trial)	20	246.731				
Task Demand * Format	4	890.644	2.437	.081	.328	.588
Error(Task Demand * Format)	20	365.483				
Task Demand * Trial	8	298.140	1.400	.226	.219	.552
Error(Task Demand * Trial)	40	212.954				
Task Demand * Trial	8	140.414	.814	.595	.140	.322
Error(Format * Trial)	40	172.586				
Task Demand * Format * Trial	16	130.815	.663	.821	.117	.404
Error(Task Demand * Format * Trial)	80	197.383				

Computed using alpha = .05

Discussion for Experiment 2

Experiment 2 was conducted to determine if pictorial and written warnings resulted in the superior format of warning presentation as it did in Experiment 1.

Hypothesis 1 on warning format predicted that participants would have significantly higher rate of compliance behavior when warnings were presented in verbal compared to written and pictorial. Results of Experiment 1 concluded that contrary to aforementioned hypotheses, pictorial and written warnings resulted in the superior format of warning presentation as compared to verbal warnings. Experiment 2 replicated Experiment 1, but with one variation, the response format was changed form pictorial to written.

Additionally, Hypothesis 2 also predicted verbal warning would be the superior mode of warning presentation because warnings presented verbally would have less interference on the operational task. Results of Experiment 2 also replicate the results of Experiment 1 in that no significant effect of format was found for the operational task.

Experiment 2 validates the pictorial and written superiority finding. Thus, altering the response format did not influence the results of the experiment on either the WCCOM task or the operational task. Verbal warning presentation remained the inferior format of warning presentation.

For the hypotheses on task demand, Hypotheses 3 and 4, it was predicted that when two or four warnings were presented performance on neither the WCCOM task nor the operational task would be affected; yet when the warning presentation increased to eight, performance on both tasks would degrade. Consistent with these hypotheses, the WCCOM task was affected by the task based stress. When only two or four warnings were presented, no differences were found, yet a significant decrease in compliance resulted when eight warnings were presented. Contrary to what was predicted, the operational task was not affected by task demand. These findings are consistent with Experiment 1.

It was predicted in Hypotheses 6 and 7 that subjective workload rating scores would be significantly lower when two warnings were presented compared to four warnings. As predicted, participants felt less workload at level two than at level four. Furthermore, it was predicted that eight warnings would exceed participants' resources and thus, eight warnings would not be associated between all levels of task demand. Results of this study did reveal that participants had the highest rate of mental workload at level eight.

Experiment 3 was conducted in order to determine if the results of Experiment 1 and 2 were sound. Experiment 3 was an exact replica of Experiment 1, thus the pictorial response mode was used for this study. Furthermore, the operational task remained the same, yet a more sensitive measure was taken, the number of enemies killed divided by the number of enemies, vice just the number of enemies killed. Additionally, a larger same size was used.

EXPERIMENT 3

Experiment 1 investigated 1) whether task-based stress (expressed by increasing working memory demand) had systematic effects on performance across the modalities of verbal, written, and pictorial presentation in simulated operational conditions, 2) if the modality of the warning presentation effected compliance behavior when the second task was predominantly a visual and spatial task, and 3) the effects of modality and task load on workload. Furthermore, accumulation of data on individual differences in working memory capacity began.

Experiment 2 replicated the objectives of Experiment 1 yet the response mode from pictorial to written on the WCCOM task was changed.

Experiment 3 and 4 were conducted in order to validate Experiment 1 and 2's experimental setting and with a larger sample size. In addition, a more sensitive measure was used to determine performance. In Experiments 1 and 2, the number of kills was used to determine performance, but in Experiments 3, 4, and 5, the kill percent was used. Kill percent was measured by the number of people killed divided by the number of people in the building.

Method

Participants

Eleven undergraduate students from the University of Central Florida (7 females and 3 males, mean age = 19.9 years (one missing demographic questionnaire)) were recruited on a voluntary basis from the university's experimental recruiting website (www.experimetrix.com/ucf). Participants were paid approximately $37.50 for their participation (based on an hourly rate of $7.50). Participants could only participate in one of the eight studies. Participants were not used in any other experiments in the series of studies so that they would not become familiar with the task or become "expert users". The number of participants in increased from six in Experiments 1-2 to eleven or twelve in Experiments 3-8 in order to get sufficient amount of participants to analyze the working memory data.

Materials

WCCOM Compliance Task

The WCCOM compliance task for experiment 3 was consistent with the general materials used for the experimental system. The keyboard in this task had keys "q" through "p" labeled with a pictorial representation of the warning portion of each combination.

Ghost Recon Task

The Ghost Recon shooting task for Experiment 3 was consistent with the general materials used for the experimental system with one exception. In Experiments 1 and 2, the number of kills was used to determine performance, but in Experiments 3, 4, and 5, the kill percent was used. Kill percent was measured by the number of people killed divided by the number of people in the building.

Procedure

The procedure for experiment 3 was consistent with the general procedure used for the experimental system.

Results for Experiment 3

Experiment 3 was conducted in order to validate Experiment 1 and 2's experimental setting and with a larger sample size. In addition, a more sensitive measure was used, the number of kills divided by the number of enemies, vice just the number of kills that were taken in Experiments 1 and 2.

WCCOM Task

In order to determine whether the size of the memory set (task-based stress) had a systematic effect on performance across varying formats of warning presentation, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on warning compliance behavior. The independent variables were format, task demand, and trial. The three levels of warning format: verbal, pictorial, and written were presented to the participant at varying levels of task demand: presentations of two, four, and eight warning-color combinations, over forty-five trials: 15 trials per block, five trials per level of task demand. The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination) at each level of task demand.

Analyses were performed using SPSS for Windows, 11.5 and an alpha level of .05 was used for all analysis. A conservative approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick & Fidell, 2001). Thus, all eleven participants were used in all analyses.

There was a significant main effect of format and task demand, Wilk's Λ = .366, $F(2, 20)$ = 11.7, p = .0005, partial η^2 = .54, and Wilk's Λ = .11, $F(2, 20)$ = 32.6, p = .0005, partial η^2 =

.77, respectively. No main effect was found for trial. A significant interaction was found for format and task demand, $F(4, 40) = 2.7$, $p = .04$, partial $\eta^2 = .21$.

A set of Fisher LSD post hoc tests for format type showed significant differences between all formats of warning presentation. Participants were significantly more likely to comply when the information was presented in pictorial (*M* = .735, *SD* = .05) than written (*M* = .63, *SD* = .057), or verbal (*M* = .506, *SD* = .04).

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two (*M* = .727, *SD* = .043) and four (*M* = .691, *SD* = .061) than at level eight (*M* = .453, *SD* = .03). No significant differences in compliance were found between levels two (*M* = .727, *SD* = .043) and four (*M* = .691, *SD* = .061).

Table 9
The ANOVA Table for the WCCOM Task for Experiment 3

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	2.171	11.725	.000	.540	.985
Error(format)	20	.185				
Task Demand	2	3.644	32.600	.000	.765	1.000
Error(Task Demand)	20	.112				
Trial	4	.081	.805	.529	.074	.234
Error(Trial)	40	.100				
Format * Task Demand	4	.123	2.728	.042	.214	.701
Error(Format*Task Demand)	40	.045				
Format * Trial	8	.040	.567	.802	.054	.245
Error(Format*Trial)	80	.071				
Task Demand * Trial	8	.078	.928	.498	.085	.403
Error(Task Demand*Trial)	80	.084				
Format * Task Demand * Trial	16	.064	.822	.659	.076	.545
Error(Format*Task Demand*Trial)	160	.077				

Computed using alpha = .05

Ghost Recon Shooting Task

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the percentage of enemies killed. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15trials. The dependent variable was the percent of enemies killed in each building (enemies in the building divided by enemies killed). There were no significant main effects of format, Wilk's Λ = .85, $F(2, 20)$ = .55, p > .05, partial η^2 = .05, Wilk's Λ = .55, task demand, $F(2, 20)$ = 1.6, p > .05, partial η^2 = .31, or trial, Wilk's Λ = .44, 3.1, $F(4, 40)$ = 1.1, p > .05, partial η^2 = .32. No significant interactions were found.

Table 10
The ANOVA Table for the Ghost Recon Shooting Task for Experiment 3

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.015	.554	.583	.053	.129
Error(Format)	20	.026				
Task Demand	2	.035	1.643	.218	.141	.305
Error(Task Demand)	20	.021				
Trial	4	.019	1.103	.368	.099	.315
Error(Trial)	40	.018				
Format * Task Demand	4	.009	.594	.669	.056	.180
Error(Format*Task Demand)	40	.015				
Format * Trial	8	.033	1.839	.082	.155	.742
Error(Format*Trial)	80	.018				
Task Demand * Trial	8	.019	.995	.447	.090	.432
Error(Task Demand*Trial)	80	.019				
Format * Task Demand * Trial	16	.018	1.064	.394	.096	.689
Error(Format * Task Demand * Trial)	160	.017				

Computed using alpha = .05

NASA-TLX

In order to determine if format of presentation affected subjective mental workload ratings, a one-way, analysis of variance (ANOVA) was conducted. The independent variable, warning format, included three levels: verbal, written and pictorial warnings. The dependent variable was the subjective mental workload effort determined by using the NASA-task load Index (NASA-TLX). The analysis yielded non-significant results, Wilk's $\Lambda = .85$, $F(2, 16) = .38$, $p > .05$, partial $\eta^2 = .05$. Thus, participants did not report feeling more mental demand in one warning format vice another format.

RSME

A three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the Rating Scale Mental Effort (RSME) subjective workload measure. The independent variables were format, task demand, and trial. The dependent variable of interest was the RSME. Results of this analysis did not yield a main effect for task demand, warning format, or trial, Wilk's $\Lambda = .62$, $F(2, 20) = 2.95$, $p > .05$, partial $\eta^2 = .23$, Wilk's $\Lambda = .80$, $F(2,20) = 2.3$, $p > .05$ partial $\eta^2 = .19$, Wilk's $\Lambda = .22$, $F(4,40) = 2.4$, $p > .05$, partial $\eta^2 = .19$, respectively. No significant interactions were found. Thus, participants did not report feeling a workload increase based on warning format, increase in task demand, or over repeated trials.

Table 11
The ANOVA Table for the RSME for Experiment 3

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Task Demand	2	2873.046	2.949	.075	.228	.510
Error(Task Demand)	20	974.116				
Format	2	8435.961	2.305	.126	.187	.412
Error(Format)	20	3659.254				
Trial	4	842.251	2.411	.065	.194	.640
Error(Trial)	40	349.269				
Task Demand * Format	4	32.520	.053	.995	.005	.060
Error(Task Demand * Format)	40	613.892				
Task Demand * Trial	8	118.970	.314	.959	.030	.145
Error(Task Demand * Trial)	80	378.937				
Task Demand * Trial	8	585.423	1.913	.069	.161	.763
Error(Format * Trial)	80	305.951				
Task Demand * Format * Trial	16	395.168	1.047	.411	.095	.680
Error(Task Demand * Format * Trial)	160	377.474				

Computed using alpha = .05

Discussion for Experiment 3

Experiment 3's results replicated those of Experiments 1 and 2. Consistent with the contradictory results found in the previous experiments' hypotheses on warning format, Experiment 3 also found verbal format to be the inferior format of warning presentations. Furthermore, format type did not affect performance on the operational task even when measures that are more sensitive were employed.

As per hypotheses 3 and 4, task demand did have a significant effect on compliance. Task demand at level two and four were not significantly different as predicted, and compliance scores were significantly lower than at level eight. Consistent with our hypotheses, the operational task was not affected by task demand at level two or four, yet inconsistent with what was predicted, Ghost Recon was not affected by task demand at level eight.

The results of this experiment are in line with the hypotheses on task demand, Hypotheses 3 and 4, that when two or four warnings were presented performance on neither the WCCOM task nor the operational task would be affected. However when the warning presentation increased to eight, performance on the WCCOM task degraded, but the operational task performance was not affected. These findings are consistent with Experiment 1 and 2.

It was predicted in Hypotheses 6 and 7 that subjective workload rating scores would be significantly lower when two warnings were presented compared to four warnings. Furthermore, it was predicted that eight warnings would exceed participants' resources and thus, eight warnings would not be associated between all levels of task demand. Results of this study did not find an effect for task demand on subjective workload.

Experiment 3 replicated Experiment 1 in order to validate Experiments 1's findings. Experiment 4 was an exact replica of Experiment 2, thus the written response format was used. Experiment 4 was conducted to validate Experiment 2's findings. In addition, Experiment 4 used a more sensitive measure, the number of kills divided by the number of enemies, vice just the number of kills that were taken in Experiments 1 and 2.

EXPERIMENT 4

Experiment 4 was a replica of Experiment 2, the response format on the WCCOM task was the written response. The Ghost Recon shooting task was also used for this experiment and performance was measured using the more sensitive measure, the number of kills divided by the number of enemies in the building.

Method

Participants

Eleven undergraduate students from the University of Central Florida (9 females and 2 males, mean age = 19.5 years) were recruited on a voluntary basis from the university's experimental recruiting website (www.experimetrix.com/ucf). Participants were paid approximately $37.50 for their participation (based on an hourly rate of $7.50). Participants could only participate in one of the eight studies. Participants were not used in any other experiments in the series of studies so that they would not become familiar with the task or become "expert users".

Materials

WCCOM Compliance Task

The WCCOM compliance task for experiment 4 was consistent with the general materials used for the experimental system. The keyboard in this task had keys "q" through "p" labeled with a written word representing the warning portion of each combination.

Ghost Recon Task

The Ghost Recon shooting task for experiment 4 was consistent with the general materials used for the experimental system.

Procedure

The procedure for Experiment 4 was consistent with the general materials for the experimental setting.

Results for Experiment 4

Experiment 4 is an exact replication of Experiment 2, including the use of written words as a response mode. Again, a larger sample size was used.

WCCOM Task

In order to determine whether the size of the memory set (task-based stress) had a systematic effect on performance across varying formats of warning presentation, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on warning compliance behavior. The independent variables were format, task demand, and trial. The three levels of warning format: verbal, pictorial, and written were presented to the participant at varying levels of task demand: presentations of two, four, and eight warning-color combinations, over forty-five trials: 15 trials per block, five trials per level of task demand. The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination) at each level of task demand.

Analyses were performed using SPSS for Windows, 11.5 and an alpha level of .05 was used for all analysis. A conservative approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick & Fidell, 2001). Thus, all eleven participants were used in all analyses.

There was a significant main effect of format and task demand, Wilk's Λ = .16, $F(2, 20) = 10.6$, $p = .001$, partial $\eta^2 = .52$, and Wilk's Λ = .08, $F(2, 20) = 87.2$, $p = .0005$, partial $\eta^2 = .90$, respectively. No main effect was found for trial and no significant interactions were found.

A set of Fisher LSD post hoc tests for format type showed that participants were significantly more likely to comply when the information was presented in written (*M* = .66, *SD* = .036) than verbal format (*M* = .46, *SD* = .03). No significant differences were found between picture format (*M* = .537, *SD* = .053) and verbal (*M* = .46, *SD* = .03) or picture (*M* = .537, *SD* = .053) and written format (*M* = .66, *SD* = .036).

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two (*M* = .718, *SD* = .028) than at level four (*M* = .532, *SD* = .035) or eight (*M* = .406, *SD* = .04). In addition, results yield a significant difference between level four (*M* = .532, *SD* = .035) and level eight (*M* = .406, *SD* = .04). Thus, as the rate of task demand increased compliance scores decreased.

Table 12
The ANOVA Table for the WCCOM Task for Experiment 4

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	1.681	10.627	.001	.515	.976
Error(format)	20	.158				
Task Demand	2	4.068	87.222	.000	.897	1.000
Error(Task Demand)	20	.047				
Trial	4	.185	1.623	.187	.140	.454
Error(Trial)	40	.114				
Format * Task Demand	4	.012	.115	.977	.011	.071
Error(Format*Task Demand)	40	.104				
Format * Trial	8	.123	1.319	.246	.117	.567
Error(Format*Trial)	80	.093				
Task Demand * Trial	8	.087	1.162	.332	.104	.503
Error(Task Demand*Trial)	80	.075				
Format * Task Demand * Trial	16	.062	.771	.717	.072	.511
Error(Format*Task Demand*Trial)	160	.080				

Computed using alpha = .05

Ghost Recon Shooting Task

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the amount of enemies killed. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15trials. The dependent variable was the percent of enemies killed in each building (enemies in the building divided by enemies killed). A significant main effect for task demand was found, Wilk's $\Lambda = .66$, $F(2, 20) = 3.8$, $p = .04$, partial $\eta^2 = .28$. There were no significant main effects of format, Wilk's $\Lambda = .81$, $F(2, 20) = .38$, $p > .05$, partial $\eta^2 = .10$ or trial, Wilk's $\Lambda = .69$, $F(4, 40) = 1.8$, $p > .05$, partial $\eta^2 = .51$. A significant interactions was found between format and task demand, $F(4, 40) = 3.1$, $p = .03$, partial $\eta^2 = .23$.

Furthermore, Fisher LSD post hoc tests for task demand did not yield any task demand level as significantly different from another level. A trend was emerging between level two (*M* = .96, *SD* = .013) and level eight (*M* = .924, *SD* = .018). However, as can be seen from the means and standard deviations, there are very little differences between the three groups.

Table 13
The ANOVA Table for the Ghost Recon Shooting Task for Experiment 4

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.011	.376	.691	.036	.102
Error(Format)	20	.029				
Task Demand	2	.072	3.844	.039	.278	.628
Error(Task Demand)	20	.019				
Trial	4	.022	1.822	.144	.154	.505
Error(Trial)	40	.012				
Format * Task Demand	4	.050	3.063	.027	.234	.757
Error(Format*Task Demand)	40	.016				
Format * Trial	8	.016	1.303	.254	.115	.561
Error(Format*Trial)	80	.012				
Task Demand * Trial	8	.012	1.153	.338	.103	.499
Error(Task Demand*Trial)	80	.011				
Format * Task Demand * Trial	16	.023	1.392	.152	.122	.832
Error(Format * Task Demand * Trial)	160	.017				

Computed using alpha = .05

NASA-TLX

In order to determine if format of presentation affected subjective mental workload ratings, a one-way, analysis of variance (ANOVA) was conducted. The independent variable, warning format, included three levels: verbal, written and pictorial warnings. The dependent variable was the subjective mental workload effort determined by using the NASA-task load Index (NASA-TLX). The analysis yielded non-significant results, Wilk's Λ = .86, F (2, 18) = .82, p>.05, partial η^2 = .84. Thus, participants did not report feeling more mental demand in one warning format vice another format.

RSME

A three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the Rating Scale Mental Effort (RSME) subjective workload measure. The independent variables were format, task demand, and trial. The dependent variable of interest was the RSME. Results of this analysis yielded a main effect for warning format Wilk's Λ = .10, $F(2,20) = 68.2$, p =.0005, partial η^2 = .87. A main effect was not found for task demand Wilk's Λ = .45, $F(2,20) = 1.9$, p >.05, partial η^2 = .16 or for trial Wilk's Λ = .31, $F(4,40) = 2.3$, p >.05, partial η^2 = .19. No significant interactions were found.

A set of Fisher LSD post hoc tests for format showed that participants subjective workload ratings were the greatest when the warning format was written (*M* = 67.32, *SD* = 3.37) compared pictorial (*M* = 51.93, *SD* = 4.66) or verbal (*M* = 30.95, *SD* = 5.14) Furthermore, significant differences in workload were found between pictorial (*M* = 51.93, *SD* = 4.66) and verbal format (*M* = 30.95, *SD* = 5.14). Therefore, participants felt that they used the most mental effort when warnings were presented in written, followed by pictorials and the least amount of workload in the verbal condition.

Table 14
The ANOVA Table for the RSME for Experiment 4

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Task Demand	2	2258.958	1.882	.178	.158	.344
Error(Task Demand)	20	1200.324				
Format	2	54974.324	68.223	.000	.872	1.000
Error(Format)	20	805.799				
Trial	4	482.804	2.284	.077	.186	.612
Error(Trial)	40	211.422				
Task Demand * Format	4	716.346	2.235	.082	.183	.602
Error(Task Demand * Format)	40	320.446				
Task Demand * Trial	8	188.949	1.050	.406	.095	.456
Error(Task Demand * Trial)	80	179.947				
Format * Trial	8	299.010	1.537	.158	.133	.647
Error(Format * Trial)	80	194.603				
Task Demand * Format * Trial	16	61.433	.387	.984	.037	.246
Error(Task Demand * Format * Trial)	160	158.649				

Computed using alpha = .05

Discussion for Experiment 4

Experiment 4 replicated the compliance results of Experiment 2. As in Experiment 2, verbal warning format yielded the lowest compliance scores. This finding was consistent across all four experiments, when the response format was either pictorial or written; the verbal presentation consistently resulted in the format that yielded the lowest compliance scores. These results may be due to the excess processing that may have occurred from the interference between the verbal presentation format and the pictorial and written responses. Additionally, results of Experiment 4 were also consistent with the previous three experiments in that contrary to what was hypothesized, no significant effect of format was found for the operational task.

Hypotheses 3 and 4 on task demand predicted that when two or four warnings were presented performance on neither the WCCOM task nor the operational task would be affected; yet when the warning presentation increased to eight, performance on both tasks would degrade.

Consistent with these hypotheses, the WCCOM task was affected by the task based stress. As the task demand increased, compliance rate decreased. Additionally, the operational task performance was also affected by task demand.

It was predicted in Hypotheses 6 and 7 that subjective workload rating scores would be significantly lower when two warnings were presented compared to four warnings. Inconsistent with this prediction, participants did not report feeling more mental workload at any level of task demand.

The verbal inferiority finding for the WCCOM task was consistent across all four experiments, when the response format was either pictorial or written; the verbal presentation consistently resulted in the format that yielded the lowest compliance scores. As discussed previously, these results may be due to the excess processing that may have occurred from the interference between the verbal presentation format and the pictorial and written responses. Therefore, Experiment 5 was conducted in order to broaden the response format to verbal response in order to determine if the pairing of presentation and response mode in the same format would result in verbal superiority effect.

EXPERIMENT 5

Experiment 5 was an exact replica of Experiments 3 and 4, but a verbal response mode was used for the WCCOM task.

Method

Participants

Twelve undergraduate students from the University of Central Florida (8 females and 4 males, mean age = 21.2 years) were recruited on a voluntary basis from the university's experimental recruiting website (www.experimetrix.com/ucf). Participants were paid approximately $37.50 for their participation (based on an hourly rate of $7.50). Participants could only participate in one of the eight studies. Participants were not used in any other experiments in the series of studies so that they would not become familiar with the task or become "expert users".

Materials

WCCOM Compliance Task

The WCCOM compliance task for Experiment 5 was consistent with the general materials used for the experimental system. Keyboard responses were not used in this task. Participants were required to respond verbally to the WCCOM color stimuli. Responses were recorded via a microphone component of the WCCOM computer.

Ghost Recon Task

The Ghost Recon shooting task for Experiment 5 was the exact same as Experiments 3 and 4 and the more sensitive measure for performance was used.

Procedure

The procedure for Experiment 5 was consistent with the general materials for the experimental setting.

Results for Experiment 5

Experiment 5 was conducted in order to determine if adding a third response mode, verbal response to the WCCOM task would affect performance across warning format and task load.

WCCOM Task

In order to determine whether the size of the memory set (task-based stress) had a systematic effect on performance across varying formats of warning presentation, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on warning compliance behavior. The independent variables were format, task demand, and trial. The three levels of warning format: verbal, pictorial, and written were presented to the participant at varying levels of task demand: presentations of two, four, and eight warning-color combinations, over forty-five trials: 15 trials per block, five trials per level of task demand. The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination) at each level of task demand.

Analyses were performed using SPSS for Windows, 11.5 and an alpha level of .05 was used for all analysis. A conservative approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick & Fidell, 2001). Thus, all twelve participants were used in all analyses.

There was a significant main effect of task demand and trial, Wilk's Λ = .14, $F(2, 22) = 26.5$, $p = .0005$, partial $\eta^2 = .77$, and Wilk's Λ = .38, $F(4, 44) = 3.3$, $p = .019$, partial $\eta^2 = .79$, respectively. No main effect was found for format Wilk's Λ = .84, $F(2, 22) = .61$, $p > .05$, partial $\eta^2 = .05$. A significant interaction was found for task demand and trial $F(8, 88) = 2.2$, $p = .04$, partial $\eta^2 = .16$.

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two (*M* = .932, *SD* = .019) than at level four (*M* = .763, *SD* = .049) or eight (*M* = .617, *SD* = .049). In addition, results yield a significant difference between level four (*M* = .763, *SD* = .049) and level eight (*M* = .617, *SD* = .049). Thus, as the rate of task demand increased compliance scores decreased.

Additionally, Fisher LSD post hoc tests for trial revealed that at trial one (*M* = .704, *SD* = .048), participants complied significantly less than at trial two (*M* = .811, *SD* = .041) or three (*M* = .806, *SD* = .036). No other significant differences were found for trial.

Table 15
The ANOVA Table for the WCCOM Task for Experiment 5

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.103	.605	.555	.052	.138
Error(format)	22	.170				
Task Demand	2	4.479	36.537	.000	.769	1.000
Error(Task Demand)	22	.123				
Trial	4	.200	3.281	.019	.230	.794
Error(Trial)	44	.061				
Format * Task Demand	4	.041	.725	.579	.062	.215
Error(Format*Task Demand)	44	.057				
Format * Trial	8	.050	.934	.493	.078	.409
Error(Format*Trial)	88	.053				
Task Demand * Trial	8	.094	2.154	.039	.164	.824
Error(Task Demand*Trial)	88	.044				
Format * Task Demand * Trial	16	.046	.961	.501	.080	.635
Error(Format*Task Demand*Trial)	176	.048				

Computed using alpha = .05

Ghost Recon Shooting Task

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the amount of enemies killed. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15trials. The dependent variable was the percent of enemies killed in each building (enemies in the building divided by enemies killed). There were no significant main effects of format, task demand, or trial, Wilk's Λ = .56, $F(2, 22) = .69$, $p > .05$, partial $\eta^2 = .06$, Wilk's Λ = .89, $F(2, 22) = 2.9$, $p > .05$, partial $\eta^2 = .21$, Wilk's Λ = .75, $F(4, 44) = 1.6$, $p > .05$, partial $\eta^2 = .46$, respectively. No significant interactions were found.

NASA-TLX

In order to determine if format of presentation affected subjective mental workload ratings, a one-way, analysis of variance (ANOVA) was conducted. The independent variable, warning format, included three levels: verbal, written and pictorial warnings. The dependent variable was the subjective mental workload effort determined by using the NASA-task load Index (NASA-TLX). The analysis yielded non-significant results, Wilk's Λ = .64, F (2, 22) = 1.1, p>.05, partial η^2 = .09. Thus, participants did not report feeling more mental demand in one warning format vice another format.

RSME

A three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the Rating Scale Mental Effort (RSME) subjective workload measure. The independent variables were format, task demand, and trial. The dependent variable of interest was the RSME. Results of this analysis yielded a main effect for task demand Wilk's Λ = .05, F(2,22) = 122.1, p =.0005, partial η^2 = .92. A main effect was not found for warning format Wilk's Λ = .80, F(2,22) = 1.2, p >.05 partial η^2 = .10 or for trial Wilk's Λ = .67, F(4,44) = .71, p >.05, partial η^2 = .06. No significant interactions were found.

A set of Fisher LSD post hoc tests for task demand showed that participants felt the least amount of mental workload when warnings were presented at level two (M = 23.15, SD = 3.66) than at level four (M = 41.17, SD = 4.24) or eight (M = 61.94, SD = 3.43). Furthermore, significant differences in workload were found between level four (M = 41.17, SD = 4.24) and eight (M = 61.94, SD = 3.43). Therefore, as the task demand increased the subjective workload ratings also increased.

Table 16
The ANOVA Table for the RSME for Experiment 5

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	2496.446	1.200	.320	.098	.235
Error(Format)	22	2080.220				
Task Demand	2	67804.326	122.121	.000	.917	1.000
Error(Task Demand)	22	555.221				
Trial	4	166.904	.713	.588	.061	.212
Error(Trial)	44	234.233				
Task Demand * Format	4	265.624	1.110	.364	.092	.320
Error(Task Demand * Format)	44	239.363				
Format * Trial	8	265.384	1.296	.256	.105	.562
Error(Format * Trial)	88	204.701				
Task Demand * Trial	8	219.057	1.185	.317	.097	.517
Error(Task Demand * Trial)	88	184.850				
Task Demand * Format * Trial	16	205.376	1.578	.079	.125	.890
Error(Task Demand * Format * Trial)	176	130.150				

Computed using alpha = .05

Discussion for Experiment 5

Experiment 5 was the first study in the series of experiments to use verbal responses vice pictorials or written. Interestingly, the results of this experiment on warning presentation format resulted in non-significant differences. Thus, compliance scores did not differ when the response was in verbal format across the three warning presentation formats, pictorial, written, or verbal. These results contradict not only the hypotheses, but also the results of the previous experiments. Like the other experiments in this line of research performance on the operational task was not affected by format or task demand.

Although format was not significant, task demand was significant. For the hypotheses on task demand, Hypotheses 3 and 4, it was predicted that when two or four warnings were presented performance on neither the WCCOM task nor the operational task would be affected; yet when the warning presentation increased to eight, performance on both tasks would degrade. Consistent with these hypotheses, the WCCOM task was affected by the task based stress at all levels of demand. Contrary to what was predicted, the operational task was not affected by task demand. These findings are consistent with Experiments 1-4.

It was predicted in Hypotheses 6 and 7 that subjective workload rating scores would be significantly lower when two warnings were presented compared to four warnings. As predicted, participants felt less workload at level two than at level four or eight. Furthermore, it was predicted that eight warnings would exceed participants' resources and thus, eight warnings would not be associated between all levels of task demand. Results of this study revealed that participants had the highest rate of mental workload at level eight.

Experiments 1-5 have identified the effects of format presentation at each of the three levels (pictorial, written, verbal). Furthermore, this line of experimentation has also investigated the effects of the three levels of response mode (pictorial, written, verbal). Results of these studies suggest that it was not only the presentation format of the warning that affects compliance, but also the response format. Although compliance on the WCCOM was affected by this combination, the operational task went unscathed. In the case that the operational task was not complex enough to produce shifts in performance, a more difficult operational task was created, a navigation task. Therefore, Experiment 6 will replicate experiment three utilizing the pictorial response format. The only variation to Experiment 6 will be the operational task.

EXPERIMENT 6

Experiment 6 through 8 replicated the response modes for the WCCOM tasks of Experiments 3-5 (respectively). The Operational task in Experiments 6-8 changed to the more difficult and challenging navigation task. The navigation task was still a visual/spatial task, but the working memory component of the task was more complex. Our intent here was to examine whether the WCCOM warning presentation format/warning response results from Experiments 1-5 would replicate using a more mentally and spatially demanding operational task. In Experiments 1-5 performance on Operational task was not affected by warning format or task demand. It was predicted that if the Ghost Recon shooting task were replaced by the more demanding navigation task, it would result in performance degradations. Specifically, it was hypothesized that interference would occur when the warnings were presented in the written and pictorial due to the interference of two visual/spatial task processing. Furthermore, since the navigation task was more mentally and spatially demanding the effects of increased task demand would deplete available resources and as a result, there would be diminishing effects on performance.

Method

Participants

Twelve undergraduate students from the University of Central Florida (8 females and 4 males, mean age = 20.1 years) were recruited on a voluntary basis from the university's experimental recruiting website (www.experimetrix.com/ucf). Participants were paid approximately $37.50 for their participation (based on an hourly rate of $7.50). Participants could only participate in one of the eight studies. Participants were not used in any other experiments in the series of studies so that they would not become familiar with the task or become "expert users".

Materials

WCCOM Compliance Task

The WCCOM compliance task for Experiment 6 was consistent with the general materials used for the experimental system. The keyboard in this task had keys "q" through "p" labeled with a pictorial representation of the warning portion of each combination.

Ghost Recon Task

The Ghost Recon navigation task was used for Experiment 6 and was consistent with the general materials used for the experimental system.

Procedure

The procedure for Experiment 6 was consistent with the general materials for the experimental setting.

Results for Experiment 6

Experiment 6 through 8 replicated the response modes for the WCCOM tasks of experiments 3-5 (respectively). Thus, in Experiment 6 the WCCOM response was a pictorial representation of the warning as it was in Experiments 1 and 3. The Operational task in Experiments 6-8 changed to a more difficult task, the navigation task.

WCCOM Task

In order to determine whether the size of the memory set (task-based stress) had a systematic effect on performance across varying formats of warning presentation, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on warning compliance behavior. The independent variables were format, task demand, and trial. The three levels of warning format: verbal, pictorial, and written were presented to the participant at varying levels of task demand: presentations of two, four, and eight warning-color combinations, over forty-five trials: 15 trials per block, five trials per level of task demand. The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination) at each level of task demand.

Analyses were performed using SPSS for Windows, 11.5 and an alpha level of .05 was used for all analysis. A conservative approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick & Fidell, 2001). Thus, all twelve participants were used in all analyses.

There was a significant main effect of format and task demand, Wilk's $\Lambda = .14$, $F(2, 22) = 42.8$, $p = .0005$, partial $\eta^2 = .80$, and Wilk's $\Lambda = .078$, $F(2, 22) = 94.7$, $p = .0005$, partial $\eta^2 = .90$, respectively. No main effect was found for trial, Wilk's $\Lambda = .37$, $F(4, 44) = 1.7$, $p > .05$. The analysis yielded a significant interaction between format and task demand, $F(4, 44) = 5.2$, $p = .002$, partial $\eta^2 = .32$.

A set of Fisher LSD post hoc tests for format type showed that participants were significantly more likely to comply when the information was presented in written ($M = .816$, $SD = .023$) and pictorial format ($M = .791$, $SD = .032$) than compared to verbal format ($M = .58$, *SD* = .032). No significant differences were found between written format (*M* = .816, *SD* = .023) and pictorials (*M* = .791, *SD* = .032).

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two (*M* = .822, *SD* = .018) and four (*M* = .819, *SD* = .031) than at level eight (*M* = .549, *SD* = .032), but no significant differences in compliance were found between levels two (*M* = .822, *SD* = .018) and four (*M* = .819, *SD* = .031).

Finally, Duncan post hoc analyses were conducted in order to determine where the differences lie in the interaction between format and task demand, $F(4, 44) = 5.2$, $p = .002$, partial $\eta^2 = .32$. Results of these analyses suggest that participants were more likely to comply when warnings were presented in written format at level two (*M* = .975, *SD* = .018) and four (*M* = .867, *SD* = .03) as compared to written warnings at level eight (*M* = .606, *SD* = .047). Significant differences were not found between written warnings at level two (*M* = .975, *SD* = .018) and four (*M* = .867, *SD* = .03).

Similarly, participants were more likely to comply when warnings were presented in verbal format at level two (*M* = .642, *SD* = .023) and four (*M* = .733, *SD* = .056) as compared to verbal warnings at level eight (*M* = .375, *SD* = .036). Significant differences were not found between verbal warnings at level two (*M* = .642, *SD* = .023) and four (*M* = .733, *SD* = .056).

Results of the Duncan post hoc analyses also revealed a significant differences between warnings presented in pictorial format at level four (*M* = .858, *SD* = .033) and warnings presented in pictorial format at level eight (*M* = .665, *SD* = .045).

Looking across warning format, post hoc analyses yield warnings presented in verbal format at level two (*M* = .642, *SD* = .023) had significantly lower compliance scores then warnings presented in pictorial (*M* = .85, *SD* = .034) or written (*M* = .975, *SD* = .018) format at the same level.

Likewise, post hoc analyses also revealed that warnings presented in verbal format at level eight (*M* = .375, *SD* = .036) had significantly lower compliance scores then warnings presented in pictorial (*M* = .665, *SD* = .045) or written (*M* = .606, *SD* = .047) format at the same level.

Table 17
The ANOVA Table for the WCCOM Task for Experiment 6

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	2.936	42.802	.000	.796	1.000
Error(format)	22	.069				
Task Demand	2	4.447	94.692	.000	.896	1.000
Error(Task Demand)	22	.047				
Trial	4	.144	1.701	.167	.134	.479
Error(Trial)	44	.085				
Format * Task Demand	4	.254	5.246	.002	.323	.953
Error(Format*Task Demand)	44	.048				
Format * Trial	8	.039	.605	.771	.052	.264
Error(Format*Trial)	88	.065				
Task Demand * Trial	8	.024	.333	.951	.029	.153
Error(Task Demand*Trial)	88	.072				
Format * Task Demand * Trial	16	.051	.827	.653	.070	.552
Error(Format*Task Demand*Trial)	176	.062				

Computed using alpha = .05

Ghost Recon Navigation Task

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way

3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the amount of enemies killed. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15 trials. The dependent variable was the time (in %) it took participants to reach all four waypoints ((actual time-goal time)/goal time X 100).

There was a significant main effect of task demand, Wilk's $\Lambda = .26$, $F(2, 22) = 9.1$, $p = .001$, partial $\eta^2 = .45$. No main effect was found for format or trial, Wilk's $\Lambda = .98$, $F(2, 22) = .74$, $p > .05$, partial $\eta^2 = .01$, Wilk's $\Lambda = .35$, $F(4, 44) = 1.9$, $p > .05$, partial $\eta^2 = .15$, respectively. No significant interactions were found.

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at level two ($M = .259$, $SD = .071$) than at level eight ($M = .507$, $SD = .097$), but no other significant differences in compliance were found.

Table 18
The ANOVA Table for the Ghost Recon Navigation Task for Experiment 6

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.073	.074	.929	.007	.060
Error(Format)	22	.986				
Task Demand	2	2.776	9.092	.001	.453	.954
Error(Task Demand)	22	.305				
Trial	4	.592	1.883	.130	.146	.525
Error(Trial)	44	.314				
Format * Task Demand	4	.497	1.362	.263	.110	.389
Error(Format*Task Demand)	44	.365				
Format * Trial	8	.244	.675	.712	.058	.294
Error(Format*Trial)	88	.361				
Task Demand * Trial	8	.227	.545	.820	.047	.238
Error(Task Demand*Trial)	88	.418				
Format * Task Demand * Trial	16	.304	.635	.852	.055	.420
Error(Format * Task Demand * Trial)	176	.479				

Computed using alpha = .05

NASA-TLX

In order to determine if format of presentation affected subjective mental workload ratings, a one-way, analysis of variance (ANOVA) was conducted. The independent variable, warning format, included three levels: verbal, written and pictorial warnings. The dependent variable was the subjective mental workload effort determined by using the NASA-task load Index (NASA-TLX). The analysis yielded non-significant results, Wilk's Λ = .73, F (2, 20) = 2.9, p>.05, partial η^2 = .22. Thus, participants did not report feeling more mental demand in one warning format vice another format.

RSME

A three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the Rating Scale Mental Effort (RSME) subjective workload measure. The independent variables were format, task demand, and trial. The dependent variable of interest was the RSME. Results of this analysis yielded a main effect for task demand Wilk's Λ = .05, F(2,22) = 156.2, p =.0005, partial η^2 = .93. A main effect was not found for warning format Wilk's Λ = .80, F(2,22) = .31, p >.05, partial η^2 = .03 or for trial, Wilk's Λ = .67, F(4,44) = .69 p >.05, partial η^2 = .06. A significant three-way interactions was found between format, task demand, and trial, F(16,176) = 2.1, p =.012, partial η^2 = .97.

A set of Fisher LSD post hoc tests for task demand showed that participants had the least amount of mental workload at level two (M = 26.15, SD = 3.19) than at level four (M = 42.75, SD = 2.05) or eight (M = 69.51, SD = 3.56). Furthermore, significant differences in workload were found between level four (M = 42.75, SD = 2.05) and eight (M = 69.51, SD = 3.56). Therefore, as task demand increased, subjective workload ratings also increased.

Table 19
The ANOVA Table for the RSME for Experiment 6

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	240.557	.306	.739	.027	.093
Error(Format)	22	785.885				
Task Demand	2	86160.480	156.213	.000	.934	1.000
Error(Task Demand)	22	551.558				
Trial	4	188.951	.685	.606	.059	.205
Error(Trial)	44	275.758				
Task Demand * Format	4	222.882	1.081	.378	.089	.312
Error(Task Demand * Format)	44	206.188				
Format * Trial	8	233.548	1.059	.399	.088	.464
Error(Format * Trial)	88	220.499				
Task Demand * Trial	8	147.144	1.249	.281	.102	.543
Error(Task Demand * Trial)	88	117.847				
Task Demand * Format * Trial	16	253.731	2.059	.012	.158	.966
Error(Task Demand * Format * Trial)	176	123.214				

a Computed using alpha = .05

Discussion for Experiment 6

Experiment 6 replicated the response mode for the WCCOM tasks of experiments 1 and 3 (pictorial). The operational task in Experiments 6 was altered to a more difficult task, a navigation task. The navigation task was still a visual/spatial task, but the working memory component of the task was more complex.

Hypothesis 1 on warning format predicted that participants would have significantly higher rate of compliance behavior when warnings were presented in verbal compared to written and pictorial. Results of Experiment 6 concluded that contrary to aforementioned hypotheses, pictorial and written warnings resulted in the superior format of warning presentation as compared to verbal warnings. These results replicate those of Experiments 1 and 3.

Additionally, Hypothesis 2 also predicted verbal warning would be the superior mode of warning presentation because warnings presented verbally would have less interference on the operational task, Ghost Recon shooting task. Even though the Ghost Recon shooting task used in Experiments 1-5 was replaced by a more complex navigation task, performance was still not affected by warning format.

For the hypotheses on task demand, Hypotheses 3 and 4, it was predicted that when two or four warnings were presented performance on neither the WCCOM task nor the operational task would be affected; yet when the warning presentation increased to eight, performance on both tasks would degrade. Consistent with these hypotheses, the WCCOM task was affected by the task based stress. When only two or four warnings were presented, no differences were found, yet a significant decrease in compliance resulted when eight warnings were presented. Operational task performance was affected by task demand in the same in the same manner the WCCOM was affected.

It was predicted in Hypotheses 6 and 7 that subjective workload rating scores would be significantly lower when two warnings were presented compared to four warnings. As predicted, participants felt less workload at level two than at level four. Furthermore, it was predicted that eight warnings would exceed participants' resources and thus, eight warnings would not be associated between all levels of task demand. Results of this study did reveal that participants had the highest rate of mental workload at level eight.

The next Experiment, Experiment 7 was conducted in order to carry out the experimentation using the Ghost Recon navigation task, yet with a different response format, the written warning response. Experiment 7 was an exact replication of Experiment 4, yet with the replacement of the navigation task vice the shooting task.

EXPERIMENT 7

Method

Participants

Twelve undergraduate students from the University of Central Florida (5 females and 7 males, mean age = 19.5 years) were recruited on a voluntary basis from the university's experimental recruiting website (www.experimetrix.com/ucf). Participants were paid approximately $37.50 for their participation (based on an hourly rate of $7.50). Participants could only participate in one of the eight studies. Participants were not used in any other experiments in the series of studies so that they would not become familiar with the task or become "expert users".

Materials

WCCOM Compliance Task

The WCCOM compliance task for Experiment 7 was consistent with the general materials used for the experimental system. The keyboard in this task had keys "q" through "p" labeled with a written word representing the warning portion of each combination.

Ghost Recon Task

The Ghost Recon navigation task was used for Experiment 7 and was consistent with the general materials used for the experimental system.

Procedure

The procedure for experiment 7 was consistent with the general materials for the experimental setting.

Results for Experiment 7

WCCOM Task

In order to determine whether the size of the memory set (task-based stress) had a systematic effect on performance across varying formats of warning presentation, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on warning compliance behavior. The independent variables were format, task demand, and trial. The three levels of warning format: verbal, pictorial, and written were presented to the participant at varying levels of task demand: presentations of two, four, and eight warning-color combinations, over forty-five trials: 15 trials per block, five trials per level of task demand. The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination) at each level of task demand.

Analyses were performed using SPSS for Windows, 11.5 and an alpha level of .05 was used for all analysis. A conservative approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick & Fidell, 2001). Thus, all twelve participants were used in all analyses.

There was a significant main effect of format and task demand, Wilk's Λ = .15, $F(2, 22) = 33.3$, $p = .0005$, partial $\eta^2 = .76$, and Wilk's Λ = .13, $F(2, 22) = 35.4$, $p = .0005$, partial $\eta^2 =$ .76, respectively. No main effect was found for trial, Wilk's Λ = .37, $F(4, 44) = .71$, $p > .05$, partial η^2 = .06,. The analysis yielded a significant interaction between format and task demand, $F(4, 44) = 3$, $p = .03$, partial η^2 = .21.

A set of Fisher LSD post hoc tests for format type showed that participants were significantly more likely to comply when the information was presented in written (*M* = .708, *SD* = .04) and pictorial format (*M* = .692, *SD* = .031) than compared to verbal format (*M* = .471, *SD* = .038). No significant differences were found between written format (*M* = .708, *SD* = .04) and pictorials (*M* = .692, *SD* = .031).

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two (*M* = .778, *SD* = .029) than at level four (*M* = .65, *SD* = .053) or eight (*M* = .443, *SD* = .029). In addition, results yield a significant difference between level four (*M* = .65, *SD* = .053) and level eight (*M* = .443, *SD* = .029). Thus, as the rate of task demand increased compliance scores decreased.

Finally, Duncan post hoc analyses were conducted in order to determine where the differences lie in the interaction between format and task demand, $F(4, 44) = 3$, $p = .03$, partial η^2 = .21. Results of these analyses suggest that participants were more likely to comply when warnings were presented in written format at level two (*M* = .925, *SD* = .035) than at level four (*M* = .687, *SD* = .065) or level eight (*M* = .51, *SD* = .049). Significant differences were not found between written warnings at level two (*M* = .925, *SD* = .035) and four (*M* = .687, *SD* = .065).

Additionally, participants were more likely to comply when warnings were presented in verbal format at level two (*M* = .575, *SD* = .033) and four (*M* = .533, *SD* = .075) as compared to verbal warnings at level eight (*M* = .304, *SD* = .033). Significant differences were not found between verbal warnings at level two (*M* = .575, *SD* = .033) and four (*M* = .533, *SD* = .075).

Results of the Duncan post hoc analyses also revealed a significant differences between warnings presented in pictorial format at level two (*M* = .833, *SD* = .036) and four (*M* = .729, *SD* = .049) as compared to warnings presented in pictorial format at level eight (*M* = .515, *SD* = .034). Significant differences were not found between pictorial warnings at level two (*M* = .833, *SD* = .036) and four (*M* = .729, *SD* = .049).

Looking across warning format, post hoc analyses yield warnings presented in verbal format at level two (*M* = .575, *SD* = .033) had significantly lower compliance scores then warnings presented in pictorial (*M* = .833, *SD* = .036) or written (*M* = .925, *SD* = .035) format at the same level.

Likewise, post hoc analyses also revealed that warnings presented in verbal format at level eight (*M* = .304, *SD* = .033) had significantly lower compliance scores then warnings presented in written (*M* = .51, *SD* = .049) format at the same level.

Table 20
The ANOVA Table for the WCCOM Task for Experiment 7

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	3.162	33.289	.000	.752	1.000
Error(format)	22	.095				
Task Demand	2	5.136	35.408	.000	.763	1.000
Error(Task Demand)	22	.145				
Trial	4	.055	.711	.589	.061	.212
Error(Trial)	44	.078				
Format * Task Demand	4	.161	2.982	.029	.213	.750
Error(Format*Task Demand)	44	.054				
Format * Trial	8	.069	.687	.702	.059	.299
Error(Format*Trial)	88	.100				
Task Demand * Trial	8	.119	1.384	.215	.112	.596
Error(Task Demand*Trial)	88	.086				
Format * Task Demand * Trial	16	.074	1.033	.425	.086	.676
Error(Format*Task Demand*Trial)	176	.072				

a Computed using alpha = .05

Ghost Recon Navigation Task

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the amount of enemies killed. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15trials. The dependent variable was the time (in %) it took participants to reach all four waypoints ((actual time-goal time)/goal time X 100).

There were no significant main effects of format, task demand, or trial Wilk's $\Lambda = .95$, $F(2, 22) = .39$, $p > .05$, partial $\eta^2 = .03$, Wilk's $\Lambda = 1.4$, $F(2, 22) = .74$, $p > .05$, partial $\eta^2 = .26$, and, Wilk's $\Lambda = .70$, $F(4, 44) = .53$, $p > .05$, partial $\eta^2 = .16$, respectively. No significant interactions were found.

NASA-TLX

In order to determine if format of presentation affected subjective mental workload ratings, a one-way, analysis of variance (ANOVA) was conducted. The independent variable, warning format, included three levels: verbal, written and pictorial warnings. The dependent variable was the subjective mental workload effort determined by using the NASA-task load Index (NASA-TLX). The analysis yielded significant results, Wilk's $\Lambda = .28$, F (2, 22) = 6.7, p =.003, partial $\eta^2 = .41$. A set of Fisher LSD post hoc tests for the NASA-TLX ratings across warning format showed that participants felt a higher degree of mental workload in the pictorial warning format ($M = 62.72$, $SD = 2.77$) as compared to the verbal ($M = 50.12$, $SD = 5.65$) or written format ($M = 50.95$, $SD = 3.02$).

Table 21
The ANOVA Table for NASA-TLX Scores for Experiment 7

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	596.211	7.656	.003	.410	.915
Error(NASA)	22	77.875				

a Computed using alpha = .05

RSME

A three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the Rating Scale Mental Effort (RSME) subjective workload measure. The independent variables were format, task demand, and trial. The dependent variable of interest was the RSME. Results of this analysis yielded a main effect for task demand Wilk's $\Lambda = .23$, $F(2,22) = 28.7$, $p = .0005$, partial $\eta^2 = .72$. A main effect was not found for warning format Wilk's $\Lambda = .65$, $F(2,22) = .30$, $p > .05$, partial $\eta^2 = .22$ or for trial, Wilk's $\Lambda = .70$, $F(4,44) = .92$ $p > .05$, partial $\eta^2 = .08$. No significant interactions were found.

A set of Fisher LSD post hoc tests for task demand showed that participants mental workload was the least at level two ($M = 23.18$, $SD = 4.03$) compared to level four ($M = 39.97$, $SD = 5.18$) or eight ($M = 61.52$, $SD = 6.54$). Furthermore, significant differences in workload were found between level four ($M = 39.97$, $SD = 5.18$) and eight ($M = 61.52$, $SD = 6.54$). Therefore, as task demand increased, subjective workload increased.

Table 22
The ANOVA Table for the RSME for Experiment 7

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	2529.306	3.012	.070	.215	.525
Error(Format)	22	839.851				
Task Demand	2	66504.156	28.668	.000	.723	1.000
Error(Task Demand)	22	2319.840				
Trial	4	144.804	.918	.462	.077	.267
Error(Trial)	44	157.660				
Task Demand * Format	4	138.661	.263	.900	.023	.102
Error(Task Demand * Format)	44	526.901				
Format * Trial	8	88.030	.553	.813	.048	.241
Error(Format * Trial)	88	159.058				
Task Demand * Trial	8	177.519	1.237	.288	.101	.538
Error(Task Demand * Trial)	88	143.557				
Task Demand * Format * Trial	16	98.139	.673	.817	.058	.447
Error(Task Demand * Format * Trial)	176	145.717				

a Computed using alpha = .05

Discussion for Experiment 7

Experiment 7 replicated the response mode for the WCCOM tasks of experiments 2 and 4 (written). The Operational task task in Experiments 7 was altered to a more difficult task, a navigation task. The navigation task was still a visual/spatial task, but the working memory component of the task was more complex.

Results of Experiment 7 concluded that contrary to hypotheses 1 and 2 on warning format which predicted verbal format as the superior format in which to present warnings. Subsequently, pictorial and written warnings resulted in the superior format of warning presentation as compared to verbal warnings. These results replicate those of Experiments 2 and 4.

Furthermore, performance on the operational task was not affected by warning presentation format. Thus, inconsistent with hypothesis 2, interference did not occur on any of the formats of warning presentation on the operational task. Even though the operational task used in Experiments 1-5 was replaced by a more complex operational task, performance was still not affected by task demand or warning format.

Task demand did affect compliance as was predicted by hypothesis 3. As task demand increased, compliance scores decreased. Yet, contrary to hypothesis 4, performance on the operational task was not affected.

It was predicted in Hypotheses 6 and 7 that subjective workload rating scores would be significantly lower when two warnings were presented compared to four warnings. As predicted, participants felt less workload at level two than at level four. Furthermore, it was predicted that eight warnings would exceed participants' resources and thus, eight warnings would not be associated between all levels of task demand. Results of this study did reveal that participants had the highest rate of mental workload at level eight, but the workload and score were associated.

Experiments 6 and 7 investigated the effects of pictorial or written response format as well as the effects of a more difficult military simulated task. Experiment 8 was conducted to determine if verbal response format differed from the effects of the pictorial or written formats. In addition, it was also of interest to see if the verbal response interfered with the operational task.

EXPERIMENT 8

Method

Participants

Twelve undergraduate students from the University of Central Florida (6 females and 6 males, mean age = 22.3 years) were recruited on a voluntary basis from the university's experimental recruiting website (www.experimetrix.com/ucf). Participants were paid approximately $37.50 for their participation (based on an hourly rate of $7.50). Participants could only participate in one of the eight studies. Participants were not used in any other experiments in the series of studies so that they would not become familiar with the task or become "expert users".

Materials

WCCOM Compliance Task

The WCCOM compliance task for Experiment 8 was consistent with the general materials used for the experimental system. Keyboard responses were not used in this task. Participants were required to respond verbally to the WCCOM color stimuli. Responses were recorded via a microphone component of the WCCOM computer.

Ghost Recon Task

The Ghost Recon navigation task was used for Experiment 8 and was consistent with the general materials used for the experimental system.

Procedure

The procedure for Experiment 8 was consistent with the general materials for the experimental setting.

Results for Experiment 8

WCCOM Task

In order to determine whether the size of the memory set (task-based stress) had a systematic effect on performance across varying formats of warning presentation, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on warning compliance behavior. The independent variables were format, task demand, and trial. The three levels of warning format: verbal, pictorial, and written were presented to the participant at varying levels of task demand: presentations of two, four, and eight warning-color combinations, over forty-five trials: 15 trials per block, five trials per level of task demand. The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination) at each level of task demand.

Analyses were performed using SPSS for Windows, 11.5 and an alpha level of .05 was used for all analysis. A conservative approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick & Fidell, 2001). Thus, all twelve participants were used in all analyses.

There was a significant main effect of task demand, Wilk's Λ = .14, $F(2, 22) = 33.8$, p = .0005, partial $\eta^2 = .76$. No main effect was found for format or trial, Wilk's Λ = .71, $F(2, 22)$ =

1.3, $p > .05$, partial $\eta^2 = .11$, Wilk's $\Lambda = .77$, $F(4, 44) = .67$, $p > .05$, partial $\eta^2 = .06$, respectively. No significant interactions were found.

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two (*M* = .975, *SD* = .013) than at level four (*M* = .791, *SD* = .058) or eight (*M* = .615, *SD* = .058). In addition, results yield a significant difference between level four (*M* = .791, *SD* = .058) and level eight (*M* = .615, *SD* = .058). Thus, as the rate of task demand increased compliance scores decreased.

Table 23
The ANOVA Table for the WCCOM Task for Experiment 8

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.145	1.287	.296	.105	.249
Error(Format)	22	.113				
Task Demand	2	5.820	33.830	.000	.755	1.000
Error(Task Demand)	22	.172				
Trial	4	.033	.672	.615	.058	.201
Error(Trial)	44	.049				
Format * Task Demand	4	.046	.828	.515	.070	.243
Error(Format*Task Demand)	44	.055				
Format * Trial	8	.051	.855	.558	.072	.374
Error(Format*Trial)	88	.060				
Task Demand * Trial	8	.077	1.859	.077	.145	.752
Error(Task Demand*Trial)	88	.042				
Format * Task Demand * Trial	16	.030	.774	.714	.066	.516
Error(Format * Task Demand * Trial)	176	.039				

a Computed using alpha = .05

Ghost Recon Navigation Task

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the amount of enemies killed. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15trials. The dependent variable was the time (in %) it took participants to reach all four waypoints ((actual time-goal time)/goal time X 100).

There was a significant main effect of task demand, Wilk's $\Lambda = .35$, $F(2, 22) = 5.9$, $p = .009$, partial $\eta^2 = .35$. No main effect was found for format or trial, Wilk's $\Lambda = .97$, $F(2, 22) = .19$, $p > .05$, partial $\eta^2 = .02$, Wilk's $\Lambda = .56$, $F(4, 44) = .50$, $p > .05$, partial $\eta^2 = .16$, respectively. No significant interactions were found.

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at level two ($M = .307$, $SD = .073$) than at level eight ($M = .539$, $SD = .094$). Additionally, participants were significantly more likely to comply at level four ($M = .407$, $SD = .095$) than at level eight ($M = .539$, $SD = .094$). No significant differences were found between levels two ($M = .307$, $SD = .073$) and four ($M = .407$, $SD = .095$).

Table 24
The ANOVA Table for the Ghost Recon Navigation Task for Experiment 8

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.112	.187	.831	.017	.075
Error(Format)	22	.601				
Task Demand	2	2.441	5.854	.009	.347	.823
Error(Task Demand)	22	.417				
Trial	4	.215	.498	.737	.043	.157
Error(Trial)	44	.431				
Format * Task Demand	4	.836	1.656	.177	.131	.467
Error(Format*Task Demand)	44	.504				
Format * Trial	8	.379	.875	.541	.074	.383
Error(Format*Trial)	88	.434				
Task Demand * Trial	8	.522	1.117	.360	.092	.488
Error(Task Demand*Trial)	88	.468				
Format * Task Demand * Trial	16	.419	1.038	.419	.086	.679
Error(Format * Task Demand * Trial)	176	.404				

a Computed using alpha = .05

NASA-TLX

In order to determine if format of presentation affected subjective mental workload ratings, a one-way, analysis of variance (ANOVA) was conducted. The independent variable, warning format, included three levels: verbal, written and pictorial warnings. The dependent variable was the subjective mental workload effort determined by using the NASA-task load Index (NASA-TLX). The analysis yielded non-significant results, Wilk's Λ = .85, F(2, 22) = .67, p>.05, partial η^2 = .06. Thus, participants did not report feeling more mental demand in one warning format vice another format.

RSME

A three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the Rating Scale Mental Effort

(RSME) subjective workload measure. The independent variables were format, task demand, and trial. The dependent variable of interest was the RSME. Results of this analysis yielded a main effect for task demand, Wilk's Λ = .24, $F(2,22) = 32.8$, $p = .0005$, partial $\eta^2 = .75$. A main effect was not found for format or trail, Wilk's Λ = .87, $F(2,22) = .82$, $p > .05$, partial $\eta^2 = .17$, Wilk's Λ = .58, $F(4,44) = 1.8$, $p > .05$, partial $\eta^2 = .50$, respectively. An interaction was found between warning format and trial, $F(8,88) = 3.3$, $p = .003$, partial $\eta^2 = .23$.

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants felt the least amount of mental workload at level of two (*M* = 19.9, *SD* = 2.83) compared to level four (*M* = 36.79, *SD* = 3.72) or eight (*M* = 58.67, *SD* = 6.26). In addition, results yield a significant difference between level four (*M* = 36.79, *SD* = 3.72) and level eight (*M* = 58.67, *SD* = 6.26). Thus, as the rate of task demand increased workload increased.

Table 25
The ANOVA Table for the RSME for Experiment 8

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	569.311	.818	.454	.069	.172
Error(format)	22	695.675				
Task Demand	2	68007.255	32.799	.000	.749	1.000
Error(Task Demand)	22	2073.461				
Trial	4	560.035	1.768	.152	.138	.496
Error(Trial)	44	316.702				
Format * Task Demand	4	264.623	1.271	.296	.104	.364
Error(Format*Task Demand)	44	208.255				
Format * Trial	8	563.219	3.273	.003	.229	.961
Error(Format*Trial)	88	172.100				
Task Demand * Trial	8	239.475	1.310	.249	.106	.567
Error(Task Demand*Trial)	88	182.823				
Format * Task Demand * Trial	16	178.199	1.308	.197	.106	.805
Error(Format*Task Demand*Trial)	176	136.233				

a Computed using alpha = .05

Discussion for Experiment 8

Experiment 8 replicated the response mode for the WCCOM tasks of experiments 5 (verbal). This experiment was the final experiment in the series of experiments utilizing the Ghost Recon navigation task.

Results of Experiment 8 replicate the results found in the previous verbal response study, Experiment 5. Contrary to hypotheses 1 and 2 on warning format, verbal warning was not the superior warning format. In fact, warning format did not significantly affect behavioral compliance on the WCCOM task. Although these results were not inline with the hypotheses, the results are consistent based on presentation/response format for the varying manipulation of multiple experiments. Consistent with hypothesis 2, operational task performance was affected by task demand. Task demand at level eight decreased performance compared to demand at levels two and four.

Task demand did affect compliance as was predicted by hypothesis 3. As task demand increased, compliance scores decreased. Yet, contrary to hypothesis 4, performance on the operational task was not affected.

It was predicted in Hypotheses 6 and 7 that subjective workload rating scores would be significantly lower when two warnings were presented compared to four warnings. As task demand increased, subjective workload ratings increased. Furthermore, it was predicted that eight warnings would exceed participants' resources and thus, eight warnings would not be associated between all levels of task demand. Results of this study did reveal that participants had the highest rate of mental workload at level eight.

COLLAPSED DATA

The data was collapsed by response mode, thus the data from Experiments 3 and 6 were combined, Experiments 4 and 7 were combined, and Experiments 5 and 8 were combined. Because a third response mode for verbal response was not conducted in the initial set of experiments, Experiments 1 and 2 were excluded from the analyses so that the number of participants would be even across experiments.

Results for Collapsed Data

WCCOM Task

In order to determine whether task-based stress (size of memory set) had a systematic effect on compliance across varying formats of warning presentation and response format, a four-way 3 (response format) X 3 (presentation format) X 3 (task demand) x 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted for the initial analysis. Trial was not a significant contributing factor in the initial analysis and was not used in the final data analysis. Although trial was used as variable in Experiments 1-8, in only two of the eight experiments it yielded significant effects. Furthermore, the focus of this analysis was not to determine if a learning curve emerged, thus it was decided that trial was not a necessary factor for the final collapsed data analysis.

Table 26
Significant Main Effects for experiments 1-8

Experiment Number	ME for format	ME for Task Demand	ME for Trial
1	*	*	X
2	*	*	*
3	*	*	X
4	*	*	X
5	X	*	*
6	*	*	X
7	*	*	X
8	X	*	X

* Significant main effects
X Non-significant effects

Thus, a three-way 3 (response format) X 3 (presentation format) X 3 (task demand) within-participants repeated measure analysis of variance (ANOVA) was conducted for the final analysis. The independent variables were response format, presentation format, and task demand. The three levels of response format were verbal, pictorial, and written. The three levels of presentation format were also: verbal, pictorial, and written. Task demand was presented at varying levels of two, four, and eight warning-color combinations. The dependent variable was the percent correct (number of times the participant correctly recalled the warning-color combination).

Analyses were performed using SPSS for Windows, 11.5 and an alpha level of .05 was used for all analysis. A conservative approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick & Fidell, 2001). Thus, all twenty-three participants were used in all analyses.

There was a significant main effect of response format, presentation format, and task demand, Wilk's Λ = .42, F(2, 44) = 15.1, p = .0005, partial η^2 = .41, Wilk's Λ = .20, F(2, 44) = 43.9, p = .0005, partial η^2 = .67, and Wilk's Λ = .04, F(2, 44) = 311.5, p = .0005, partial η^2 = .93, respectively. The analysis yielded a significant interaction between response and presentation, response and task demand, and presentation and task demand, F(4, 88) = 11.7, p = .0005, partial η^2 = .35, F(4, 88) = 5.2, p = .001, partial η^2 = .19, and F(4, 88) = 3.3, p = .01, partial η^2 = .13, respectively. A three-way interaction was also found between response, presentation, and task demand, F(8, 176) = 2.1, p = .04, partial η^2 = .09.

A set of Fisher LSD post hoc tests for response format showed that participants were significantly more likely to comply when the response format was verbal (*M* = .789, *SD* = .028) and pictorial (*M* = .679, *SD* = .026) than compared to written response format (*M* = .589, *SD* = .023). Additionally, significant differences were found between verbal (*M* = .789, *SD* = .028) and pictorial response format (*M* = .679, *SD* = .026). Thus, verbal response format resulted in the superior response mode when the experimental data was collapsed.

A set of Fisher LSD post hoc tests for presentation format showed that participants were significantly more likely to comply when the presentation format was written (*M* = .745, *SD* = .017) or pictorial (*M* = .716, *SD* = .023) as compared to verbal presentation format (*M* = .595, *SD* = .012). No significant differences were found between written (*M* = .745, *SD* = .017) and pictorial response format (*M* = .716, *SD* = .023). Thus, both written and pictorial presentation format were superior to verbal warning presentations when experimental data was collapsed.

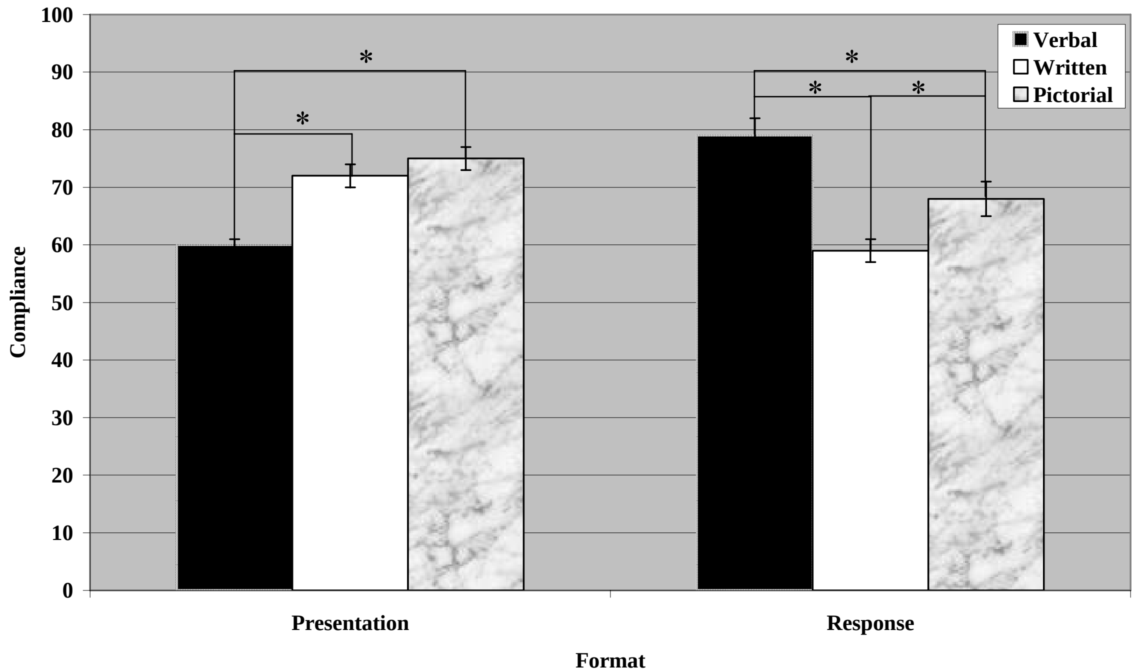


Figure 16: Compliance Scores for the WCCOM task for both the Main Effects of Presentation and Response Format.

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two (*M* = .826, *SD* = .012) than at level four (*M* = .711, *SD* = .019) or eight (*M* = .52, *SD* = .017). In addition, results yield a significant difference between level four (*M* = .711, *SD* = .019) and level eight (*M* = .52, *SD* = .017). Thus, as the rate of task demand increased compliance scores decreased.

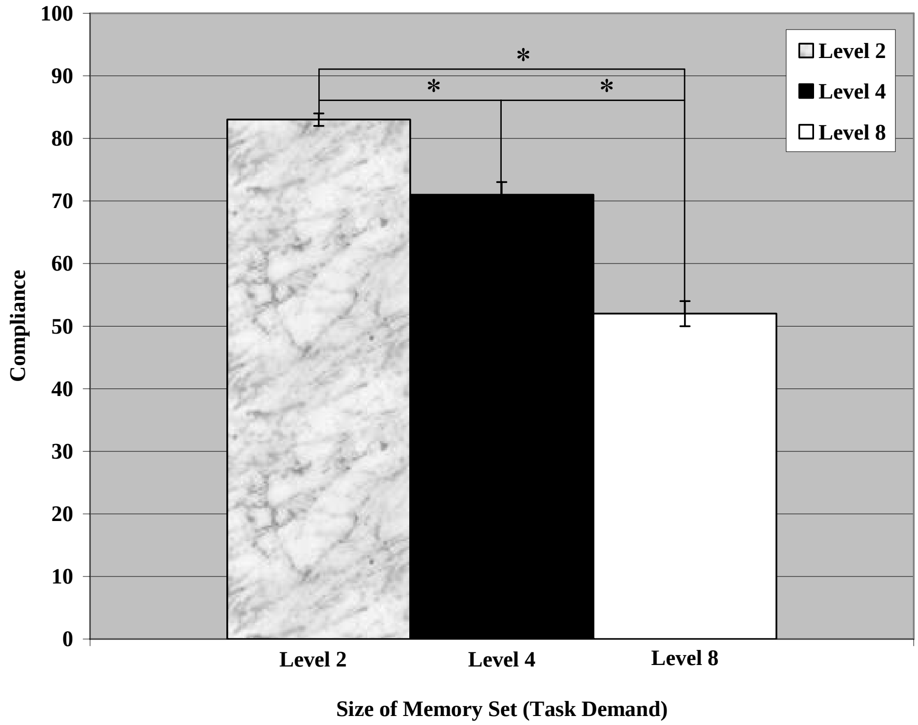


Figure 17: Compliance Scores at each Level of Task Demand for the WCCOM Task

In addition, Duncan post hoc analyses were conducted on presentation format by response format interaction, $F(4, 88) = 11.7$, $p = .0005$, partial $\eta^2 = .35$. Analysis yielded differences between response formats when the warning presentation format remained constant. Specifically, when the presentation format and the response format were both verbal ($M = .774$, $SD = .034$) it resulted in higher compliance as did when the presentation was verbal and the response format was written ($M = .465$, $SD = .024$) and when the presentation was verbal and the response was pictorial ($M = .546$, $SD = .026$).

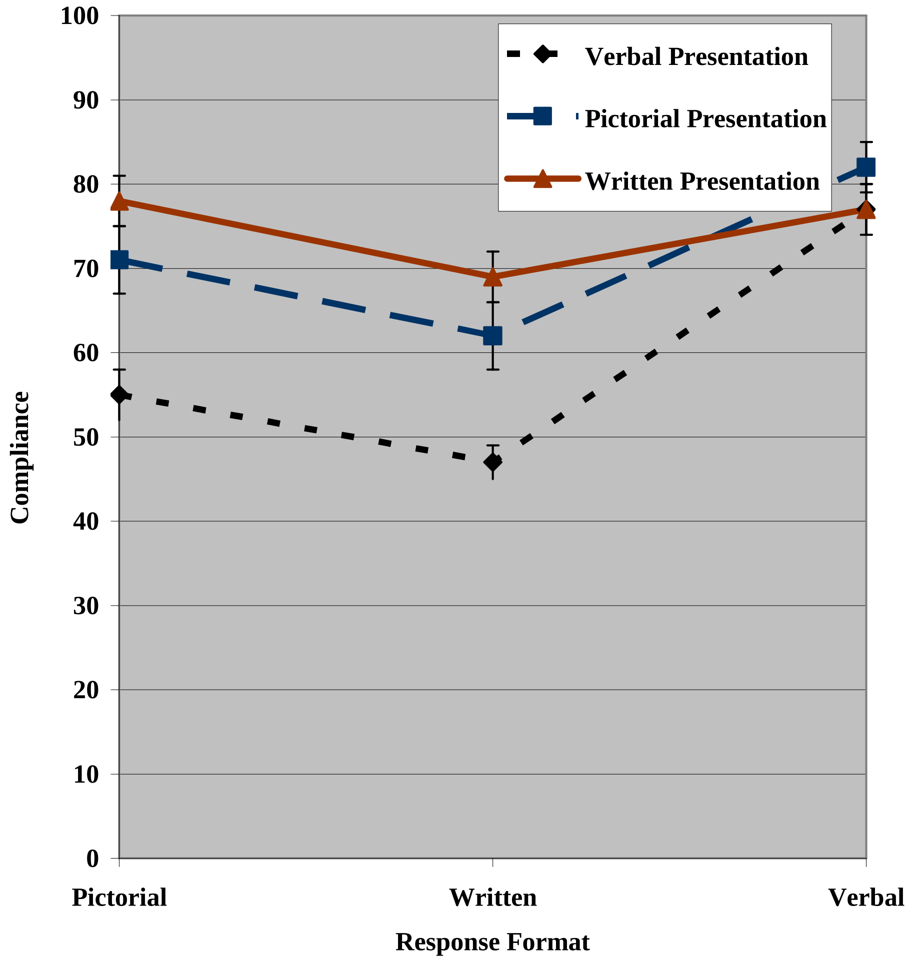


Figure 18: Presentation by Response Format Interaction for the WCCOM Task

When the presentation format and the response format were both pictorial (*M* = .713, *SD* = .036) participants complied more often then when the presentation was pictorial and the response was written (*M* = .617, *SD* = .034). Yet, when the warning presentation format was pictorial and the response format was verbal (*M* = .819, *SD* = .029) compliance was greater then when the presentation was pictorial and the response was pictorial (*M* = .713, *SD* = .036) or written (*M* = .617, *SD* = .034). No differences were found when the warning presentation was written.

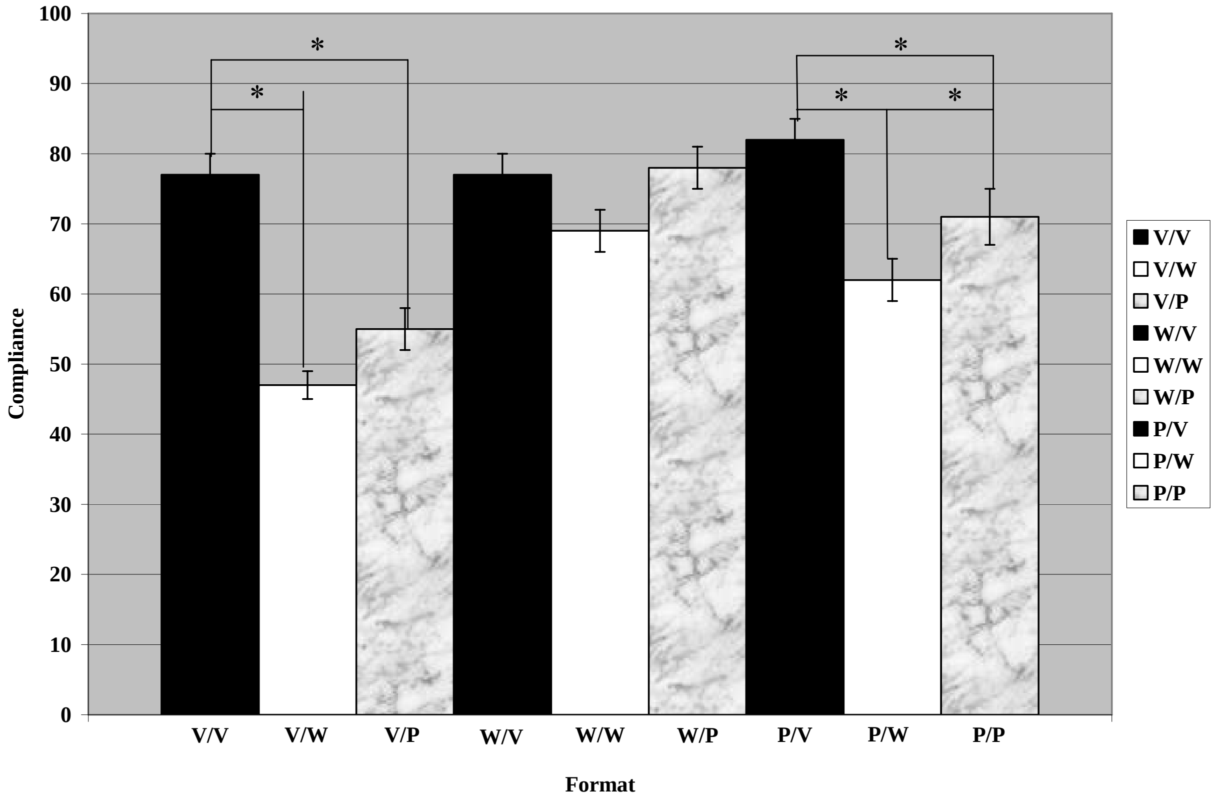


Figure 19: Significant Differences for Presentation when the Response Mode Varied.

Significant differences were also found when the presentation differed, yet the response mode was the same. Written presentation matched with written responses (*M* = .685, *SD* = .027) significantly differed from verbal presentations with written responses (*M* = .465, *SD* = .024). When written warning presentation was coupled with a pictorial response (*M* = .777, *SD* = .027) participants complied more often then when the presentation was verbal and the response was pictorial (*M* = .546, *SD* = .026). Additionally, pictorial presentations coupled with written response (*M* = .617, *SD* = .034) resulted in greater behavioral compliance than verbal presentations with written response format (*M* = .465, *SD* = .024).

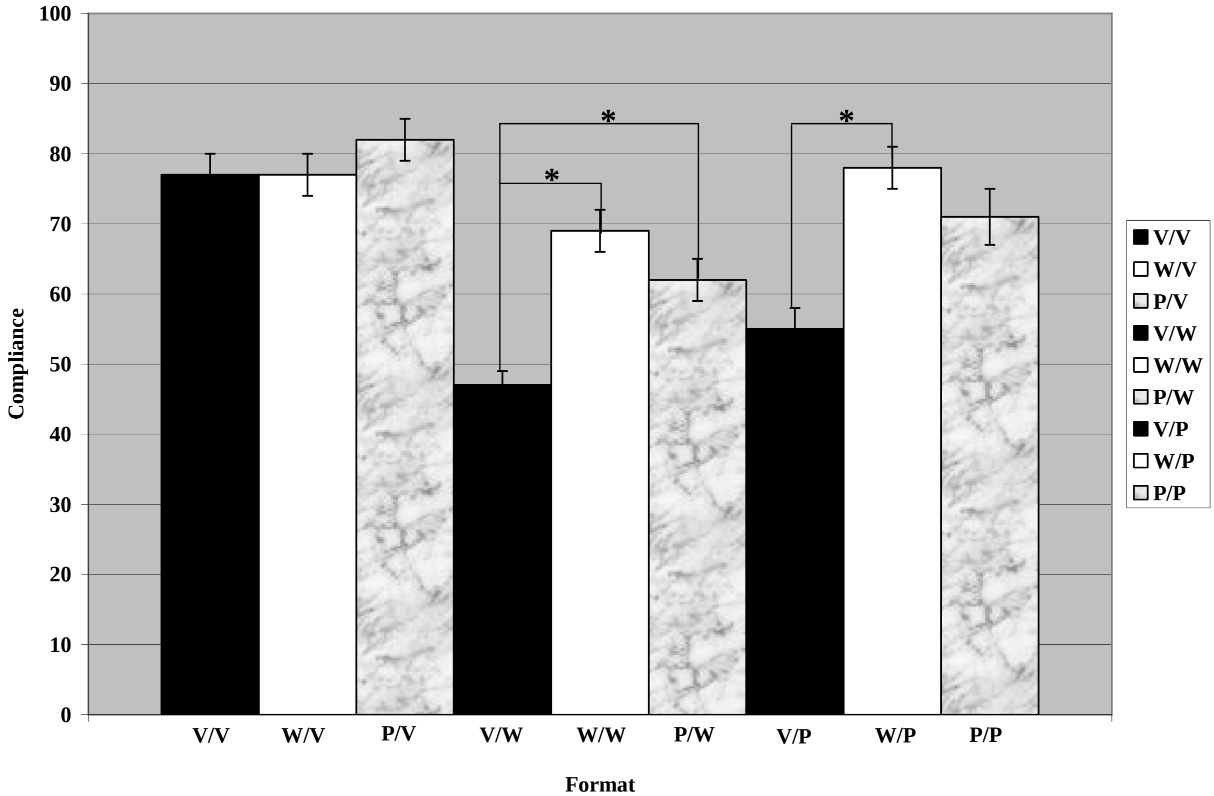


Figure 20: Significant Differences for Response Mode when the Presentation Varied.

Duncan post hoc analyses were also conducted on interaction between presentation format and task demand. Task demand across all formats of warning presentation yielded the same results. Compliance for written, verbal, and pictorials at level two (*M* = .907, *SD* = .021; *M* = .727, *SD* = .030; *M* = .851, *SD* = .019, respectively) were significantly greater than at levels four (*M* = .76, *SD* = .02; *M* = .646, *SD* = .018; *M* = .727, *SD* = .030, respectively) and eight (*M* = .571, *SD* = .027; *M* = .421, *SD* = .013; *M* = .567, *SD* = .021, respectively). Significant differences were also found between level four (*M* = .76, *SD* = .02; *M* = .646, *SD* = .018; *M* = .727, *SD* = .030, respectively) and level eight (*M* = .571, *SD* = .027; *M* = .421, *SD* = .013; *M* = .567, *SD* = .021, respectively).

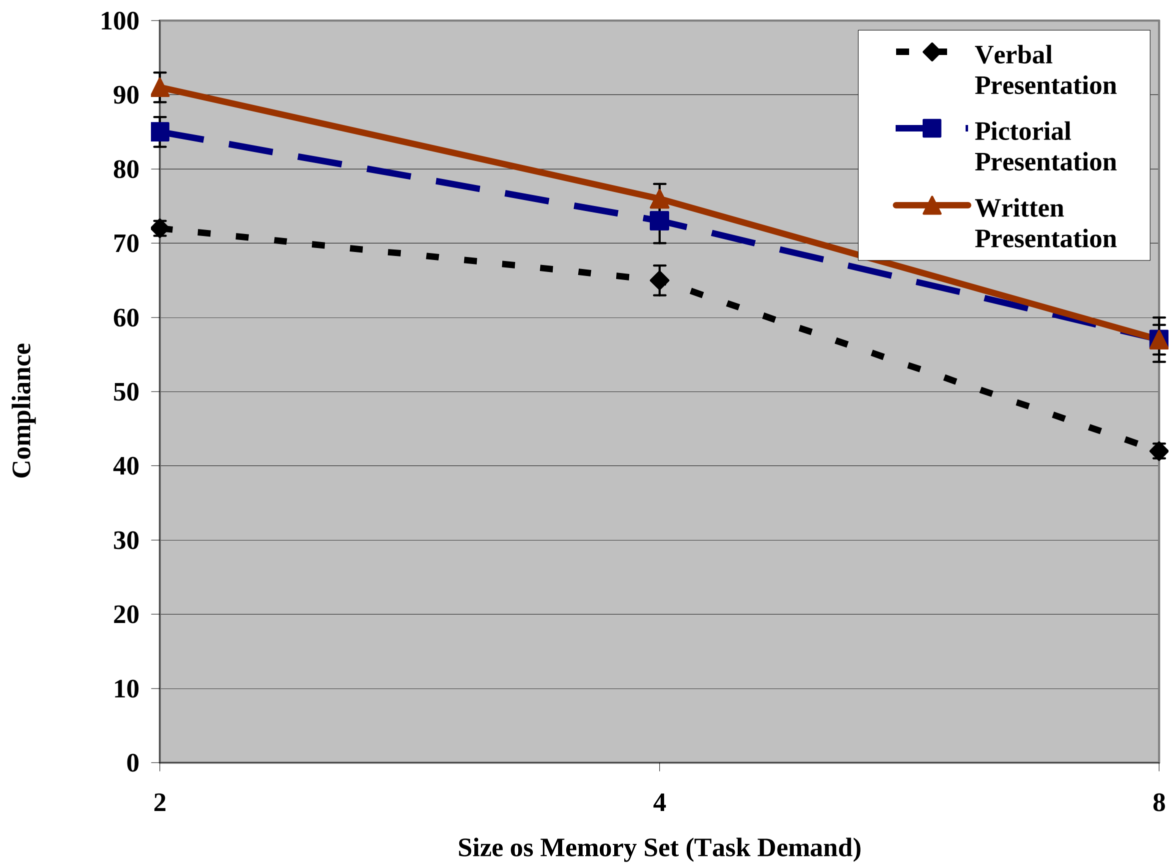


Figure 21: Presentation by Task Demand Interaction for the WCCOM

Post hoc analyses also revealed a difference between presentation formats at the same level of task demand. Written and pictorial warning presentations at level two (*M* = .907, *SD* = .021, *M* = .851, *SD* = .019, respectively) yielded greater compliance than verbal presentation at level two (*M* = .719, *SD* = .013). Similarly, written and pictorial warning presentations at level four (*M* = .76, *SD* = .02, *M* = .727, *SD* = .030, respectively) yielded greater compliance than verbal presentation at level four (*M* = .646, *SD* = .018). Yet again, written and pictorial warning presentations at level eight (*M* = .567, *SD* = .021, *M* = .571, *SD* = .027, respectively) yielded greater compliance than verbal presentation at level eight (*M* = .421, *SD* = .013).

Furthermore, post hoc analyses were conducted for the interaction for warning response format and task demand. Analyses revealed differences between varying levels of task demand when the warning format was the same. For instance, pictorial responses at level two (*M* = .776, *SD* = .024) and level four (*M* = .757, *SD* = .036) were significantly higher than compliance at level eight (*M* = .503, *SD* = .024). Compliance for verbal responses at level two (*M* = .953, *SD* = .013) were significantly greater than at level four (*M* = .782, *SD* = .039). In addition, verbal responses at level four (*M* = .782, *SD* = .039) were greater than at level eight (*M* = .631, *SD* = .038). Finally, written responses at level two (*M* = .748, *SD* = .021) were significantly higher than at level four (*M* = .594, *SD* = .034) or at level eight (*M* = .425, *SD* = .024). Participants also complied more often at level four (*M* = .594, *SD* = .034) for written responses than at eight (*M* = .425, *SD* = .024).

Results for the Duncan post hoc analyses also yielded significant differences across response format type at the same level of task demand. When the task demand was at level two, verbal response (*M* = .953, *SD* = .013) resulted in higher compliance than did pictorial (*M* = .776, *SD* = .024). At task demand level four, pictorial response (*M* = .757, *SD* = .036) had higher compliance scores than did written (*M* = .594, *SD* = .034) responses. Verbal response at level eight (*M* = .631, *SD* = .038) yielded higher compliance than did pictorial responses at that same level (*M* = .503, *SD* = .024).

Table 27
The ANOVA Table for WCCOM Task for the Collapsed Data

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Response Format	2	2.071	15.112	.000	.407	.999
Error(RESP)	44	.137				
Presentation Format	2	1.306	43.940	.000	.666	1.000
Error(PRES)	44	.030				
Task Demand	2	4.956	311.461	.000	.934	1.000
Error(WM)	44	.016				
Response * Presentation	4	.297	11.719	.000	.348	1.000
Error(Response *Presentation)	88	.025				
Response * Task Demand	4	.124	5.176	.001	.190	.961
Error(Response*Task Demand)	88	.024				
Presentation * Task Demand	4	.041	3.322	.014	.131	.824
Error(Presentation*Task Demand)	88	.012				
Response * Presentation* Task Demand	8	.025	2.064	.042	.086	.824
Error(Response*Presentation*Task Demand)	176	.012				

a Computed using alpha = .05

A second analysis on the WCCOM task was conducted in order to examine the differences between warning presentations, response mode and trial on reaction time. A three-way 3 (format) X 2 (response format) 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on warning compliance reaction time. The independent variables were format, response mode, task demand, and trial. The three levels of warning format: verbal, pictorial, and written were presented to the participant at varying levels of task demand: presentations of two, four, and eight warning-color combinations, over 15 trials per block. Two levels of response mode were also factors of interest, written and pictorial formats. The dependent variable was reaction time.

There was a significant main effect of presentation format, task demand, and trial Wilk's Λ = .81, $F(2, 40) = 3.641$, $p = .035$, partial $\eta^2 = .154$, and Wilk's Λ = .12, $F(2, 40) = 95.9$, p =

.0005, partial η^2 = .83, Wilk's Λ = .54, $F(4, 80)$ =6.26, p < .0005, partial η^2 = .99, respectively. No main effect for response format was found, Wilk's Λ = .54, $F(1, 20)$ =.000, p > .05, partial η^2 = .000. A significant interaction was found between presentation and task demand, $F(4, 80)$ =6.3, p< .0005, partial η^2 = .24.

Table 28
ANOVA Table for WCCOM task Reaction Time

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Presentation	2	22365454.702	3.638	.035	.154	.638
Response	1	307.051	.000	.991	.000	.050
Task Demand	2	65445560.181	95.873	.000	.827	1.000
Trial	4	3159763.562	6.261	.000	.238	.985
Presentation * Response	2	1708815.551	.471	.628	.023	.122
Presentation * Task Demand	4	2760184.074	6.335	.000	.241	.986
Response * Task Demand	2	691227.517	1.208	.309	.057	.249
Presentation * Response * Task Demand	4	615868.771	1.227	.306	.058	.368
Presentation * Trial	8	793710.029	1.662	.112	.077	.714
Response * Trial	4	699971.740	1.980	.105	.090	.571
Presentation * Response * Trial	8	267301.600	.568	.803	.028	.257
Task Demand * Trial	8	72372.482	.241	.983	.012	.123
Task Demand * Presentation * Trial	16	388608.685	1.253	.226	.059	.800
Response * Trial * Task Demand	8	341410.979	.842	.567	.040	.383
Presentation * Response * Trial* Task Demand	16	320697.979	.786	.702	.038	.541

A set of Fisher LSD post hoc tests for presentation format yielded a significant difference between verbal (*M* =2297.2ms, *SD* = 116.1) presentation format and written format (*M*

=2673.2ms, *SD* = 132.7). No other differences were found for presentation format. Thus, responses were faster when warnings were presented in verbal format compared to written format.

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants responded significantly faster at levels of two (*M* = 2136.5ms, *SD* = 82.9) and four (*M* = 2522.9ms, *SD* = 88.1) than at level eight (*M* = 2776.5ms, *SD* = 76.87). Significant differences in reaction time were also found between levels two (*M* = 2136.5ms, *SD* = 82.9) and eight (*M* = 2776.5ms, *SD* = 76.87).

A set of Fisher LSD post hoc tests for trial yielded a significant difference between trial one (*M* =2634.01ms, *SD* = 63.6) and trails two (*M* =2486.9ms, *SD* = 88.2), three (*M* =2436.3ms, *SD* = 83.8), and four (*M* =2411.8ms, *SD* = 87.6). No other differences were found between other trials. Thus, reaction time was slower in trial one compared to all other trials.

Ghost Recon Tasks

Since both operational tasks were used for this line of research and have different measures of performance, the data was analyzed separately. Thus, the shooting task data consists of Experiments 3, 4, and 5. The navigation task data consisted of Experiments 6, 7, and 8. Again, Experiments 1 and 2 were excluded form these analyses so that the so that the number of participants would be even across experiments.

Ghost Recon Shooting Task

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way

3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the amount of enemies killed. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15 trials. The dependent variable was the percent of enemies killed in each building (enemies in the building divided by enemies killed).

There was a significant main effects of task demand, Wilk's $\Lambda = .71$, $F(2, 66) = 6.3$, $p = .003$, partial $\eta^2 = .16$. Format or trial did not result in a main effect, Wilk's $\Lambda = .99$, $F(2, 66) = .13$, $p > .05$, partial $\eta^2 = .004$, Wilk's $\Lambda = .84$, $F(4, 132) = 1.3$, $p > .05$, partial $\eta^2 = .04$, respectively. No significant interactions were found.

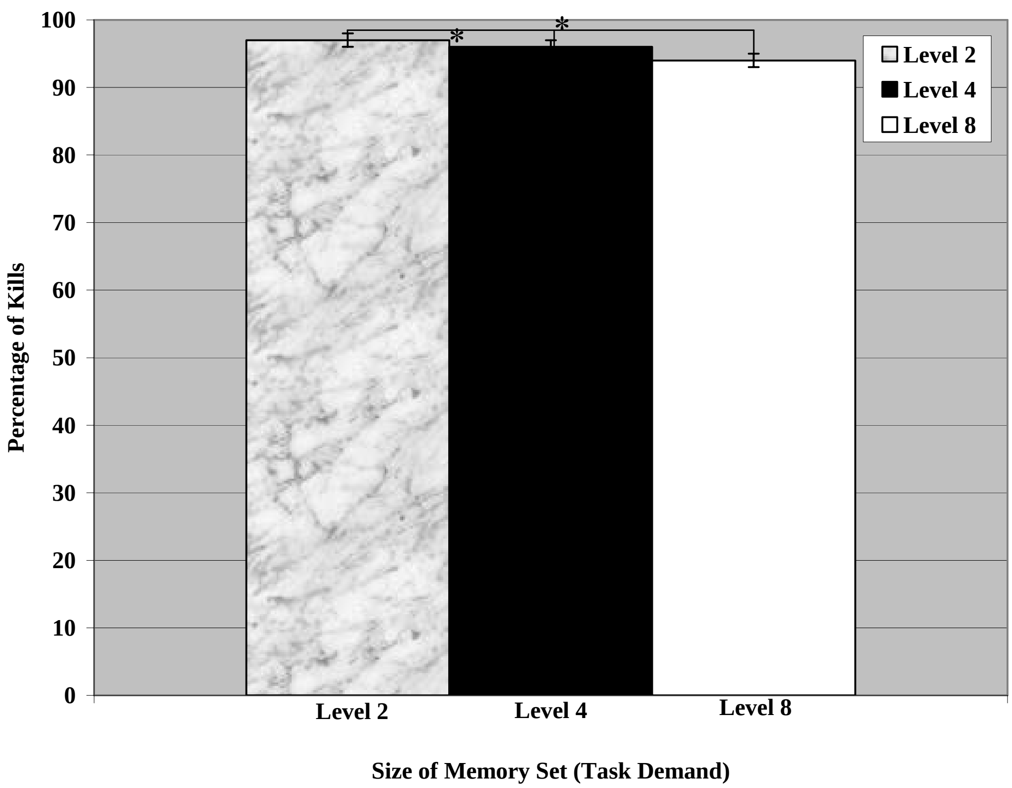


Figure 22: Main Effect for Task Demand at each Level for the Ghost Recon Shooting Task

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants were significantly more likely to comply at levels of two (*M* = .965, *SD* = .006) than at level four (*M* = .958, *SD* = .006) or eight (*M* = .939, *SD* = .009). No significant difference between level two (*M* = .965*, SD* = .006) and level four (*M* = .958, *SD* = .006) were found.

Table 29
ANOVA Table for the Ghost Recon Shooting Task for the Collapsed Data

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.003	.125	.883	.004	.068
Error(Format)	66	.021				
Task Demand	2	.097	6.291	.003	.160	.884
Error(Task Demand)	66	.015				
Trial	4	.018	1.339	.259	.039	.409
Error(TRIAL)	132	.014				
Format * Task Demand	4	.030	1.971	.103	.056	.580
Error(Format * Task Demand)	132	.015				
Format * Trial	8	.013	.871	.541	.026	.403
Error(Format * Trial)	264	.015				
Task Demand * Trial	8	.013	.905	.513	.027	.419
Error(Task Demand * Trial)	264	.014				
Format * Task Demand * Trial	16	.015	.984	.473	.029	.676
Error(Format * Task Demand * Trial)	528	.016				

a Computed using alpha = .05

Ghost Recon Navigation Task

Two measures were of interest for the navigation task, performance and time. Specifically, performance was measured by the number of waypoints that were reached in each trial (0-4). Additionally, a temporal measurement was also considered, the time that it took to reach all waypoints compared top the “goal time,” the fastest amount of time to navigate to all four waypoints.

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on total number of waypoints reached. The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15 trials. The dependent variable was the number of waypoints reached (0-4).

There was a significant main effect of task demand, Wilk's Λ = .58, $F(2, 58) = 6.95$, p = .002, partial η^2 = .193. No main effect was found for format or trial, Wilk's Λ = .97, $F(2, 58)$ = .41, $p > .05$, partial η^2 = .014, Wilk's Λ = .90, $F(4, 116)$ = .927, $p > .05$, partial η^2 = .03, respectively. No significant interactions were found.

Table 30
The ANOVA Table for the Ghost Recon Navigation Task for the Collapsed Data

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.033	.410	.666	.014	.113
Error(Format)	58	.081				
Task Demand	2	.250	6.947	.002	.193	.912
Error(Task Demand)	58	.036				
Trial	4	.034	.927	.451	.031	.286
Error(TRIAL)	116	.037				
Format * Task Demand	4	.038	.827	.510	.028	.258
Error(Format * Task Demand)	116	.046				
Format * Trial	8	.025	.605	.773	.020	.278
Error(Format * Trial)	232	.041				
Task Demand * Trial	8	.040	1.021	.421	.034	.471
Error(Task Demand * Trial)	232	.039				
Format * Task Demand * Trial	16	.024	.589	.893	.020	.406
Error(Format * Task Demand * Trial)	464	.041				

Computed using alpha = .05

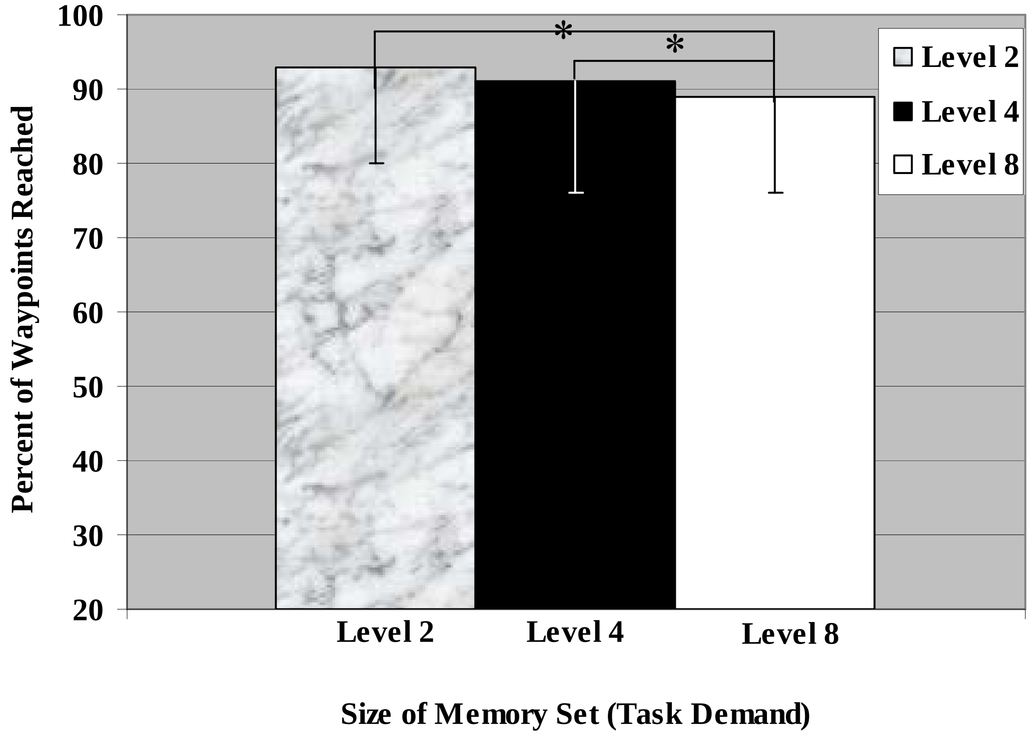


Figure 23: Main Effect for Task Demand at each Level for the Ghost Recon Navigation Task for Performance

In order to determine if the format of the warning presentation or task demand affected compliance behavior when the second task was predominantly a visual/spatial task, a three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the time it took to reach all 4 waypoints (temporal measurement). The independent variables included format: verbal, pictorial, and written, task demand: presentations of two, four, and eight, and trial: three blocks of 15 trials. The dependent variable was the time (in %) it took participants to reach all four waypoints ((actual time-goal time)/goal time X 100).

There was a significant main effect of task demand, Wilk's Λ = .51, $F(2, 70) = 13.8$, p = .0005, partial η^2 = .99. No main effect was found for format or trial, Wilk's Λ = .99, $F(2, 70)$ =

.17, $p > .05$, partial $\eta^2 = .07$, Wilk's $\Lambda = .87$, $F(4, 140) = 1.5$, $p > .05$, partial $\eta^2 = .04$, respectively. No significant interactions were found.

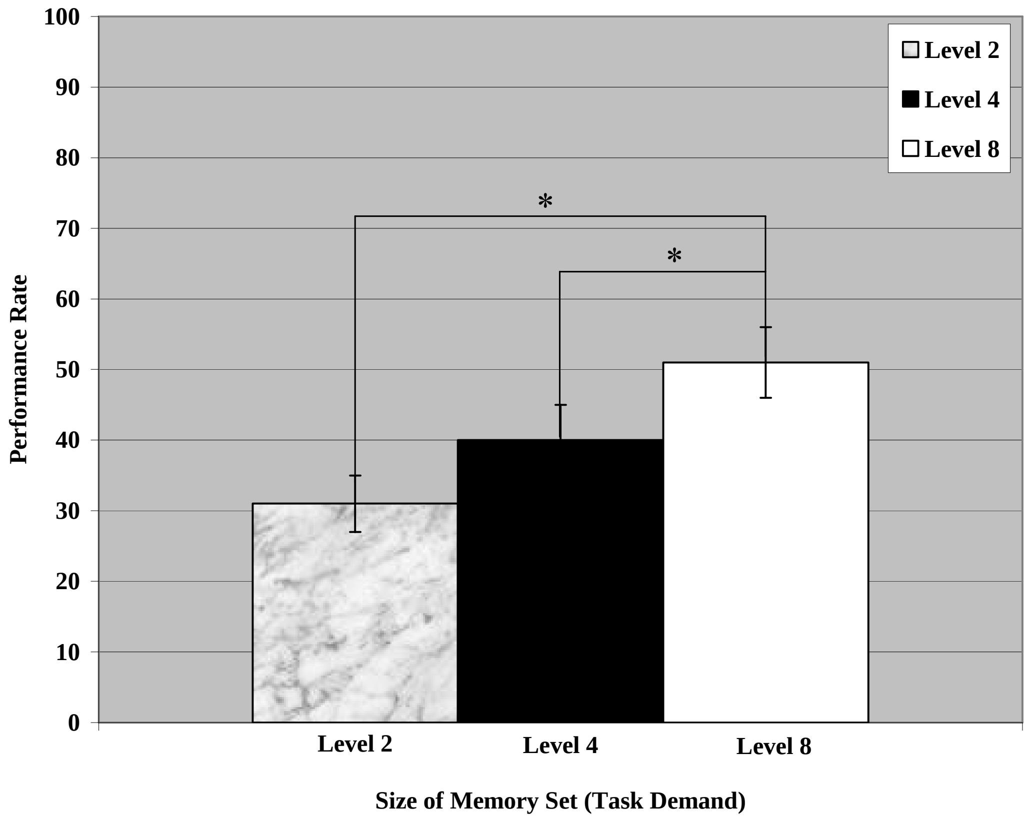


Figure 24: Main Effect for Task Demand at each Level for the Ghost Recon Navigation Task

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants took significantly longer at level eight (*M* = .508, *SD* = .05) than at level four (*M* = .398, *SD* = .05) and at level 2 (*M* = .311, *SD* = .04). In addition, a significant difference between level four (*M* = .398, *SD* = .05) and eight (*M* = .508, *SD* = .05) were found.

Table 31
The ANOVA Table for the Ghost Recon Navigation Task for the Collapsed Data

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.169	.173	.841	.005	.076
Error(Format)	70	.975				
Task Demand	2	5.262	13.844	.000	.283	.998
Error(Task Demand)	70	.380				
Trial	4	.539	1.487	.209	.041	.452
Error(TRIAL)	140	.363				
Format * Task Demand	4	.594	1.105	.357	.031	.341
Error(Format * Task Demand)	140	.538				
Format * Trial	8	.388	.905	.512	.025	.420
Error(Format * Trial)	280	.429				
Task Demand * Trial	8	.655	1.523	.149	.042	.678
Error(Task Demand * Trial)	280	.430				
Format * Task Demand * Trial	16	.145	.329	.994	.009	.221
Error(Format * Task Demand * Trial)	560	.441				

a Computed using alpha = .05

NASA-TLX

In order to determine if format of presentation affected subjective mental workload ratings, a 3 (response format) X 3 (presentation format) analysis of variance (ANOVA) was conducted. The independent variables were warning format and presentation format. Warning response format included three levels: verbal, written and pictorial warnings. Presentation format also included three levels: verbal, written and pictorial warnings. The dependent variable was the subjective mental workload effort determined by using the NASA-Task Load Index (NASA-TLX).

No main effect for response format or presentation format was found, Wilk's Λ = .85, $F(2, 36) = 1$, $p > .05$, partial $\eta^2 = .37$, Wilk's Λ = .80, $F(2, 36) = 2.4$, $p > .05$, partial $\eta^2 = .45$, respectively. Thus, participants did not report feeling more mental demand in one response or presentation format vice another format.

The non-significant findings for the NASA-TLX across warning presentation and response format was expected since all but one of the individual experiments resulted in non significant results (see table 30).

Table 32
Results of the NASA-TLX across all Eight Experiments

Experiment #	Number of Participants	Results of NASA-TLX	Post Hoc Results
1	6	ns	NA
2	6	ns	NA
3	11	ns	NA
4	11	ns	NA
5	12	ns	NA
6	12	ns	NA
7	12	Significant effect	Pictorials > Written or Verbal format
8	12	ns	NA

Due to the design of the study, the NASA-TLX results may be due to the combination of the varying presentations and response formats that were mixed (e.g. verbal, written, pictorial presentation with pictorial response (Experiment 3)). Thus, it deemed necessary to analyses the scores of the NASA-TLX on the effects of the presentation, response, and the operational task since all of these variables may have contributed to workload. Thus a 3 (response format) X 3 (presentation format) X 2 (operational task) analysis of variance (ANOVA) was conducted. The independent variables were warning format, presentation format, and the operational tasks. Warning response format included three levels: verbal, written and pictorial warnings. Presentation format also included three levels: verbal, written and pictorial warnings. The operational task consisted of the shooting task and the navigation task. The dependent variable was the subjective mental workload effort determined by using the NASA-Task Load Index.

No main effect for presentation format was found, Wilk's $\Lambda = .72$, $F(2, 20) = 1.9$, $p > .05$, partial $\eta^2 = .16$. A main effect for Response mode was found, Wilk's $\Lambda = .48$, $F(2, 20) = 4.2$, $p = .03$, partial $\eta^2 = .3$. No interactions were found.

Table 33
The ANOVA Table for the Main Effect of Response Format for NASA-TLX

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Presentation	2	664.472	1.888	.177	.159	.345
Response	2	376.786	4.175	.031	.295	.667
Operational Task	1	121.903	.161	.697	.016	.065
Presentation * Response	4	214.898	1.699	.169	.145	.473
Presentation * Operational Task	2	1232.056	3.027	.071	.232	.521
Response * Operational Task	2	31.069	.312	.736	.030	.093
Presentation * Response * Operational Task	4	129.920	1.345	.270	.119	.380

RSME

A three-way 3 (format) X 3 (task demand) X 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the Rating Scale Mental Effort (RSME) subjective workload measure. The independent variables were format, task demand, and trial. The dependent variable of interest was the RSME. Results of this analysis yielded a main effect for task demand, Wilk's Λ = .06, $F(2,44) = 203.7$, $p =.0005$, partial $\eta^2 = .9$. A main effect was not found for response format or presentation format, Wilk's Λ = .98, $F(2,44) = .19$, $p >.05$, partial $\eta^2 = .01$, Wilk's Λ = .87, $F(2,44) = 1.3$, $p >.05$, partial $\eta^2 = .27$, respectively. No interaction was found.

Furthermore, Fisher LSD post hoc tests for task demand revealed that participants had the least amount of mental workload at level of two (*M* = .271, *SD* = .156) compared to level four (*M* = .431, *SD* = .165) or eight (*M* = .625, *SD* = .168). In addition, results yield a significant difference between level four (*M* = .431, *SD* = .165) and level eight (*M* = .625, *SD* = .168). Thus, as the rate of task demand increased subjective workload ratings increased.

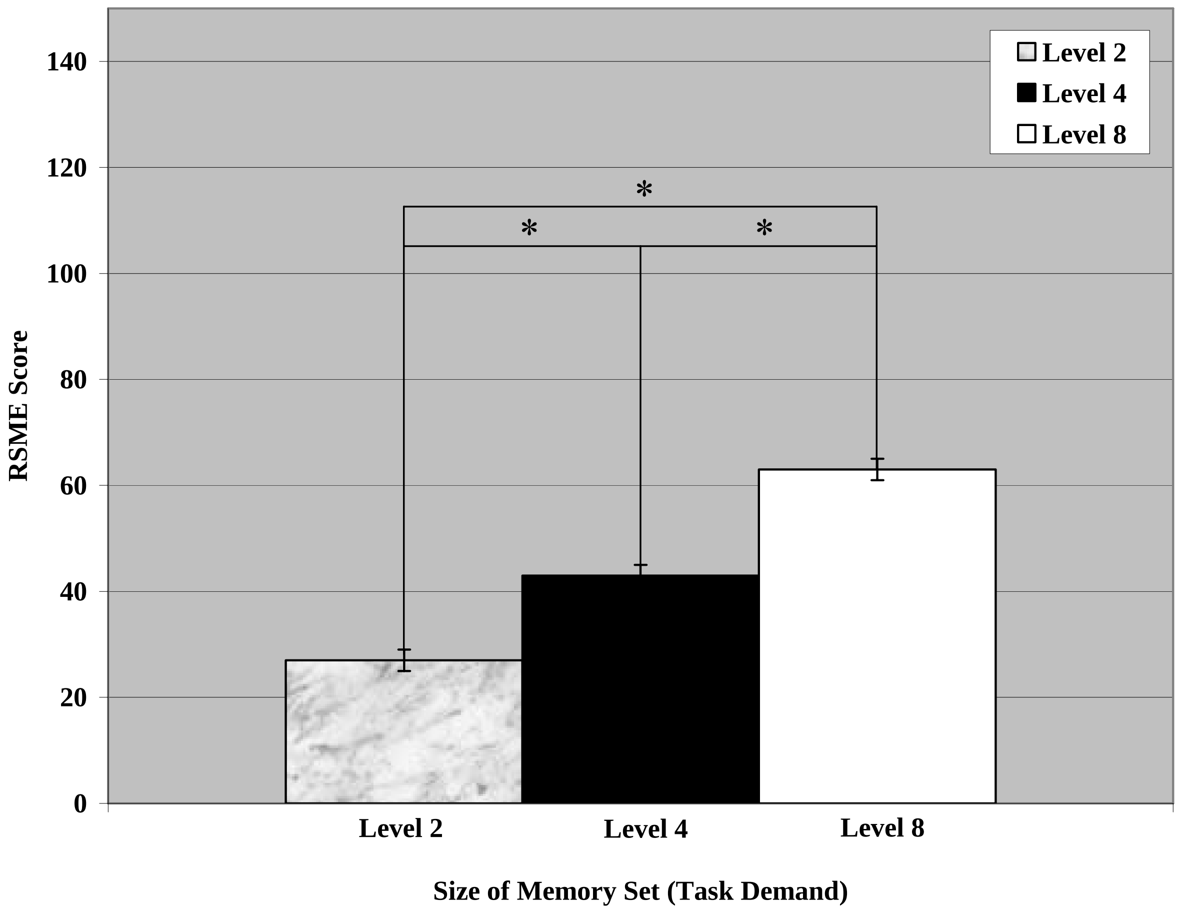


Figure 25: Main Effect for Task Demand at each Level for the RSME

Table 34
The ANOVA Table for the RSME for the Collapsed Data

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Presentation Format	2	2487.555	1.341	.272	.057	.274
Error(Presentation)	44	1855.211				
Response Format	2	48.156	.189	.828	.009	.078
Error(Response)	44	254.663				
Task Demand	2	64867.721	203.672	.000	.903	1.000
Error(Task Demand)	44	318.492				
Presentation * Response	4	375.989	1.690	.160	.071	.499
Error(Presentation * Response)	88	222.530				
Presentation * Task Demand	4	538.313	1.325	.267	.057	.398
Error(Presentation * Task Demand)	88	406.181				
Response * Task Demand	4	65.887	1.030	.396	.045	.313
Error(Response * Task Demand)	88	63.949				
PRES * Response * Task Demand	8	44.084	.607	.771	.027	.276
Error(Presentation * Response * Task Demand)	176	72.591				

a Computed using alpha = .05

Additionally, it was of interest to determine if any relationship existed between the workload measures, the NASA-TLX and the RSME. Therefore, the scores on the mental effort subscale of the TLX were correlated with the RSME score. Experiments 3-8 were individually analyzed. Results of this correlation only yielded significant correlations on Experiment 3 between mental effort on the TLX and the RSME scores.

Table 35
Correlation Table for Experiment 3 for the NASA-TLX and RSME Scores

	NASA-TLX Pictorial Rating	NASA-TLX Written Rating	NASA-TLX Verbal Rating	RSME Pictorial Rating	RSME Written Rating	RSME Verbal Rating
NASA-TLX Pictorial Rating	1					
NASA-TLX Written Rating	.528	1				
NASA-TLX Verbal Rating	.891(**)	.718(*)	1			
RSME Pictorial Rating	.696(*)	.655(*)	.814(**)	1		
RSME Written Rating	.325	.745(**)	.619(*)	.805(**)	1	
RSME Verbal Rating	.645(*)	.678(*)	.850(**)	.828(**)	.754(**)	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Working Memory

It was the interest of this study to examine if individual differences in working memory played a role in determining if verbal and spatial working memory capacities predicted performance in different modalities. In order to do this, examination of the separability of the verbal and spatial resources, which were tapped using the four working memory spans had to be performed, thus correlations were performed between the working memory spans (spatial span, reading span, verification arrow, and verification word). Secondly, in order to determine if working memory predicted performance across warning format correlations were performed between the working memory spans and warning presentation format (verbal, written, and pictorial); a second set of correlations were performed between the working memory spans and the nine combinations of warning presentation/response format (pictorial presentation with pictorial, written, and verbal response; written presentation with pictorial, written, and verbal response; and verbal presentation with pictorial, written, and verbal response).

The analysis on the working memory spans included all participants from the eight studies that were conducted, equating to an $\underline{N} = 83$. Over the eight experiments, only one participants' reading and verification word span, and three verification arrow spans were not included in the final analysis due to computer malfunctions (missing data). A lenient cutoff for significant correlations was set at .10 for all analyses.

The results of the correlation analysis between working memory spans (spatial span, reading span, verification arrow, and verification word) yielded reading span as significantly correlated with the verification word span $r(82) = .42$, $p < .0005$. Thus, the reading span correlation with the verification word span was .42, accounting for 58% of the variance. Thus, verbal working memory spans, reading and verification word span support complex language processing. Furthermore, verification arrow was significantly correlated with spatial span $r(80) = .21$, $p < .05$ and with the verification word $r(80) = .21$, $p < .05$. Therefore, the verification arrow, which taps both language and spatial processing, is correlated with both the spatial span task and the language-processing task, verification word. The correlation between the reading span and verification arrow was not significantly significant, $r(80) = .13$, $p > .10$.

The correlations between the span tasks (spatial span, reading span, verification arrow, and verification word) and warning presentation format (verbal, written, and pictorial) resulted in significant correlations between reading span and verbal presentation $r(81) = .15$, $p = .095$, no other significant correlations were found for reading span. Spatial span was significantly correlated with verbal $r(82) = .18$, $p = .06$ and pictorial $r(82) = .26$, $p < .05$ presentations. Significant correlations were found between verification word and verbal $r(81) = .22$, $p <.05$ and pictorial presentations $r(81) = .21$, $p < .05$. Verification arrow was also significantly correlated with verbal presentations $r(79) = .28$, $p < .05$and pictorial $r(79) = .3$, $p <.05$ (.3). None of the span tasks were significantly correlated with written presentation.

Table 36
Correlation Table for Span Tasks by Warning Presentation Format

	Reading Span	Spatial Span	Verification Word	Verification Arrow
1. Reading Span	_	_	_	_
2. Spatial Span	-.016	_	_	_
3. Verification Word	.419(**)	-.078	_	_
4. Verification Arrow	.133	.214(*)	.205(*)	_
5.Verbal Presentation	.147	.178	.217(*)	.277(**)
6. Pictorial Presentation	.111	.263(**)	.207(*)	.301(**)
7. Written Presentation	.042	.005	.055	.112

** Correlation is significant at the 0.01 level (1-tailed).
* Correlation is significant at the 0.05 level (1-tailed).

Correlations were performed between the working memory spans and the nine combinations of warning presentation/response format (pictorial presentation with pictorial, written, and verbal response; written presentation with pictorial, written, and verbal response; and verbal presentation with pictorial, written, and verbal response). The results of this analysis revealed that reading span was correlated with picture presentation/picture response format $r(22)$ $=.3$, $p<.10$, written presentation/verbal response $r(22) = .37$, $p < .05$, and verbal presentation/ verbal response $r(22) = -.31$, $p < .10$. Spatial span was significantly correlated with picture presentation/picture response $r(23) = .43$, $p < .05$, picture presentation/written response $r(23) =$ $.32$, $p < .10$, and written presentation/picture response $r(23) = .31$, $p < .10$. Verification word span was only significantly correlated with picture presentation/verbal response $r(22) = .43$, p

$<.05$. The verification arrow task was not significantly correlated with any of the nine warning combinations.

Table 37
Correlation Table for Span Tasks by Presentation/Response Format

	Reading Span	Spatial Span	Verification Word	Verification Arrow
1. Reading Span	_	_	_	_
2. Spatial Span	.055	_	_	_
3. Verification Word	.495(**)	-.082	_	_
4. Verification Arrow	.213(*)	.226(*)	.181	_
5. Pictorial Presentation/Verbal Response	.179	.256	.430(*)	.161
6. Pictorial Presentation/Pictorial Response	.296	.428(*)	.383(*)	-.267
7. Pictorial Presentation/Written Response	.136	.315	.217	-.236
8. Written Presentation/Verbal Response	.372(*)	-.026	.105	.054
9. Written Presentation/Pictorial Response	.351	.309	.160	.001
10. Written Presentation/Written Response	.248	-.092	.177	.088
11. Verbal Presentation/Verbal Response	-.309	.119	-.225	-.045
12.Verbal Presentation/Pictorial Response	-.087	-.024	-.081	-.032
13.Verbal Presentation/Written Response	-.127	-.170	-.154	-.143

** Correlation is significant at the 0.01 level (1-tailed).
* Correlation is significant at the 0.05 level (1-tailed).

Gender

Historically, gender differences in spatial ability have been found to favor men. Linn and Peterson (1985) conducted a meta- analysis of the gender differences in abstract spatial tasks, they found that men outperform women on mental rotation and spatial perception tasks. This trend is also found in real world spatial tasks (Beatty & Troster, 1987). Thus, it was the interest of this line of research to determine if a gender played a role in performance on the spatial tasks. Therefore a three-way 3 (response format) X 3 (presentation format) X 3 (task demand) within-participants repeated measure analysis of covariance (ANCOVA) was conducted on warning compliance behavior. Gender was the covariate. Results of this analysis yielded gender as a non-significant factor, $F(1,20) = .09$, $p > .05$, partial $\eta^2 = .01$.

DISCUSSION

Warning Format

The first two hypotheses looked at differences between warning presentation format. Hypothesis 1 predicted that participants would have a higher rate of compliance when warnings were presented in verbal compared to written and pictorial format. This prediction was made based on the supporting literature in the memory and warnings domain. Although not all evidence in the memory literature supports the verbal superiority effect, auditory modes of communication were generally found to be superior (Penney, 1975; Murdock, 1968; Watkins & Watkins, 1980). In addition, the literature in the warnings domain has also found greater behavioral compliance when warnings were presented in a verbal format in various environmental settings (Jaynes & Boles, 1990; Wogalter, Rashid, Clarke, & Kalsher, 1991; Wogalter & Young, 1991). Thus, it was predicted that results of the eight experimental studies would replicate the results of the previous literature on format differences. Subsequently, none of the experiments yielded verbal format as the superior mode of warning presentation. The presentation format that resulted in the highest behavioral compliance was either pictorial or written.

Although the results of the previous literature on memory and warnings were based on single task performance, it was also taken into account that warnings in this study would be presented in a dual task paradigm. Firstly, Broadbent, Vines, and Broadbent (1978) and Gardiner, Thompson, and Maskarinec (1974) found that the task that fills the interval of time between presentation and recall is dependent on how much information is remembered. They found that if the interval between presentation and recall were silent or if non-verbal distractions were present, auditory information was recalled at a greater rate than visual information. Recall was reduced for verbal recall when an auditory distracter was presented in the time interval between presentation and response more than when non-verbal distracters were present. Both distracters, verbal or visual reduced the rate of recall on visually presented words. In our study, an interval of time between warning presentation and warning recall existed, the operational task (a visual-spatial task) that filled the interim was important to consider.

Secondly, the research by Wickens, Sandry, and Vidulich, (1983) was also taken into account when considering the mode of presentation that will result in the greatest compliance since warnings are presented while performing the operational task. They found that cross-modal timesharing is better than intra-modal (Wickens, Sandry, & Vidulich, 1983), thus verbal warnings should have resulted in the highest behavioral compliance.

Therefore, Hypothesis 2 also predicted verbal warning would be the superior mode of warning presentation since the operational task in this study, that filled the time interval between warning presentation and recall, was a visual/spatial task. Contradictory to our predictions and the theories that supported them (Broadbent, Vines, & Broadbent, 1978; Gardiner, Thompson, & Maskarinec, 1974; Wickens, Sandry, & Vidulich, 1983) it was found that verbal warnings were the inferior mode of warning presentation across all eight experiments (or at least not significantly different from written or pictorial formats in Experiments 5 and 8 where the response mode was verbal).

Inconsistent with our predictions, verbal warning presentations did not result in the superior format. Furthermore, the operational task, which was preformed in the interim, did not degrade performance on the visual warning presentations (pictorials and written warnings) more than then the verbal warnings. In addition, the format of warning presentation did not affect performance on the shooting or the navigation operational tasks as predicted. Thus, the visual/spatial operational task, regardless of its complexity did not interfere in timesharing with intra-modal warning presentations or cross modal timesharing.

To get a better look at the experiments overall, the data was collapsed by response mode. Thus, Experiments 3 and 6, 4 and 7, and 5 and 8 (respectively) were combined. Response mode played an integral role in the experimental design. No specific predictions were made on response format because currently, no literature exists on the effects they may have on performance. When the data was collapsed, presentation format still yielded written and pictorial warnings as the superior format. Yet, the analyses for response format concluded that participants were significantly more likely to comply when the response format was verbal compared to pictorial or written.

Reaction time data was also considered in order to determine if a speed/accuracy tradeoff emerged. Compliance and reaction time for the WCCOM task were separately analyzed in the current study in order to determine if a speed/accuracy tradeoff surfaced. Since presentation format was a significant factor in both WCCOM analyses, it was possible to compare the percentage of correct compliance scores with the speed of response. Compliance scores for presentation resulted in significant differences between verbal format and both pictorial and written, yielding the lowest compliance behavior for verbal presentations. No differences resulted between compliance scores for written and pictorial formats. Subsequently, the analysis on reaction time yielded a significantly higher reaction time for verbal presentation formats compared to written. No other differences were found for reaction time between pictorial and written or verbal presentations. What was concluded from the comparison of the compliance scores and reaction times for presentation format is that a speed/accuracy tradeoff did exist for verbal presentations. Although verbal responses had faster reaction times than written, it verbal presentation yielded lower compliance scores than written or pictorials; thus, the fast rate of response resulted in compliance errors.

It was of interest of the current study to determine if response format affected compliance. Additionally, speed/accuracy tradeoffs were also investigated. Results of the WCCOM compliance data suggests that participants were significantly more likely to comply when the response format was verbal (M = .789, SD = .028) and pictorial (M = .679, SD = .026) compared to written response format (M = .589, SD = .023). Additionally, significant differences were found between verbal (M = .789, SD = .028) and pictorial response format (M = .679, SD = .026). Thus, verbal response format resulted in the superior response mode when the experimental data was collapsed.

Compliance scores differed on the WCCOM task resulting in verbal response having the highest compliance, followed by pictorials than written. Although the reaction time data for the verbal response format was not all available, nine participants' data was retrieved. Based on this limited amount of data for verbal response format, analysis were still conducted. Note, that this data is not one one-hundred percent reliable as a measure. Results of this analysis found verbal response as having the longest reaction time (M = 3210.7ms, SD = 917.1) compared to pictorial (M = 2414.1ms, SD = 455.7) or written (M = 2691.8ms, SD =451.4). In light of the results on compliance and reaction time, a speed/accuracy tradeoff yields true for verbal format. In no other instance of response was that the case. The reaction time for verbal may be longer than the other warning formats due to faulty equipment, such that a response that was not recognized resulted in the highest score of 5000ms.

The speed/accuracy tradeoff was not found for task demand for the WCCOM task. Again, compliance and reaction time were compared. As for reaction time, as task demand increased, reaction time increased. Thus, the greater the size of the memory set, the longer the response time. Compliance scores for the WCCOM task resulted in a decrease in compliance as the size of the memory set increased.

The results of the verbal combination formats yielding higher compliance scores may have been an artifact of the study. Technical constraints caused the verbal response format to be scored less conservatively than the pictorial or written. For the pictorial and written responses, which were exact and coded by the computer, reaction time was limited to five hundred milliseconds. Only a small portion of the verbal response were coded for reaction time, such that only a small portion of the sample was used in the reaction time data for response mode. Therefore, the reaction time data for verbal response should be cautiously interpreted.

The proximity compatibility principle (PCP) (Barnett & Wickens, 1988; Wickens and Andre, 1990; Wickens & Carswell, 1995) takes into account the ways that multiple display channels can be integrated. This principle takes into account the spatial compatibility of displays, yet does not consider the combination of the format of presentation and the response mode. A gap in the literature exists as far as the combinations of presentation and response mode that result in the greatest performance. In light of the limited research on presentation/response format, our study found that the presentation and response mode must be taken into consideration when developing warnings and may be task dependent. This area of research is in need of further development.

Task Based Stress

Hypotheses 3, 4 and 5 were specific to the effects of task demand on performance. Two separate areas of task based stress were of interest in the current study, modality and the size of memory set. More specifically, it was of interest to determine the optimal amount of warning information that could be presented before the effects of stress affected performance on both the WCCOM task and the operational task. Secondly, the format the warnings were presented were also considered a task stress. Determining the effects that written, pictorial or verbal tasks had on the operational task was also considered.

The number of warnings presented was determined based on the work by Miller (1956) on the limitations of working memory. Miller's work on memory suggests that only 5-9 bits of information can successfully processed, retained, and recalled form working memory. Therefore, if the memory set was small to moderate, two or four warning combinations, then participants would be able to recall (comply with) the warnings. Yet, if eight warning-color combinations were presented, participants would no longer be able to comply as often as they did in the lower memory sets. Although all three memory sets (2, 4, and 8) are within Miller's "magical numbers" the operational task must be taken into consideration. Since the operational task will also tap working memory, it was predicted that the smaller memory sets (2 and 4) would result in higher compliance (recall) than the larger memory set (8 warning-color combinations).

As predicted, the size of the memory set affected compliance on the WCCOM task across all eight experiments. Results of the collapsed data revealed that as task demand increased, performance on the WCCOM task decreased. When the task-based stress was at level two (two warning-color combinations) participants could still cope with the stress and comply at a rate of

83% (out of 100%). Similarly, when the task-based stress was at level four (four warning-color combinations) compliance was still relatively high at a rate of 71%. Yet, compliance scores dropped off dramatically when the level of demand is at eight (eight warning-color combinations). Participants only complied at a rate of 52% when eight warnings were presented.

These results, as predicted can be described using the Hancock and Warm model of stress (1989). When the size of the memory set was low to moderate (2 or 4 warning-color combinations) participants performing the task could adapt to the task demand and thus, performance and workload had true associations. Yet, when task demand was at a high level (eight warning-color combinations) participants performing the task could no longer adapt to the task demand and dissociations or insensitivities occurred.

The size of the memory set (task demand) also had an effect on the operational tasks. Two separate analyses were conducted for the collapsed data. One set of analyses were conducted on the shooting task and the other on the navigation tasks since the performance measures differed. The analyses on the shooting task comprised Experiments 3, 4, and 5. Results of these analyses revealed that when the size of the memory set was two or four, compliance did not significantly differ, but performance was significantly lower when the size of the memory set was eight. Although statistical differences were found, a ceiling effect occurred resulting in performance scores that ranged from 94-97% of enemies killed.

The second set of analyses were conducted on the navigation task which included Experiments, 6, 7, and 8. The first measure of interest was number of waypoints reached. Here again as the size of the memory set increased performance decreased. No significant differences were found between two warning-color combinations and four, but differences were found between two and eight and four and eight. A ceiling effect also appeared in the navigation task, with the lowest performance score of 88%.

The second measure of interest for the navigation task was the time it took to complete the task. The results of the analysis yielded a significant effect for size of memory set, specifically as the demand increased the time it took to complete the task also increased. When the size of the memory set was four, it took 1.3 times longer to reach all four waypoints as compared to the time at level two. Similarly, when participants were to remember and recall eight warning-color combinations (level 8) it took them 1.7 times longer than it did at level two. Thus, the time it took nearly doubled when the size of the memory set increased from two warning-color combinations to eight.

The effect of the size of the memory set (task demand) was much greater on the navigation task than it was on the shooting task. The navigation task was designed to be more mentally complex than the shooting task. Navigating from waypoint one to waypoint two entails retrieving the relevant information from working memory in order to complete the task (Tversky, 2003). Participants had to visualize or construct a representation of the environment in their working memory to accomplish the navigation task. To fulfill the task objectives in the shooting task, participants had little strain on memory and no need to reconstruct the environment. Thus, it was expected that the navigation task would be affected by the task-based stress more so than the shooting task because more resources would be needed and thus depleted as the demand increased.

The effects that modality as a task based stress was also of interest in the current study. It was predicted that warning presentation would affect task demand due to format interference with the operational task. Previous research on modality, (Wickens, 1992; Wickens, Sandry, & Vidulich, 1983) has indicated that in general resources were utilized better when divided across modalities (auditory and visual stimuli, for example) rather than displayed via two auditory or two visual channels. In accordance with the multiple resource theory, it was hypothesized that the visual modality in which warnings were presented (written and pictorial), would have affected performance on the operational task because it was predominantly a visual and spatial task. Thus, it was hypothesized verbal warnings would not have interfered with the secondary visual-spatial task because the codes would not have competed for resources. Thus, it was hypothesized that since verbal warnings should not interfere with the operational task, they would result in the lowest subjective workload rating compared to pictorial or written warnings since they share the same working memory code.

The results of the current study do not support Wickens' theory. No format effect was found. Thus two spatial tasks, the simultaneous interaction of storing and processing the pictorial warnings and interacting with the operational task, did not result in degraded performance. Therefore, modality as a task based stress did not affect performance.

Subjective Workload Ratings

Hypotheses 5, 6, 7, and 8 were predictions made about the effects of task demand on subjective workload. The predictions for the previous section on task based stress focused on performance on the WCCOM task and the operational task. The subjective workload rating section hypotheses are similar to the task based hypotheses and follow the same theoretical underpinnings, but focused on the subjective rating of the task based stress.

Separate predictions were made regarding modality and the size of memory set on subjective workload. It was predicted that subjective workload measures for varying sizes of memory set would affect workload ratings such that workload scores for two warning presentations would be significantly lower compared to workload on four warning presentations. Additionally, it was predicted that subjective workload and size of memory set would be correlated in conditions when the number of warning presentations was two or four because the amount of demand and the subjective rating would be associates (amply amount of resources available). Workload measures for conditions with eight warning presentations would exceed the resources available and the size of memory set would not be associated with workload measures (dissociation or insensitivities will occur).

It was also predicted that warning presentation would affect task demand due to format interference with the operational task. Since verbal warnings should not interfere with the secondary visual-spatial task, it was hypothesized that verbal warnings would result in the lowest subjective workload rating compared to pictorial or written warnings since they share the same working memory code.

The Rating Scale Mental Effort scores were taken after each trial in order to observe the variations in the size of memory set. As mentioned previously, the data was collapsed across response mode to get a better look at the global effects of task demand. Results of the analyses revealed that task demand affected subjective workload. Results yielded that as the task demand increased, the subjective workload ratings also increased. The rating scale ranged from 0-150mm, 150mm being the highest workload rating. The results of the collapsed data yielded level two at a score of 27mm. Level four almost doubled the score at 43mm, and level eight more than doubled the score of level two with a score of 63mm.

Although these scores significantly differed from each other and increased as the demand increased, eight warnings did not exceed the resources available. No dissociations or insensitivities were found in the results of the current study. We predicted that when workload increased and went beyond the resources available, performance would either increase, or no change would occur (dissociations and insensitivities, “a” or “c”). Subsequently what happened in this study is, as the task demand increased workload also increased, yet compliance scores decreased. In relation to the Hancock and Warm model (1989), this is known as association (region “d” in the figure below).

Table 38:
Predictions of Workload and Performance Relationships based on the Hancock and Warm Extended-U Model (Oron-Gilad, Hancock, Stafford, & Szalma, in review)

Region	Workload	Performance	Relationship
(a)	⇧	⇧	Dissociation
(b)	No change	No change	Control
(c)	⇧	No change	Performance Insensitivity
(d)	⇧	⇩	Association
(e)	Plateau 100%	⇩	Workload Insensitivity
(f)	⇩	⇩	Dissociation

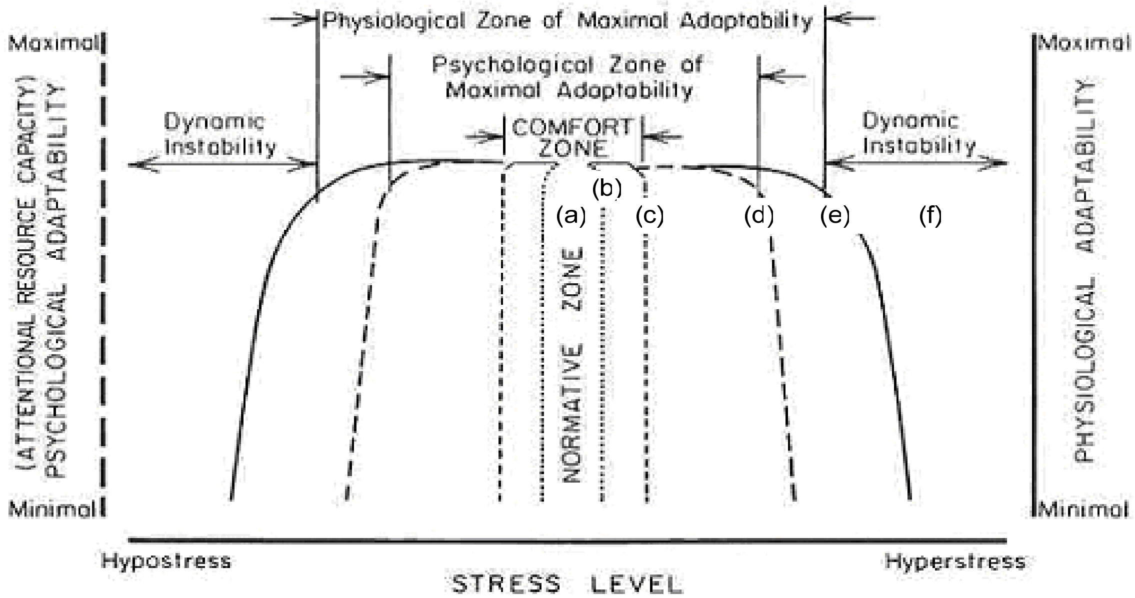


Figure 26: The Identification of the Regions of Hancock and Warm Model of Stress and Attention (1989).

Hypothesis 5 predicted that when the format of warning presentation was verbal it would have less interference on the operational task. In order to measure subjective workload for modality, the NASA-TLX was administered following the last trial in the block to get an overall rating of modality. The NASA-TLX was used to score the workload measures between the three formats of warning presentation; pictorial, written, and verbal. Contrary to hypothesis 5, no differences were found between warning formats for the collapsed data. Thus, participants did not feel that they had to exert more effort for warnings presented in written and pictorial compared to verbal warnings. Looking back at the scores from the individual experiments, it can also be seen that seven out of eight experiments found no differences in format (see table 33).

Table 39
Results of the NASA-TLX over all Eight Experiments

Experiment #	Number of Participants	Results of NASA-TLX	Post Hoc Results
1	6	ns	NA
2	6	ns	NA
3	11	ns	NA
4	11	ns	NA
5	12	ns	NA
6	12	ns	NA
7	12	Significant effect	Pictorials > Written or Verbal format
8	12	ns	NA

Although the results of the individual experiments (7 out of 8) and collapsed data analyses reveal that workload was not significant across modality, this may be an artifact of the design. Although the intent was for the NASA-TLX to identify differences between modality, the response mode variations may have been an experimental confound. Therefore, the results of this measure are not a true representation of presentation workload, but a combination of presentation and response variation.

In order to identify if any differences did exist between presentation formats, further analysis were conducted. Presentation and response mode were divided so that all combinations could be analyzed, thus nine combinations were compared (Verbal Presentation-Pictorial Response, Pictorial-Pictorial, Written-pictorial, Verbal-Written, Pictorial-Written, Written-Written, Verbal-Verbal, Verbal-Pictorial, and Verbal-Written). It was also of interest to investigate the effects of the operational tasks on workload. The results of this analysis revealed a main effect for response format, but significant differences were found for presentation or operational task. Thus, even though the navigation task was designed to be more cognitively complex, participants did not subjectively rate it as needing more mental effort.

The results of presentation response also yielded no differences between formats. Yet, response mode did have significant effects. Differences were found between written response mode and pictorials and verbal format. No differences were found between pictorial and verbal response mode. No explanation is currently available for why the differences were found between response formats, especially considering pictorial and verbal responses yielded no differences. Future work in response mode is needed.

Working Memory

Hypotheses 9-12 were predicted in order to determine if individual differences in working memory played an influential role in the warning format that resulted in the highest behavioral compliance. It was predicted that a) individuals low in both verbal and spatial working memory abilities would yield non-significant differences between warning presentation/format types; b) individuals high in both verbal and spatial working memory abilities would yield non-significant differences between warning presentation/format types; c) individuals high in verbal and low in spatial working memory abilities would perform significantly better in the auditory and written condition than in the pictorial condition; d) individuals high in spatial and low in verbal working memory abilities would perform significantly better in the pictorial condition than in the auditory or written condition.

Firstly, a correlation analysis on the four working memory spans, which were adopted from Shah and Miyake's (1996) experiment, were performed to determine the separability of spatial and working memory. If working memory has a separate pool of resources for verbal and spatial memory, then reading span (Daneman & Carpenter, 1980), a measure of functional working memory capacity for language, should correlate with verification word span (Shah & Miyake, 1996), a second measure of language processing that also included a processing and storage component. Furthermore, these two spans should correlate because the processing and storage involved in both spans are of the same modality, verbal. Additionally, the spatial span task (Shah & Miyake, 1996), a spatial measure of functional working memory capacity should not correlate with either of the language processing tasks because they each tap separate working memory resources, spatial and verbal. The third measure, the verification arrow task (Shah & Miyake, 1996), used separate modes for storage and processing, thus the processing portion of the task entails language processing and the storage component taps spatial thinking. As Shah and Miyake found (1996), this task should correlate with the spatial span and not the verbal span.

Consistent with what was predicted, the reading span and the verification word span were moderately correlated. These results support the separability hypotheses, in that the demands that the verbal spans were driven by the simultaneous demand that each span task imposed on the processing and storage of same modality information (verbal). To further support the separability hypothesis, the spatial span and the reading span were not correlates. This study also replicated the Shah and Miyake's (1996) findings that the verification arrow task was correlated with the spatial span task. Significant correlations were also found between verification arrow and the verification word, this result may be due to the processing component of both tasks which was both verbal. These results replicate those of Shah and Miyake (1996) supporting the separability hypothesis and the examination of the importance of the modality of the processing and storage component of the span task.

Yet, it was the interest of this study to determine not only are the correlations between spans in line with the separability hypothesis, but if the spans are predictive of presentation format. It was hypothesized that scores on the reading span and the verification word spans would predict how well participants did on the varying formats of the WCCOM task. It was predicted that participants high in language processing would do better in the WCCOM tasks when warnings were presented in verbal or written formats. It was also hypothesized high scores on the spatial span task would result in better performance on the pictorial warning presentations in the WCCOM task. Finally, it was predicted that the verification arrow task, because it used dual modes for processing would correlate with the warnings that matched that pattern, verbal processing and spatial storage. Yet, the results of the correlations between working memory spans and presentation type were not consistent with predictions. The pattern of results from the correlation between working memory spans and warnings collapsed by presentation (verbal, pictorial, and written) were not as predicted. The three warning formats did not correlate with the predicted spans; this could be an artifact of the presentation/response combination, which was mixed across formats. Thus, a closer look at the presentation/response format was in need.

Since the aforementioned results may be due to the mixed combination of the presentation and response format, correlations were conducted between the span measures and the nine warning presentation/response formats. Unfortunately, no consistent pattern was found for the presentation/response combinations. Since there are nine combinations, it is easiest to pinpoint the warning combinations that matched, such as the verbal/verbal, written/written, and pictorial/pictorial to determine if they are correlated with the predicted span. It was hypothesized that the verbal spans would be predictive of tasks that involved language processing, such as warnings of verbal presentation/verbal response, but the correlations were not significant between the verbal/verbal warning format and the verification word task. Similarly, the written presentation, written response combination was not correlated with any span measure. Furthermore, the pictorial presentation, pictorial response warning was significantly correlated with all but the verification arrow task. In light of these findings, it is impossible to predict the format of presentation/response that would result in the highest compliance behavior, because no pattern emerges. Even when the presentation and response format match, the pattern of correlations is not consistent. Thus based on this evidence the separability of working memory processing when applied was not supported.

Communication-Human Information Processing

A warning message has to be processed at each stage of this model in order to produce the desired compliance behavior. Bottlenecks may occur at any given stage in this model, thus leading to the possibility of non-compliance. In order to have a comprehensive view of the factors that effect compliance behavior, this line of research set out to determine if stress affected compliance behavior. In addition, it was of interest in the current study to determine if bottlenecks in processing were due to individual difference variables (receiver) and/or the format (channel) in which the warning information was presented.

Based on the results of this line of research, bottlenecks may due to the format (channel) of the warning information or the task demand. Channel bottleneck may be due to the format of presentation and response. Thus, it was found that in order to reduce the bottlenecks in this stage of processing, the presentation and response mode should match formats. Secondly, task demand may also cause a bottleneck and reduce the rate of compliance. Task demand can be seen from two perspectives in this model, it can be a form of stress that affects the receiver causing bottlenecks and from different perspective as a memory component. As a memory component, results of this line of research revealed that as task demand increases, the amount of information that must be processed and stored in memory increases, compliance behavior decreases. Thus, the amount of warning information that is presented must be regulated in order to decrease the amount of bottlenecks at this stage of processing. The C-HIP figure below (Figure 25) highlights the areas that bottlenecks may occur if not monitored and designed to suit the receiver.

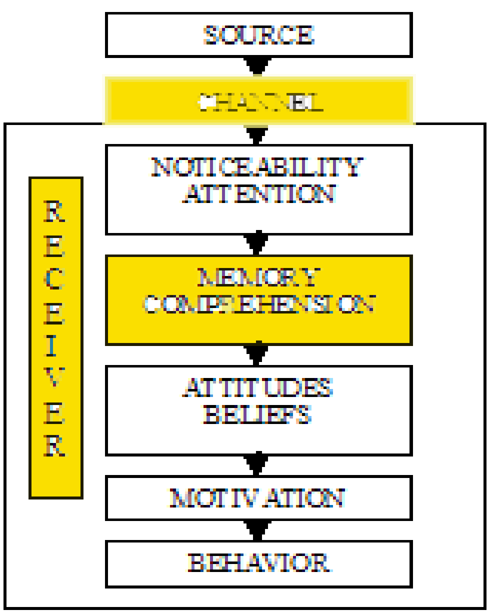


Figure 27: The C-HIP model with the stages of processing that may incur bottlenecks.

PRACTICAL IMPLICATIONS

The results of this line of research on warning compliance and the effects of stress and memory are applicable to a wide range of domains. Firstly, the warnings arena can benefit from this experimentation in a fashion of ways. The collapsed data on warning compliance for the WCCOM task revealed that it is not only the format of the warning presentation, but the combination of the presentation and response format. Additionally, when the matched presentation and response mode were analyzed, auditory superiority effect was not found. Incidentally, pictorial presentation/ pictorial response combination resulted in a compliance rate of 82%, which is the highest rate of behavioral compliance across warning formats. Verbal presentation coupled with verbal response yielded a compliance rate of 77%. Finally, written presentation/written response combination resulted in a compliance rate of 69%.

Therefore, the current and future applications of warnings need to consider the matching of warning presentation and response mode in order to get the highest rate of behavioral compliance. Contrary to previous literature on warning format, the matching of presentations response format suggests that pictorials result in the highest behavioral compliance across formats. Thus, pictorials may be coded both verbally and specially, thus increasing the chances that they are remembered over the auditory or written warnings.

This series of experimentation also revealed that stress does affect warning compliance. As the task demand increases, compliance decreases. In tasks that are more complex, stress affects performance at an earlier rate. Thus, complex tasks are more sensitive to the amount of warning presented than simple tasks. As the number increased from two to four warnings degradations were found in the complex task, but the simple task was robust to the task demand until it increased to eight warning presentations.

Warnings are presented in many real world environments that should consider the warning design recommendations such as cockpit design, automobile dashboard design, industrial work environments, and consumer products, etc. The aforementioned examples can all benefit from recommendations because they involve, a) a single or dual task environment, b) an environment that may be affected by stress, and c) environments where single, or multiple warnings may be presented.

Design Recommendations:

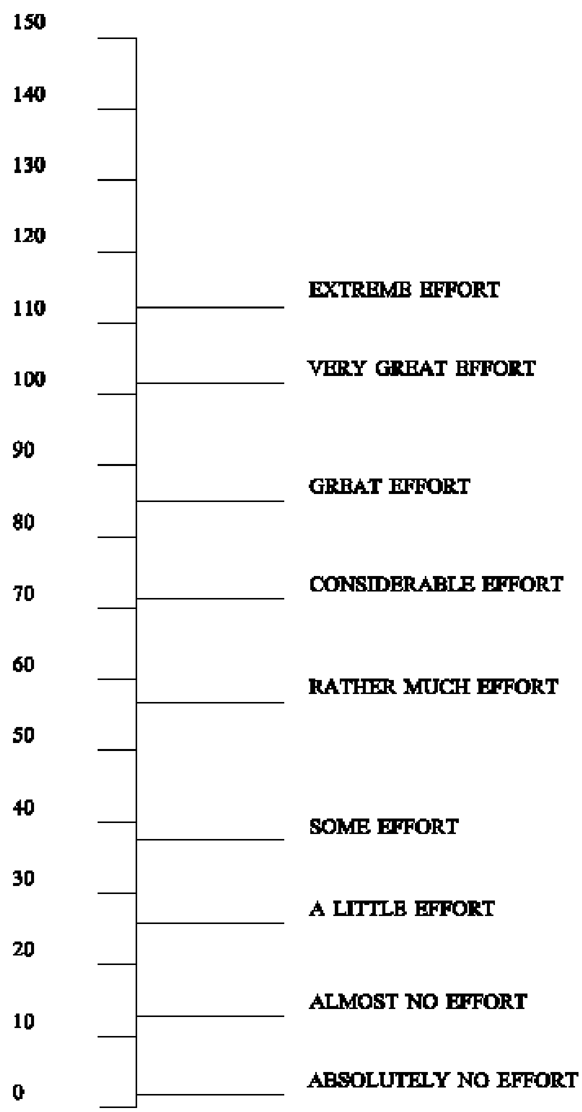
1. Consider the warning presentation and response format.
2. Utilize pictorials warnings; they may be coded both verbally and spatially.
3. Refine the amount of warnings that are presented to four in simple tasks and less than four in complex tasks.

APPENDIX A

RATING SCALE MENTAL EFFORT

Rating Scale Mental Effort

Please indicate, by marking the vertical axis below, how much effort it took for you to complete the task you've just finished



APPENDIX B

NASA TASK LOAD INDEX

Part I Instructions

Rating Scales. We are not only interested in assessing your performance but also the experiences you had during the experiment. In the most general sense, we are examining the "workload" you experienced. Since workload is something that is experienced individually by each person, there are no set "rulers" that can be used to estimate the workload associated with different activities. One way to find out about workload is to ask people to describe the feelings they experienced while performing a task. The set of six rating scales that I will give you was developed for you to use in evaluating your experiences during this task. Please read the descriptions of the scales carefully. If you have any questions about any of the scales in the table, please ask me about them.

For each of the six scales, you will evaluate the task by typing in a multiple of 5 that can range from 0 to 100 to reflect the point that matches your experience. Pay close attention to each scale's endpoint descriptions when making your assessments. Please note that when the rating scale for PERFORMANCE appears, a low score means you think you did well, while a high score means that you think you did poorly.

Upon completing each scale, use the mouse to click on the "Next" button to go on to the next scale. Read the description for each scale again before making your rating.

Part 1: Rating Scale

MENTAL DEMAND

LOW = 0 ------------------- 100 = HIGH

PHYSICAL DEMAND

LOW = 0 ------------------- 100 = HIGH

TEMPORAL DEMAND

LOW = 0 ------------------- 100 = HIGH

PERFORMANCE

POOR = 0 ------------------- 100 = GOOD

EFFORT

LOW = 0 ------------------- 100 = HIGH

FRUSTRATION LEVEL

LOW = 0 ------------------- 100 = HIGH

Part II Instructions

Pairwise Comparisons. Rating scales of this sort are extremely useful, but their usefulness is limited by the tendency people have to interpret them in different ways. People differ in which scales they think were the most important contributors to workload for a task.

The next step in your evaluation is to assess the relative importance of the six factors in determining how much workload you experienced. You will be presented with pairs of rating scale titles (e.g. EFFORT vs. MENTAL DEMAND) and asked to choose which of the two items was more important to your experience of workload in the task that you just performed.

Please consider your choices carefully and try to make them consistent with your scale ratings. Refer back to the rating scale definitions if you need to as you proceed. There is no correct pattern of responses. We are only interested in your opinions.

Do you have any questions?

Part II: Paired Comparisons

Please choose the more important contributor to workload:
1 = FRUSTRATION
2 = TEMPORAL DEMAND

Please choose the more important contributor to workload:
1 = EFFORT
2 = PHYSICAL DEMAND

Please choose the more important contributor to workload:
1 = EFFORT
2 = PERFORMANCE

Please choose the more important contributor to workload:
1 = PERFORMANCE
2 = PHYSICAL DEMAND

Please choose the more important contributor to workload:
1 = EFFORT
2 = MENTAL DEMAND

Please choose the more important contributor to workload:
1 = PERFORMANCE

2 = TEMPORAL DEMAND

Please choose the more important contributor to workload:
1 = TEMPORAL DEMAND
2 = PHYSICAL DEMAND

Please choose the more important contributor to workload:
1 = FRUSTRATION
2 = PERFORMANCE

Please choose the more important contributor to workload:
1 = MENTAL DEMAND
2 = FRUSTRATION

Please choose the more important contributor to workload:
1 = PHYSICAL DEMAND
2 = FRUSTRATION

Please choose the more important contributor to workload:
1 = EFFORT
2 = FRUSTRATION

Please choose the more important contributor to workload:
1 = MENTAL DEMAND
2 = TEMPORAL DEMAND

Please choose the more important contributor to workload:
1 = MENTAL DEMAND
2 = PHYSICAL DEMAND
Please choose the more important contributor to workload:
1 = TEMPORAL DEMAND
2 = EFFORT

Please choose the more important contributor to workload:
1 = MENTAL DEMAND
2 = PERFORMANCE

RATING SCALE DEFINITIONS

TITLE	ENDPOINTS	DESCRIPTIONS
MENTAL DEMAND	LOW/HIGH	How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
PHYSICAL DEMAND	LOW/HIGH	How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	LOW/HIGH	How much time pressure did you feel due to the rate or pace at which the task or parts of the task occurred? was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	LOW/HIGH	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals? **Note**: A **low** number means you thought you did **well**, while a **high** rating means you think you did **poorly**.
EFFORT	LOW/HIGH	How hard did you have to work (mentally and/or physically) to accomplish your level of performance?
FRUSTRATION LEVEL	LOW/HIGH	How insecure, discouraged, irritated stressed, and annoyed versus secure, gratified, content, relaxed, and complacent did you feel during the task?

APPENDIX C

TABLES

Experiment 1 Tables

Table 40
WCCOM: Means and Standard Deviations for Warning Format for Experiment 1

Format	Mean	Std. Error
Verbal	.526	.054
Written	.721	.056
Pictorial	.733	.061

Table 41
WCCOM: Means and Standard Deviations for Task Demand for Experiment 1

Task Demand	Mean	Std. Error
2	.789	.035
4	.697	.070
8	.494	.062

Table 42
Ghost Recon Shooting Task: Means and Standard Deviations for Warning Format for Experiment 1

Format	Mean	Std. Error
Verbal	.938	.020
Written	.948	.018
Pictorial	.942	.016

Table 43
Ghost Recon Shooting Task: Means and Standard Deviations for Task Demand for Experiment 1

Task Demand	Mean	Std. Error
2	.946	.021
4	.962	.007
8	.920	.024

Table 44
The ANOVA Table for the NASA-TLX for Experiment 1

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	26.395	.545	.596	.098	.116
Error(Format)	10	48.460				

a Computed using alpha = .05

Table 45
NASA-TLX scores for Format for Experiment 1

Format	Mean	Std. Deviation
Pictorial	52.9783	19.65507
Verbal	57.1117	14.31790
Written	55.6650	18.72939

Table 46
RSME: Means and Standard Deviations for Warning Format for Experiment 1

Format	Mean	Std. Error
Verbal	25.756	9.785
Pictorial	42.322	9.369
Written	59.200	7.389

Table 47
RSME: Means and Standard Deviations for Task Demand for Experiment 1

Task Demand	Mean	Std. Error
2	44.111	9.871
4	40.189	7.133
8	42.978	9.672

Experiment 2 Tables

Table 48
WCCOM: Means and Standard Deviations for Warning Format for Experiment 2

Format	Mean	Std. Error
Verbal	.549	.042
Written	.729	.068
Pictorial	.719	.073

Table 49
WCCOM: Means and Standard Deviations for Task Demand for Experiment 2

Task Demand	Mean	Std. Error
1	.806	.023
2	.678	.079
3	.514	.069

Table 50
The ANOVA Table for the Ghost Recon Shooting Task for Experiment 2

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.018	.854	.454	.146	.158
Error(Format)	10	.022				
Task Demand	2	.094	2.791	.109	.358	.427
Error(Task Demand)	10	.034				
Trial	4	.028	1.465	.250	.227	.371
Error(Trial)	20	.019				
Format * Task Demand	4	.023	.865	.502	.147	.226
Error(Format*Task Demand)	20	.027				
Format * Trial	8	.024	1.975	.075	.283	.732
Error(Format*Trial)	40	.012				
Task Demand * Trial	8	.013	.807	.600	.139	.319
Error(Task Demand*Trial)	40	.016				
Format * Task Demand * Trial	16	.019	1.358	.185	.214	.782
Error(Format * Task Demand * Trial)	80	.014				

a Computed using alpha = .05

Table 51
Ghost Recon Shooting Task: Means and Standard Deviations for Warning Format for Experiment 2

Format	Mean	Std. Error
Verbal	.942	.012
Written	.941	.016
Pictorial	.917	.032

Table 52
Ghost Recon Shooting Task: Means and Standard Deviations for Task Demand for Experiment 2

Task Demand	Mean	Std. Error
2	.955	.010
4	.949	.012
8	.896	.038

Table 53
The ANOVA Table for the NASA-TLX for Experiment 2

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	52.723	.328	.728	.062	.089
Error(Format)	10	160.553				

a Computed using alpha = .05

Table 54
NASA-TLX Means and Standard Deviations for Format for Experiment 2

Format	Mean	Std. Deviation
Pictorial	59.1083	21.01518
Verbal	58.5517	13.17975
Written	63.9417	16.03658

Table 55
RSME: Means and Standard Deviations for Warning Format for Experiment 2

Format	Mean	Std. Error
Verbal	39.172	8.440
Pictorial	36.811	7.325
Written	34.967	5.929

Table 56
RSME: Means and Standard Deviations for Task Demand for Experiment 2

Task Demand	Mean	Std. Error
2	11.967	4.607
4	34.900	7.818
8	64.083	11.153

Experiment 3 Tables

Table 57
WCCOM: Means and Standard Deviations for Warning Format for Experiment 3

Format	Mean	Std. Error
Verbal	.506	.040
Written	.630	.057
Pictorial	.735	.050

Table 58
WCCOM: Means and Standard Deviations for Task Demand for Experiment 3

Task Demand	Mean	Std. Error
2	.727	.043
4	.691	.061
8	.453	.030

Table 59
Ghost Recon Shooting Task: Means and Standard Deviations for Format for Experiment 3

Format	Mean	Std. Error
Verbal	.930	.018
Written	.948	.009
Pictorial	.939	.013

Table 60
Ghost Recon Shooting Task: Means and Standard Deviations for Task Demand for Experiment 3

Task Demand	Mean	Std. Error
2	.956	.010
4	.932	.012
8	.929	.017

Table 61
The ANOVA Table for the NASA-TLX for Experiment 3

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	35.564	.379	.691	.045	.101
Error(Format)	16	93.904				

a Computed using alpha = .05

Table 62
NASA-TLX Means and Standard Deviations for Warning Format for Experiment 3

Format	Mean	Std. Deviation
Pictorial	57.9422	18.28458
Verbal	56.4478	17.72854
Written	60.3856	14.26214

Table 63
RSME: Means and Standard Deviations for Warning Format for Experiment 3

Format	Mean	Std. Error
Verbal	41.188	4.714
Pictorial	46.821	5.032
Written	55.388	7.891

Table 64
RSME: Means and Standard Deviations for Task Demand for Experiment 3

Task Demand	Mean	Std. Error
2	50.300	4.638
4	50.115	6.121
8	42.982	4.276

Experiment 4 Tables

Table 65
WCCOM; Means and Standard Deviations for Warning Format for Experiment 4

Format	Mean	Std. Error
Verbal	.460	.030
Pictorial	.537	.053
Written	.660	.036

Table 66
WCCOM: Means and Standard Deviations for Task Demand for Experiment 4

Task Demand	Mean	Std. Error
2	.718	.028
4	.532	.035
8	.406	.040

Table 67
Ghost Recon: Means and Standard Deviations for Warning Format for Experiment 4

Format	Mean	Std. Error
Verbal	.957	.012
Written	.949	.015
Pictorial	.940	.016

Table 68
Ghost Recon: Means and Standard Deviations for Task Demand for Experiment 4

Task Demand	Mean	Std. Error
2	.960	.013
4	.961	.006
8	.924	.018

Table 69
The ANOVA Table for the NASA-TLX for Experiment 4

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	93.547	.824	.455	.084	.169
Error(Format)	18	113.531				

a Computed using alpha = .05

Table 70
NASA-TLX Means and Standard Deviations for Warning Format for Experiment 4

Format	Mean	Std. Deviation
Pictorial	62.9270	12.93258
Verbal	61.7320	17.16727
Written	57.1340	7.60052

Table 71
RSME: Means and Standard Deviations for Warning Format for Experiment 4

Format	Mean	Std. Error
Verbal	30.956	5.145
Pictorial	51.930	4.660
Written	67.320	3.378

Table 72
RSME: Means and Standard Deviations for Task Demand For Experiment 4

Task Demand	Mean	Std. Error
2	51.033	5.206
4	53.191	4.529
8	45.982	4.096

Experiment 5 Tables

Table 73
WCCOM: Means and Standard Deviations for Task Demand for Experiment 5

Task Demand	Mean	Std. Error
2	.932	.019
4	.763	.049
8	.617	.049

Table 74
WCCOM: Means and Standard Deviations for Trial for Experiment 5

Trial	Mean	Std. Error
1	.704	.048
2	.811	.041
3	.806	.036
4	.761	.043
5	.771	.039

Table 75
The ANOVA Table for the Ghost Recon Shooting Task for Experiment 5

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.007	.689	.513	.059	.151
Error(Format)	22	.010				
Task Demand	2	.021	2.932	.074	.210	.513
Error(Task Demand)	22	.007				
Trial	4	.016	1.623	.185	.129	.458
Error(Trial)	44	.010				
Format * Task Demand	4	.015	1.194	.327	.098	.343
Error(Format*Task Demand)	44	.013				
Format * Trial	8	.007	.502	.852	.044	.220
Error(Format*Trial)	88	.013				
Task Demand * Trial	8	.007	.496	.856	.043	.217
Error(Task Demand*Trial)	88	.013				
Format * Task Demand * Trial	16	.013	1.005	.454	.084	.661
Error(Format * Task Demand * Trial)	176	.013				

a Computed using alpha = .05

Table 76
Ghost Recon Shooting Task: Means and Standard Deviations for Warning Format for Experiment 5

Format	Mean	Std. Error
Verbal	.979	.005
Written	.979	.008
Pictorial	.961	.008

Table 77
Ghost Recon Shooting Task: Means and Standard Deviations for Task Demand for Experiment 5

Task Demand	Mean	Std. Error
2	.979	.008
4	.967	.007
8	.973	.008

Table 78
The ANOVA Table for the NASA-TLX for Experiment 5

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	136.111	1.098	.351	.091	.218
Error(Format)	22	123.990				

a Computed using alpha = .05

Table 79
NASA-TLX Means and Standard Deviations for Format for Experiment 5

Format	Mean	Std. Deviation
Pictorial	33.3333	32.00379
Verbal	30.8333	23.91589
Written	26.6667	22.69695

Table 80
RSME: Means and Standard Deviations for Warning Format for Experiment 5

Format	Mean	Std. Error
Verbal	40.167	3.033
Written	39.722	3.222
Pictorial	46.383	6.371

Table 81
RSME: Means and Standard Deviations for Task Demand for Experiment 5

Task Demand	Mean	Std. Error
2	23.156	3.662
4	41.176	4.246
8	61.940	3.433

Experiment 6 Tables

Table 82
WCCOM: Means and Standard Deviations for Warning Format for Experiment 6

Format	Mean	Std. Error
Verbal	.583	.032
Pictorial	.791	.032
Written	.816	.023

Table 83
WCCOM: Means and Standard Deviations for Task Demand for Experiment 6

Task Demand	Mean	Std. Error
2	.822	.018
4	.819	.031
8	.549	.032

Table 84
Ghost Recon: Means and Standard Deviations for Warning Format for Experiment 6

Format	Mean	Std. Error
Verbal	.398	.102
Pictorial	.361	.108
Written	.395	.088

Table 85
Ghost Recon: Means and Standard Deviations for Task Demand for Experiment 6

Task Demand	Mean	Std. Error
2	.259	.071
4	.388	.089
8	.507	.097

Table 86
The ANOVA Table for the NASA-TLX for Experiment 6

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	118.153	2.874	.080	.223	.499
Error(NASA)	20	41.105				

a Computed using alpha = .05

Table 87
NASA-TLX Means and Standard Deviations for Format for Experiment 6

Format	Mean	Std. Deviation
Pictorial	68.2136	10.47082
Verbal	71.7582	10.36093
Written	65.2109	12.94398

Table 88
RSME: Means and Standard Deviations for Warning Format for Experiment 6

Format	Mean	Std. Error
Verbal	46.561	2.006
Pictorial	44.833	3.370
Written	47.028	3.784

Table 89
RSME: Means and Standard Deviations for Task Demand for Experiment 6

Task Demand	Mean	Std. Error
2	26.156	3.190
4	42.750	2.052
8	69.517	3.565

Experiment 7 Tables

Table 90
WCCOM: Means and Standard Deviations for Warning Format for Experiment 7

Format	Mean	Std. Error
Verbal	.471	.038
Pictorial	.692	.031
Written	.708	.040

Table 91
WCCOM: Means and Standard Deviations for Task Demand for Experiment 7

Task Demand	Mean	Std. Error
2	.778	.029
4	.650	.053
8	.443	.029

Table 92
The ANOVA Table for the Ghost Recon Navigation Task for Experiment 7

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	.569	.389	.682	.034	.105
Error(Format)	22	1.462				
Task Demand	2	.588	1.343	.282	.109	.259
Error(Task Demand)	22	.438				
Trial	4	.193	.527	.717	.046	.164
Error(Trial)	44	.366				
Format * Task Demand	4	.640	.892	.477	.075	.260
Error(Format*Task Demand)	44	.717				
Format * Trial	8	.188	.353	.942	.031	.161
Error(Format*Trial)	88	.532				
Task Demand * Trial	8	.530	1.245	.283	.102	.541
Error(Task Demand*Trial)	88	.426				
Format * Task Demand * Trial	16	.457	1.068	.389	.089	.695
Error(Format * Task Demand * Trial)	176	.428				

a Computed using alpha = .05

Table 93
Ghost Recon Navigation Task: Means and Standard Deviations for Warning format for Experiment 7

Format	Mean	Std. Error
Verbal	.462	.111
Pictorials	.430	.101
Written	.353	.074

Table 94
Ghost Recon Navigation Task: Means and Standard Deviations for Task Demand for Experiment 7

Task Demand	Mean	Std. Error
2	.368	.065
4	.397	.083
8	.479	.073

Table 95
NASA-TLX: The Means and Standard Deviations for Warning Format for Experiment 7

Format	Mean	Std. Error
Pictorial	62.727	2.771
Verbal	50.128	5.654
Written	50.950	3.026

Table 96
RSME: Means and Standard Deviations for Warning Format for Experiment 7

Format	Mean	Std. Error
Verbal	44.728	4.599
Pictorial	37.422	5.338
Written	42.533	4.468

Table 97
RSME: Means and Standard Deviations for Task Demand for Experiment 7

Task Demand	Mean	Std. Error
2	23.183	4.038
4	39.972	5.188
8	61.528	6.545

Experiment 8 Tables

Table 98
WCCOM: Means and Standard Deviations for Warning Format for Experiment 8

Format	Mean	Std. Error
Verbal	.774	.055
Pictorial	.826	.043
Written	.781	.038

Table 99
WCCOM: Means and Standard Deviations for Task Demand for Experiment 8

Task Demand	Mean	Std. Error
2	.975	.013
4	.791	.058
8	.615	.058

Table 100
Ghost Recon: Means and Standard Deviations for Warning Format for Experiment 8

Format	Mean	Std. Error
Verbal	.398	.111
Pictorial	.446	.069
Written	.409	.090

Table 101
Ghost Recon: Means and Standard Deviations for Task Demand for Experiment 8

Task Demand	Mean	Std. Error
2	.307	.073
4	.407	.095
8	.539	.094

Table 102
The ANOVA Table for the NASA-TLX Scores for Experiment 8

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Format	2	131.797	.667	.523	.057	.148
Error(NASA)	22	197.648				

a Computed using alpha = .05

Table 103
NASA-TLX Means and Standard Deviations for Warning Format for Experiment 8

Format	Mean	Std. Deviation
Pictorial	52.4492	21.10024
Verbal	51.2500	11.82617
Written	57.4950	22.48053

Table 104
RSME: Means and Standard Deviations for Warning Format for Experiment 8

Format	Mean	Std. Error
Verbal	36.640	2.750
Pictorial	40.194	4.399
Written	38.537	4.350

Table 105
RSME: Means and Standard Deviations for Task Demand for Experiment 8

Task Demand	Mean	Std. Error
2	19.906	2.837
4	36.793	3.723
8	58.674	6.263

Collapsed Data Tables

Table 106
WCCOM: Means and Standard Deviations for Response Format for the Collapsed Data

Response Format	Mean	Std. Error
Pictorial	.679	.026
Written	.589	.023
Verbal	.789	.028

Table 107
WCCOM: Means and Standard Deviations for Presentation Format for the Collapsed Data

Presentation Format	Mean	Std. Error
Verbal	.595	.012
Pictorial	.716	.023
Written	.745	.017

Table 108
WCCOM: Mean and Standard Deviation for Task Demand for the Collapsed Data

Task Demand	Mean	Std. Error
2	.826	.012
4	.711	.019
8	.520	.017

Table 109
Means and Standard Deviations for the Interaction between Response Format and Presentation Format

Presentation Format	Response Format	Mean	Std. Error
Verbal	Pictorial	.546	.026
Pictorial	Pictorial	.713	.036
Written	Pictorial	.777	.027
Verbal	Written	.465	.024
Pictorial	Written	.617	.034
Written	Written	.685	.027
Verbal	Verbal	.774	.034
Pictorial	Verbal	.819	.029
Written	Verbal	.773	.033

Table 110
Means and Standard Deviations for the Interaction between Presentation and Task Demand

Presentation Format	Task Demand	Mean	Std. Error
Verbal	2	.719	.013
Verbal	4	.646	.018
Verbal	8	.421	.013
Pictorial	2	.851	.019
Pictorial	4	.727	.030
Pictorial	8	.571	.027
Written	2	.907	.021
Written	4	.760	.020
Written	8	.567	.021

Table 111
Means and Standard deviations for the Interaction between Response format and Task Demand.

Response Format	Task Demand	Mean	Std. Error
Pictorial	2	.776	.024
Pictorial	4	.757	.036
Pictorial	8	.503	.024
Written	2	.748	.021
Written	4	.594	.034
Written	8	.425	.024
Verbal	2	.953	.013
Verbal	4	.782	.039
Verbal	8	.631	.038

Table 112
WCCOM: Mean and Standard Deviation for Task Demand for the Collapsed Data

Task Demand	Mean	Std. Error
2	.965	.006
4	.958	.006
8	.939	.009

Table 113
Ghost Recon: Means and Standard Deviations for Task Demand for the Collapsed Data

Task Demand	Mean	Std. Error
2	.311	.040
4	.398	.050
8	.508	.050

Table 114
The ANOVA Table for the NASA-TLX for the Collapsed Data

Source	df	Mean Square	F	Sig.	Partial Eta Squared	Observed Power(a)
Presentation Format	2	1353.270	2.383	.107	.117	.450
Error(PRES)	36	567.975				
Response Format	2	83.701	1.024	.369	.054	.215
Error(RESP)	36	81.720				
Presentation * Response	4	391.936	3.957	.006	.180	.887
Error(Presentation * Response)	72	99.051				

a Computed using alpha = .05

Table 115
RSME: Means and Standard Deviations for Task Demand for the Collapsed Data

Task Demand	Mean	Std. Error
2	.271	.156
4	.431	.165
8	.625	.168

Table 116
Correlation Tables for the RSME and NASA-TLX for Experiment 4

	NASA-TLX Pictorial Rating	NASA-TLX Written Rating	NASA-TLX Verbal Rating	RSME Pictorial Rating	RSME Written Rating	RSME Verbal Rating
NASA-TLX Pictorial Rating	1					
NASA-TLX Written Rating	.422	1				
NASA-TLX Verbal Rating	1.000(**)	.422	1			
RSME Pictorial Rating	.124	-.396	.124	1		
RSME Written Rating	.273	-.163	.273	.690(*)	1	
RSME Verbal Rating	.341	-.172	.341	.481	.850(**)	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table 117
Correlation Tables for the RSME and NASA-TLX for Experiment 5

	NASA-TLX Pictorial Rating	NASA-TLX Written Rating	NASA-TLX Verbal Rating	RSME Pictorial Rating	RSME Written Rating	RSME Verbal Rating
NASA-TLX Pictorial Rating	1					
NASA-TLX Written Rating	.422	1				
NASA-TLX Verbal Rating	1.000(**)	.422	1			
RSME Pictorial Rating	.124	-.396	.124	1		
RSME Written Rating	.273	-.163	.273	.690(*)	1	
RSME Verbal Rating	.341	-.172	.341	.481	.850(**)	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table 118
Correlation Tables for the RSME and NASA-TLX for Experiment 6

	NASA-TLX Pictorial Rating	NASA-TLX Written Rating	NASA-TLX Verbal Rating	RSME Pictorial Rating	RSME Written Rating	RSME Verbal Rating
NASA-TLX Pictorial Rating	1					
NASA-TLX Written Rating	.422	1				
NASA-TLX Verbal Rating	1.000(**)	.422	1			
RSME Pictorial Rating	.124	-.396	.124	1		
RSME Written Rating	.273	-.163	.273	.690(*)	1	
RSME Verbal Rating	.341	-.172	.341	.481	.850(**)	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Table 119
Correlation Tables for the RSME and NASA-TLX for Experiment 7

	NASA-TLX Pictorial Rating	NASA-TLX Written Rating	NASA-TLX Verbal Rating	RSME Pictorial Rating	RSME Written Rating	RSME Verbal Rating
NASA-TLX Pictorial Rating	1					
NASA-TLX Written Rating	.749(**)	1				
NASA-TLX Verbal Rating	.500	.527	1			
RSME Pictorial Rating	.309	.073	.163	1		
RSME Written Rating	.280	.275	.205	.793(**)	1	
RSME Verbal Rating	.322	.265	.315	.829(**)	.796(**)	1

** Correlation is significant at the 0.01 level (2-tailed).

Table 120
Correlation Tables for the RSME and NASA-TLX for Experiment 8

	NASA-TLX Pictorial Rating	NASA-TLX Written Rating	NASA-TLX Verbal Rating	RSME Pictorial Rating	RSME Written Rating	RSME Verbal Rating
NASA-TLX Pictorial Rating	1					
NASA-TLX Written Rating	.749(**)	1				
NASA-TLX Verbal Rating	.500	.527	1			
RSME Pictorial Rating	.309	.073	.163	1		
RSME Written Rating	.280	.275	.205	.793(**)	1	
RSME Verbal Rating	.322	.265	.315	.829(**)	.796(**)	1

** Correlation is significant at the 0.01 level (2-tailed).

REFERENCES

American National Standards Institute (ANSI). (1991). Product Safety Signs and Labels. ANSI Z535.4-1991. New York: ANSI.

AS 1319-1979. (1979). Standards Association of Australia*, Rules for the Design and Use of Safety Signs for the Occupational Environment*, SAA, Sydney.

Ayers, T.J., Gross, M.M., Wood, C.T., Horst, D.P., Beyer, R.R., & Robinson, J.N. (1989). In: *Proceedings of the Human factors Society 34th Annual Meeting* (pp. 426-430). Santa Monica, CA: Human Factors Society.

Baddeley, A.D. (1986). *Working Memory*. Oxford University Press : Oxford.

Baddeley, A.D. (1992). *Working memory*. Science, 255, 556-559.

Baddeley, A. (1993). Working memory and conscious awareness. In: A.F. Colins, S.E. Gathercole, M.A. Conway, P.E. Morris (Eds.), *Theories of memory,* pp. 11–28. Hove: Erlbaum.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology*, 49(a), 5-28.

Baddeley, A., & Hitch, G.J., (1974). Working memory. In: G. Brower (Ed.), *Recent advances in learning and motivation*, 8, pp. 47-90. New York: Academic Press.

Barnett, B.J. & Wickens, C.D. (1988). Displaying proximity in multicue information integration: The benefits of boxes. *Human Factors*, 30(1), 15-24.

Beatty, W. W., & Troster, A. I. (1987). Gender differences in geographical knowledge. Sex Roles, 16, 565-590.

Boersema, T. & Zwaga, H.J.G. (1989). Selecting comprehensible warning symbols for swimming pool slides. In: *Proceedings of the Human Factors Society 33rd Annual Meeting*, pp. 994-998. Santa Monica, CA: Human Factors Society.

Brainerd, C.J., & Kingma, J. (1985). On the independence of short-term memory and working memory in cognitive development. *Cognitive Psychology*, 17(2), 210-247.

Braun C.C., Silver, N.C.,& Stock, B.R. (1992). Likelihood of reading warnings: The effect of font and font sizes. *Proceedings of the Human Factors Society 36th Annual Meeting*. Santa Monica, CA: Human Factors Society, 926-930.

Braun, C.C., & Silver, N.C. (1999). Behavior. *Warnings and Risk Communication*, 245-262.

Broadbent, D.E. (1963). Flow of information within the organism. *Journal of Verbal Learning and Verbal Behavior*, 2, 34-39.

Broadbent, D., Vines, R., & Broadbent, M. (1978). Recency effects in memory, as a function of modality of intervening events. *Psychological Research*, 40(1), 5-13.

Cairney, P., & Sless, D. (1982). Communication effectiveness of symbolic safety signs with different user groups. *Applied Ergonomics*, 13(2), 91-97.

Cannon, W.B. (1915). *Bodily changes in pain, hunger, fear, and rage; an account of recent research into the function of emotional excitement*. Birmingham, Classics of Psychiatry.

Chen, J.Y.C. (2000). *Noise effects of low-criticality warnings*. Unpublished doctoral dissertation, University of Central Florida, Orlando.

Clark, D.M. (1983). On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures. *Advances in Behavior Research & Therapy*, 5(1), 27-49.

Collins, B.L. (1983). *Use of Hazard Pictorials/Symbols in the Mineral Industry*. Washington, D.C.: National Bureau of Standards, NBSIR 83-2732.

Collins, B.L. & Lerner, N.D. (1982). Assessment of fire-safety symbols. *Human Factors*, 24(1), 75-84.

Collins, B.L., Lerner, N.D. and Pierman, B.C. (1982). Symbols for Industrial Safety. National Bureau of Standards, Washington, D.C., 1-102, NSBIR 82-2485.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, 104(2), 163-191.

Cowan, E.W. (1995). The client-therapist encounter: A comparative analysis of intersubjectivity and dialogical theory. *Dissertation Abstracts International: Section B: The Sciences & Engineering*, 56(6-B), 3440.

Craik, F.I. (1969). Modality effects in short-term storage. *Journal of Verbal Learning & Verbal Behavior*, 8(5), 658-664.

Daneman, M., & Carpenter, P.A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning & Verbal Behavior*, 19(4), 450-466.

Daneman, M., & Tardiff, T. (1987). Working memory and reading skill reexamined. In: M Coltheart (Eds.), *Attention and Performance XII: The psychology of reading,* pp. 491-508. Hillsdale, NJ: Erlbaum.

Davies, D.R., Parasuraman, R. (1982). *The psychology of vigilance*. London: Academic Press.

Desaulniers, D.R. (1987). Layout, organization and the effectiveness of consumer product warning. In: *Proceedings of the Human Factors and Ergonomic Society 31st Annual Meeting*. Santa Monica, CA, pp. 56-60.

Dorris, A.L, Purswell, J.L. (1978). Human factors in the design of effective product warnings. In: *Proceedings of the Human Factors Society 22nd Annual Meeting*, pp. 343-346. Santa Monica, C.A: Human Factors Society.

Driskell, J.E., & Olmstead, B. (1989). Psychology and the military: Research applications and trends. *American Psychologist*, 44(1), 43-54.

Duffy, R, Kalsher, M.J., & Wogalter, M.S. (1995) Interactive warning: An experimental examination of effectiveness. *International Journal of Industrial Ergonomics*, 15,159-166.

Easterby, R., & Zwaga, H. (1984). *Information Design: The Design and Evaluation of Signs and Printed Material*. New York: Wiley.

Engle, R.W., Cantor, J., & Carullo, J.J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 972-992.

Engle, R.W., Kane, M.J., & Tuholski, S.W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In: A. Miyake, P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control*, pp. 102-134. New York: Cambridge University Press.

Frances, W.N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin Company.

Friedmann, K. (1988). The effect of adding symbols to written warning labels on user behavior and recall. *Human Factors*, 30, 507-515.

Gardiner, J. M., Passmore, C., Herriot, P., & Klee, H. (1977). Memory for remembered events: Effects of response mode and response-produced feedback. *Journal of Verbal Learning and Verbal Behavior, 16,* 45-54.

Gardiner, J.M., Thompson, C.P., & Maskarinec, A.S. (1974). Negative recency in initial free recall. *Journal of Experimental Psychology*, 103(1), 71-78.

Goodie, A.S., Crooks, C.L. (2004). Time-pressure effects on performance in a base-rate task. *Journal of General Psychology*, 131(1), 18-28.

Hancock, P.A., & Ganey, H.C.N. (2003). From the inverted-U to the extended-U: The evolution of a law of psychology. *Human Performance in Extreme Environments, 7* (1) 5-14.

Hancock, P.A., Ganey, H.C.N & Szalma, J.L. (2002). Performance under stress: A reevaluation of a fundamental law of psychology. In: *Proceedings of the 23rd Annual Army Science Conference*: Orlando, FL.

Hancock, P.A., Ward, P., Szalma, J.L., Stafford, S., & Ganey, H.C.N. (2002). Stress and human information processing: A descriptive framework presented in a novel manner. Paper presented at the 23rd Army Science Conference, Orlando, FL, December, 2002.

Hancock, P.A. & Warm, J.S. (1989). A dynamic model of stress and sustained attention. *Human Factors*, 31(5), 519-537.

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

Hiburn, B., & Jorna, P.G.A.M. (2001). Workload and air traffic control. In: P.A. Hancock & P.A. Desmond (Eds.), *Stress, workload, and fatigue*, pp. 384-394. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Hockey, R. (1978). Attentional selectivity and the problems of replication: A reply to Forster & Grierson. *British Journal of Psychology*, 69(4), 499-503.

Hockey, R. & Hamilton, P. (1983). The cognitive patterning of stress states. In: G.R.J. Hockey (Ed.), *Stress and Fatigue in Human Performance*. New York: John Wiley & Sons, pp. 331-362.

Jaynes, L.S., & Boles, D.B. (1990). The effects of symbols on warning compliance. In: *Proceedings of the Human factors Society 35th Annual Meeting.* Santa Monica, CA: Human Factors Society, pp. 984-987.

Kahneman, D. (1973). *Attention and Effort*. Englewood Cliffs, NJ: Prentice Hall.

Kanki, B.G. (1996). Stress and aircrew performance: A team-level perspective. In: J.E. Driskell, E. Salas (Eds.), *Stress and human performance*, pp. 127-162. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Kieras, D.E., Meyer, D.E., Mueller, S., Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In: A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control,* pp. 183-223. New York: Cambridge University Press.

Laughery, K.R., & Hammond, A. (1999). Overview. In: M.S. Wogalter, D.M. DeJoy & K.R. Laughery (Eds.). *Warnings and Risk Communication,* pp. 3-13.

Laux, L.F. & Brelsford, J.W. (1989). Locus of control, risk perception, and precautionary behavior. *In Proceedings of Interface*, 89, 121-124.

Lasswell, H.D. (1948**)**. The structure and function of communication in society. In: L. Bryson Ed., *Communication of Ideas*. New York: Harper.

Lehto, M.R., & Miller, J.M. (1986). Warnings: Fundamentals, Design, and Evaluation Methodologies. *Fuller Technical Publications*, Ann Arbor, Michigan.

Leonard, S.D., Matthews, D., & Karnes, E. W. (1986). How does the population interpret warning signals? Proceedings of the Human Factors Society 34th Annual Meeting. Santa Monica, CA: Human factors and Ergonomic Society, 984-987.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. Child Development, 56, 1479-1498.

Magurno, A.B., & Wogalter, M.S. (1994). Behavioral compliance with warnings: effects of stress and placement. In: *Proceedings of the Human factors Society 38th Annual Meeting,* pp. 826-830. Santa Monica, CA: Human Factors Society.

Mayer, R.E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist*, 32(1), 1-19.

McDonald, D.P. (2001). *Perceptaul and semantic responses to multiple alarms*. Unpublished doctoral dissertation, University of Central Florida, Orlando.

Messick, Huet, M., & Wickens, C.D. (Eds.). (1993). *Workload Transition: Implications for Individual and Team Performance*. Washington, D.C.: National Academy Press.

Miller, G.A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 63, 81-97.

Millisecond Software (2002). Inquisit, Version 1.32. Seattle, WA: Millisecond Software.

Murdock, B.B. (1968). Modality effects in short-term memory: storage or retrieval? *Journal of Experimental Psychology*, 77(1), 79-86.

Norman, D.A., & Bobrow, D.B. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*, 44-64.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review*, 86(3), 214-255.

O'Donnell, R.D., & Eggemeir, F.T. (1986). Workload assessment methodology. In: K.R. Boff, L. Kaufman, & J.P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 4/1-42/49). New York: Wiley.

O'Reilly, R.C., Braver, T.S., & Cohen, J.D. (1999). A biologically based computational model of working memory. In: A. Miyake, P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control,* pp. 375-411. New York: Cambridge University Press.

Oron-Gilad, T., Hancock P.A., Stafford S.C. and Szalma J.L. (submitted 2005).On the Relationship between Workload and Performance, International Journal of Applied Psychology.

Otsubo, S.M. (1988). A behavioral study of warning labels for consumer products: perceived danger and use of pictograms. In: *Proceedings of the Human Factors Society 32nd Annual Meeting,* pp. 536-540. Santa Monica, CA: Human Factors Society.

Paivio, A., Rogers, T.B., & Smythe, P.C. (1968). Why are pictures easier to recall than words? *Psychonomic Science*, 11(4), 137-138.

Parasuraman, R., & Hancock, P.A. (2001). Adaptive control of mental workload. In: P.A. Hancock & P.A. Desmond (Eds.), *Stress, workload, and fatigue*, pp. 305-320. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Penney, C.G. (1975). Modality effects in short-term verbal memory. *Psychological Bulletin*, 82(1), 68-84.

Penney, C.G., & Butt, A.K. (1986). Within- and between-modality associations in probed recall: A test of the separate-streams hypothesis. *Canadian Journal of Psychology*, 40(1), 1-11.

Peters, R. (1984). Nucleo-cytoplasmic flux and intracellular mobility in single hepatocytes measured by fluorescence microphotolysis. *EMBO J.* 3:1831-1836.

Reason, J.T. (1990) *Human Error.* Cambridge University Press.

Rogers, W.A., Lamson, N., & Rousseau, G.K. (2000). Warning research: An integrative perspective. *Human Factors*, 42(1), 102-139.

Rogers, W.A., Rousseau, G.K., Lamson, N. (1999). Maximizing the effectiveness of the warning process: Understanding the variables that interact with age. In: D.C. Park & R.W. Morrell (Eds.), *Processing of Medical Information in Aging Patients: Cognitive and Human Factors Perspectives*, pp. 267-290. Mahwah, New Jersey: Lawrence Erlbaum Associates.

Rousseau, G.K, Lamson, N., Rogers, W.A. (1998). Designing warnings to compensate for age-related changes in perceptual and cognitive abilities. *Psychology and Marketing*, 15(7), 643-662.

Selye, H. (1956). *The Stress of Life*. New York, NY: McGraw-Hill.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual difference approach. *Journal of Experimental Psychology: General, 125*, 4-27.

Shannon, C.E., & Weaver, W. (1949). *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Silver, N.C., & Wogalter, M.S. (1989). Broadening the range of signal words. *Proceedings of the Human Factors Society 33rd Annual Meeting*. Santa Monica, CA: Human Factors Society, 555-559.

Spielberger, C., Gorusch, R., & Lushene, R. (1970). *State-Trait Anxiety Inventory: Test Manual for Form X.* Palo Alto, California: Consulting Psychologist Press.

Standing, L., Conezio, J., & Haber, R.N. (1970). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, 19(2), 73-74.

Stanton, N.A., (1994). Alarm initiated activities. In: N. A. Stanton (Ed.), *Human factors in alarm design*. London: Taylor & Francis.

Strawbridge, J.A. (1986). The influence of position, highlighting, and embedding on warning effectiveness. In: *Proceedings of the Human Factors Society 30th Annual Meeting,* pp. 716-720. Santa Monica, CA: Human Factors Society.

Szalma, J.L., Warm, J.S., Matthews, G., Dember, W.N., Weiler, E.M., Meier, A., & Eggemeier, F.T. (2004). Effects of sensory modality and task duration on performance, workload, and stress in sustained attention, *Human Factors*, *46(2),* 219-233.

Tabachnick, B.G. & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Turner, M.L., & Engle, R.W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, 28(2), 127-154.

Turvey, R. (1969). Marginal Cost. *Economic Journal*, 79(314), 282-299.

Tversky, B. (2003). Navigating by mind and by body. In: C. Freksa, W. Brauer, C. Habel, and K.F. Wender (Eds.), *Spatial Cognition III. Routes and navigation, human memory and learning, spatial representation and spatial learning,* pp. 1-10. Berlin: Springer.

Twerski, A.D., Weinstein, A.S., Donaher, W.A., & Piehler, H.R. (1976). The use and abuse of warnings in product liability: Design defect litigation comes of age. *Cornell Law Review*, 61(4), 495-540.

Ursic, M. (1984). The impact of safety warnings on perception and memory. *Human Factors*, 26(6), 677-682.

Warm, J.S. (1984). *Sustained Attention in Human Performance*. New York: John Wiley.

Warm, J.S., Dember, W.N., Gluckman, J.P., & Hancock, P.A. (1991). Vigilance and workload. *Proceedings of the Human Factors Society*, *35*, 980-981.

Warm, J.S., Dember, W.N., & Hancock, P.A. (1996). Vigilance and workload in automated systems. In: R. Parasuraman and M. Mouloua (Eds.), *Automation and human performance.* (pp. 183-200). Mahwah, NJ: Erlbaum.

Watkins, O.C., & Watkins, M.J. (1980). The modality effect and echoic persistence. *Journal of Experimental Psychology: General*, 109, 251-278.

Weaver, J.L., Gerber, T. N., Hancock, P.A., & Ganey, H.C.N. (2003). Individual differences in behavioral compliance to warnings representing varying degrees. *Journal of Safety and Ergonomics*, 9 (2), 149-160.

Weinstein, A.K. (1977). Foreign investments by service firms: The case of the multinational advertising agencies. *Journal of International Business Studies*, Spring/Summer, 83-91.

Wickens, C.D. (1980). The structure of attentional resources. In: R.S. Nickerson (Ed.), *Attention and performance VIII* (pp. 239-257). Erlbaum. Hillsdale, N.J.

Wickens, C.D. (1984). Processing resources in attention. In: R. Parasuraman, R. Davies (Eds.), *Varieties of Attention*. Orlando, FL: Academic Press.

Wickens, C.D. (1992). *Engineering psychology and human performance (2nd ed.).* New York: HarperCollins Publishers.

Wickens, C.D. & Carswell, C.M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37(3), 473-494.

Wickens, C.D., & Holland, J.G. (2000). *Engineering psychology and human performance (3rd ed.)*. Upper Saddle River, NJ: Prentice Hall.

Wickens, C.D., Sandry, D.L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25(2), 227-248.

Wogalter, M. (1994). Factors influencing the effectiveness of warnings, *Public Graphics*, Lunteren, The Netherlands.

Wogalter, M.S., Allison, S.T., & McKenna, N.A., (1989). Effects of cost and social influence on warning compliance. *Human Factors,* 31, 133-140.

Wogalter, M.S., Dejoy, D.M., & Laugherty, K.R. (1999). Organizing Theoretical Framework: A consolidated communication-human information processing (C-HIP) model. In: M.S. Wogalter, D.M. Dejoy & K.R. Laughery (Eds.), *Warnings and Risk Communication*, pp. 15-23. London: Taylor and Frances Ltd.

Wogalter, M.S., Fontenelle, G.A., & Laughery, K.R. (1985). Behavioral effectiveness of warnings. *Human Factors Perspectives on Warnings*, 679-683.

Wogalter, M.S., Godfrey, S.S., Fontenelle, G.A., Desaulniers, D.R., Rothstein, P.R., & Laughery, K.R. (1987). Effectiveness of warnings. *Human Factors*, 29, 599-612.

Wogalter, M.S., Magurno A.B., Rashid R., & Klein K.W. (1998). The influence of time stress and location on behavioral warning compliance. *Safety Science, 29 (2),* 143-158.

Wogalter, M.S. & Usher, M.O. (1999). Effects of concurrent cognitive task loading on warning compliance behavior. In: *Proceedings of the Human factors Society 43rd Annual Meeting,* pp. 525-529. Santa Monica, CA: Human Factors Society.

Wogalter, M.S., Rashid, R., Clarke, S.W., & Kalsher, M. J. (1991). Evaluating the behavioral effectiveness of multi-modal voice warning sign in a visually cluttered environment. In: *Proceedings of the Human factors Society 36th Annual Meeting,* pp. 718-722. Santa Monica, CA: Human Factors Society.

Wogalter, M.S., & Silver, N.C. (1990). Arousal strength of signal words. *Forensic Reports*, 3, 407-420.

Wogalter, M.S., & Silver, N.C. (1995). Warning signal words: Connoted strength and understandability by children, elders, and non-native English speakers. *Ergonomics, 38*, 2188-2206.

Wogalter, M.S., & Young, S.L. (1991). Behavioral compliance to voice and print warnings.

Ergonomics, 34 (1) p. 79-89.

Wolff, J.S., & Wogalter, M. S. (1993). Test and development of pharmaceutical pictorials. In: *Proceedings of the Interface '93,* pp. 187-192. Santa Monica, CA: Human Factors Society.

Yeh, Y. Y., & Wickens, C. D. (1988). Dissociations of performance and subjective measures of workload. *Human Factors*, *30*, 111-120.

Yerkes, R.M., & Dodson, J.D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology & Psychology*, 18, 459-482.

Young, S.L., & Wogalter, M.S. (1990). Comprehension and memory pf instruction manual warnings: Conspicuous print and pictorial icons. *Human Factors*, 32(6), 637-649.

Young, S.L., & Wogalter, M.S. (1991). Memory of instruction manual warnings: Effects of pictorial icons and conspicuous print. In: *Proceedings of the Human factors Society 36th Annual Meeting,* pp. 984-987. Santa Monica, CA: Human Factors Society.

Zijlstra, F. R. H., & Van Doorn, L.(1985). *The construction of a scale to measure perceived effort.* Delf, The Netherlands, Department of Philosophy and Social Science, Delf University of Technology.