Applied Ergonomics 45 (2014) 1343–1352

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Coding warnings without interfering with dismounted soldiers' missions

Tal Oron-Gilad [a,*], P.A. Hancock [b], Jessica Helmick-Rich [b]

[a] Dept. of Industrial Engineering and Management, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel
[b] Department of Psychology, University of Central Florida, USA

ARTICLE INFO

Article history:
Received 25 July 2012
Accepted 19 August 2013

Keywords:
Military
Warning
Retention
Task demand

ABSTRACT

Objectives: Warnings are an effective way to communicate hazard, yet they can also increase task demand when presented to operators involved in real-world tasks. Furthermore, in military-related tasks warnings are often given in codes to avoid counter-intelligence, which may foster additional working memory load.
Background: Adherence to warnings in the military domain is crucial to promote safety and reduce accidents and injuries. The empirical question arises as to how aspects of coding the warning may interfere with the primary task the individual is currently performing and vice versa.
Method: Six experimental conditions were designed to assess how warning-code storage format, response format, and increasing working memory demand (retention) affected both performance on the primary task and the rate of compliance to warnings, considered here as the secondary task.
Results: Results revealed that the combination of warning-code storage and response format affected compliance rate and the highest compliance occurred when warnings were presented as pictorials and responses were coded verbally. Contrary to the proposed hypotheses, warning storage format did not affect performance on the primary task, which was only affected by the level of working memory demand. Thus, the intra-modal warning storages did not interfere with the visual/spatial nature of the primary operational task. However, increase in working memory demand, by increasing the number of memorized warning codes, had an effect on both compliance rate and primary task performance.
Conclusions: Rather than warning code storage alone, it is the coupling of warning storage and response format that has the most significant effect on compliance.

1. Coding warnings without interfering

I have a strong memory from the "Yom-Kipur" war of the radio broadcaster repeating the code "viper-snake" "viper-snake". This combination of words indicated that an alarm was about to be heard and that we should head down to the bomb shelter in fear of an air raid (first author, personal memory from an early age).

Indeed, in military operations, warnings are often pre-coded and pre-memorized. Thus, it is not the warning itself but a code which represents it, that is given in real operational conditions. The use of codes is not a conventional way to use warnings. Commonly in workplaces, warnings are explicitly given and intentionally emphasized to promote safe behavior (e.g., Wogalter, 1994; Wogalter et al., 1992). Nevertheless, in operational settings, when the operator is outside, in various locations, oftentimes in hostile areas and with some degree of secrecy – implicit hints must be given instead of the traditionally known warning presentation. It is in this realm that we wish to examine how distribution of warnings is most efficiently conveyed.

Nowadays, field operators as dismounted soldiers are equipped with information technologies such as mobile phones and small scale computing devices, which makes it easier to convey to them short messages in more than one possible way (e.g., by voice, by text, or by pictorials). Under those circumstances, warnings are most likely to be given while the individual is engaged in performing an operational mission. In the military operation realm, the mission could be consisting of surveillance, recon or conflict-solving components. The dynamics of such missions may change rapidly and little is known to the distributer of the warning, who is remote from the field operator, about the particular state of the current task. Furthermore, since there is no direct contact between the receiver of the information and the transmitter, an acknowledgment of compliance may also be needed. In such cases, storage

* Corresponding author. Tel.: +972 8 6472227; fax: +972 8 6472958.
E-mail address: orontal@bgu.ac.il (T. Oron-Gilad).

0003-6870/$ – see front matter © 2013 Elsevier Ltd and The Ergonomics Society. All rights reserved.
http://dx.doi.org/10.1016/j.apergo.2013.08.011

and processing of pre coded warnings may generate conditions that become formally equivalent to dual-task performance. It is already known that distraction and disruption to operational tasks may occur when the operator experiences high cognitive load and task criticality (Woods, 1995). The question arises therefore, as how the storage and retention of coded warning, as well as their mode of acknowledgment, act to interfere with the primary mission that the individual is currently performing and vice versa, where the latter affects warning compliance rate.

It is known in the literature that coding pose more challenges for the operator than being given explicit instructions. To communicate information effectively one aims to create direct associations between the sign and its referent (Familant and Detweiler, 1993), which enables information in the signal to be directly mapped to a certain referent in the world. For example, the word "fire" shouted in a theater (signal), has the potential of being interpreted as a condition manifested by flames (referent). Direct associations are those involving one referent (e.g., a picture of protective glasses denotes that one should wear protective glasses). For those, the denotative referent is also the sign referent (i.e., the referent to which the signal directly refers). Indirect associations, in contrast, are those involving at least two referents. That is, a signal and a denotative referent may be linked via a chain of referent relations or via two or more parallel referent relations. Whether a warning is accurately identified is likely to depend on several factors describing the extent to which the signal resembles (i.e., shares features with) those that the operator has learned to associate with the referent.

Unfortunately with military-based coding, the strength of the relationships between signs and referents are *intentionally* poor to avoid hostile forces from deciphering them. Studies that have looked at learning and retention of unrelated associations and their effect on recognizing auditory icons (Keller and Stevens, 2004) and their denotative referents (Stephan et al., 2006) have consistently shown that relative to direct and related associations, participants in unrelated conditions consistently demonstrated poor performance, made the most errors, and responded most slowly. Furthermore, Keller and Stevens (2004) found that random relations took over three times more exposure to learn than direct relations, and required over twice as much exposure to reach the same level of compliance. In military applied settings, where coding is frequent, such influences may have dramatic consequences.

Thus, aside from dual-task interference (e.g., Driskell and Olmstead, 1989; Kanki, 1996) between the appearance of a warning and the operational task, the use of coding may cause the compliance to the warning to be poor or erroneous, as deciphering the meaning of the message requires more processing. Furthermore, when codes are being used and the retention of warnings is necessary, additional load on working memory exists which may affect both mission performance and compliance. Thus, in addition to their specific form, the sheer number of retained warnings could affect performance by increasing the level of task demand imposed on the operator. Working memory capacity is limited and the amount of information that can be retained and recalled in real-world situations is similarly restricted (Miller, 1956; but see Ericsson and Staszewski, 1989). Thus, the number of remembered warnings might be reduced when working memory is divided between the operational task and the increasing spectrum of warning information. However, there are also reasons whereby the remembered warnings would not be reduced, when LTM schemata is used (e.g., if different warnings could be integrated in some way, or if cues were available to prompt the different warnings) but these may not necessarily be used the military context due to their increased notice ability by hostiles over time.

Previous research on warnings suggests that the mode of information presentation affects compliance (Lehto and Miller, 1986; Rogers et al., 2000; Wogalter and Usher, 1999; Ursic, 1984). For example, verbal information is remembered and recalled more often than written or pictorial information (Penney, 1975; Murdock, 1968; Watkins and Watkins, 1980) and verbal warnings have been found to communicate hazards better than pictorial or written formats (Jaynes and Boles, 1990; Wogalter and Young, 1991). Combined print and voice warnings have been shown to be more effective than either alone (e.g., Conzola and Wogalter, 1998). The visual presentation usually allows receivers to review a message for longer time or if it was not attended to or comprehended initially (Conzola and Wogalter, 2001). The auditory information may attune the operator toward the warning initially, but is not available for long duration (unless a repeat element is available). However, as noted by Conzola and Wogalter (2001) field experimentation on warnings in actual occupational settings is rather limited. Consequently, these results may well not apply in real-world circumstances where individuals most often receive warning information as a subsidiary form while focusing on their primary duties. Since the ability to expose real operators to hazards is limited, virtual reality is one promising direction (Duarte and Rebelo, 2007), to allow more realistic contextual setup of warnings.

Table 1
Overview of the experimental conditions.

Experimental condition	Compliance task storage and retention	Compliance task response mode	Operational task
1	Pictorial, Written and Verbal in all 6 conditions	Pictorial	Suspect detection
2		Written	Suspect detection
3		Verbal	Suspect detection
4		Pictorial	Navigation
5		Written	Navigation
6		Verbal	Navigation

Individual difference in WM can also affect the way warnings should be coded. Engle et al. (1999) proposed that WM capacity reflects one's ability to maintain information in the focus of attention in the face of distracting or interfering stimuli. "…*WM capacity is not about individual differences in how many items can be stored per se but about differences in the ability to control attention to maintain information in an active, quickly retrievable state*" (Engle, 2002, p.20). Furthermore, the relationship between working memory and cognitive abilities has been proposed to involve domain-specific resources (i.e., verbal and visuospatial; Baddeley, 1986). Consistently, Shah and Miyake (1996) have demonstrated the separability of spatial and verbal working memory resources using the spatial span test (recalling the orientation of a series of rotated letters presented sequentially). They showed that scores on the spatial span test correlate positively with spatial ability measures but not with verbal ability measures. Thus, individual difference in WM may have predictive power of the memory variables on situation awareness performance in complex task settings (see Sohn and Doane, 2003). This type of personality related information may be useful, with the flexibility of presentation format via personal mobile devices. However, very little is known about the impact of individual differences in complex operational task setups such as the one examined in the current study.

Therefore, the formal goal of the present study was to determine the effects of coding presentation and retention in a dual task paradigm with various task-induced levels of demand. Six experimental conditions were designed in order to investigate the effects of a) pairing storage formats with response modes (pictorial, written, and verbal storages with pictorial, written, verbal response modes); b) increasing the number of memorized coded warnings; c) altering the complexity and demand of the operational task; and d) evaluating the effect of individual differences in WM on compliance.

2. Experimental method

2.1. Overview

An overview of the six experimental conditions is given in Table 1. Conditions 1–3 used a simulated MOUT (military operation in urban terrain) suspect detection mission as the primary operational task. Conditions 4–6 used an open area way-point to waypoint navigation mission. These two missions were selected as they are both common tasks that dismounted soldiers do. In the suspect detection mission, soldiers are required to survey a building in search of reported suspects. This can be a demanding task, since the level of details necessary to attend to is large, and the environment consists of many paths and limited view areas. Opposed to buildings, navigation in open areas entails a clearer view of the environment and more certainty as to where dangers may appear from. Yet, navigating from waypoint to waypoint requires retrieving relevant information from working memory in order to complete the task (Tversky, 2003). Thus, navigation mission performance may be more susceptible to changes in WM demand. It was therefore assumed that differences in operational performance and compliance to warnings will emerge from the inherent differences between the two tasks.

The respective conditions included a warning storage and retention component in each one of three modalities of storage; pictorial, written, and verbal. What varied among conditions was the acknowledgment of the warning (i.e., response mode), which was pictorial (conditions 1 and 4), written (conditions 2 and 5), and verbal (conditions 3 and 6). Also, the number of retained warnings was manipulated in three levels (2, 4 or 8 coded warnings). Thus, altogether in each condition there were 45 trials; 15 for each storage format (pictorial, written and verbal), of those, 5 for each level of demand (2, 4, or 8). Specific hypotheses are presented in Table 2.

2.2. Experimental participants

Participants were undergraduate students from the University of Central Florida (UCF) and were recruited on a voluntary basis from the university's experimental recruiting website. They were paid approximately $37.50 for their participation (based on an hourly rate of $7.50). Eleven (mean age = 20 years) participated in Condition 1 and 2, respectively. Twelve participated in Condition 3 (mean age = 21 years), Condition 4 (mean age = 20.1 years), Condition 5 (mean age = 19.5 years) and Condition 6 (mean age = 22.3 years). All participants were treated in accordance to relevant APA standards and all procedures were pre-screened by the Internal Review board (IRB) before testing.

2.3. Experimental materials

The experimental system consisted of two separate tasks: a) the WCCOM (Warning-Color COding Modality) compliance task and b) The operational tasks developed in Tom Clancy's Ghost Recon® game simulation. Tasks were presented on two separate desktop computers with two monitors (17″ and 19″ flat screens), two keyboards, and the attached mouse controllers. The computer used for the WCCOM had two speakers and a microphone which were used to present and respond to auditory information when applicable. The two monitors were placed side-by-side in order for participants to easily view both monitors simultaneously, when required.

Table 2
Hypotheses by number and category.

No.	Category	Hypotheses
1	Warning Compliance/Format of Storage	It was hypothesized that participants would have a significantly higher rate of compliance behavior when coding and retrieval were verbal compared to written and pictorial formats because the WCCOM (Warning-Color COding Modality) information would have less interference with the operational task.
2	Warning Compliance/Format of Storage	It was predicted that compliance behavior would be significantly higher in the pictorial warning condition than in the written warning condition.
3	WM Demand	It was predicted that when the number of warnings retained was two or four, performance on neither the warning compliance task nor the operational task would be affected.
4	WM Demand	When eight warnings were retained both 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 with the operational tasks which are visual-spatial in nature.
6	Workload	Subjective workload estimates for two warning retained would be significantly lower compared to conditions with four warning.
7	Workload	Subjective workload and task demand would be associated (high workload with lower performance) in conditions where the number of warning storages was two or four.
8	Workload	Workload measures for conditions with more eight warning storages would exceed the resources available and dissociation or insensitivities will occur.
9	Individual differences in WM	Participants with high spatial span scores will benefit from pictorial coding while participants with higher verbal span scores will benefit from written and verbal coding.

2.4. WCCOM (Warning-Color COding Modality) compliance task

The WCCOM task had both storage and retention components. Each warning item was paired with one of ten possible colors. The storage component required memorizing the color associated with each warning symbol (e.g., boots – black). The retention component involved recalling the stored symbol from the color presented (e.g., black means boots). Both components of the task, the warning item and the color, were displayed in the same modality. There were three options of storage; pictorial, written or verbal as shown Fig. 1. The written format 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) and was paired with a written color (spelled out in the color of the pair) in the same size and font. The verbal WCCOM was presented via speakers.

The warning database consisted of ten different warnings: Boots, Earmuffs, Glasses, Gloves, Helmet, Shield, Suit, Respirator, Meter, and Mask. All ten warning messages were chosen from a pool of occupational warnings used to promote mandatory action in the ANSI standards (ANSI Z535.3, 1991) and the Australian Standard (AS 1319, 1979). Six of these symbols have been tested previously for ease of recognition by Cairney and Sless (1982).

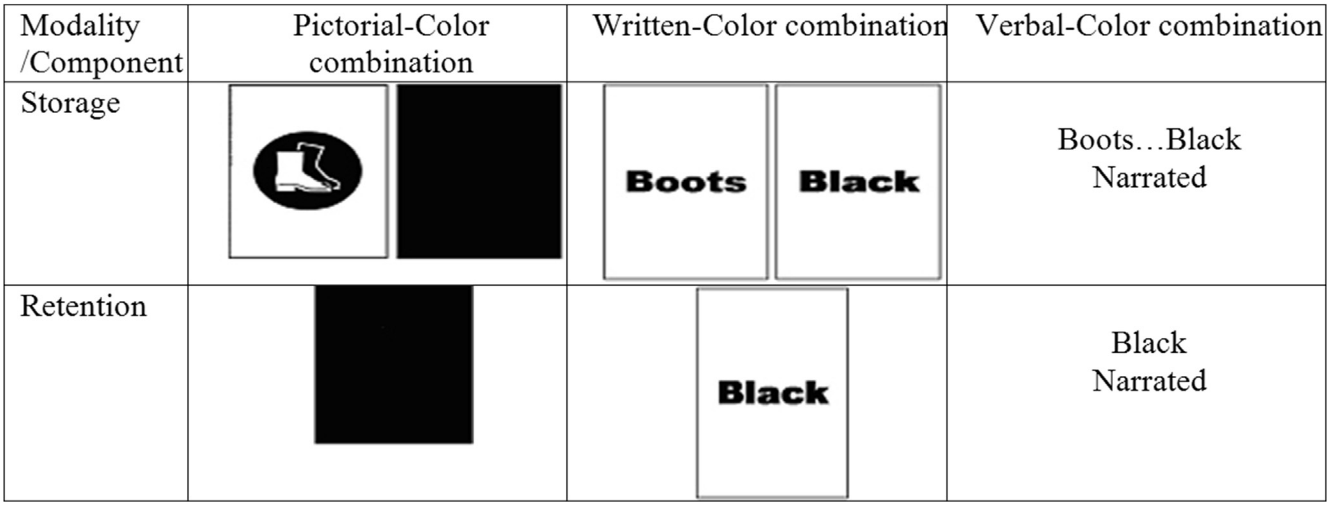

Fig. 1. Example of the pictorial, written and verbal WCCOM (top) and the color stimulus (bottom) that elicited the key press response.

Warnings were paired with one of ten colors: red, blue, green, orange, purple, black, white, gray, brown, or yellow. Colors were chosen from ANSI standards (ANSI Z535.2, 1991). Colors were pretested in order to determine if they were readily distinguishable on the specific monitor used.

Warning-color combinations were randomly selected for each trial; hence 'Boots' could be associated with the color black in one experimental trial and then be associated with the color purple on a subsequent block. A card-sorting task was used between trials to mitigate the retention of previous combinations. Working memory demand was manipulated by changing the number of memorized warnings sequentially from two to four to eight dyads, respectively. For the storage components, the pictorial and written warning-color combinations were presented for 5 s in the center of the computer screen. Then, after a brief pause the following combination appeared on the screen for the same duration of time. The verbal warning-color combination was presented verbally via speakers in the same manner that the written and pictorials were presented (yet verbal warnings could not be presented for the full 5 s and lasted just as long as it normally took to verbally pronounce the word). For all three storage modalities a short beep sounded preceding the WCCOM storage.

Participants' task was to remember the correct pairing of the warning and color combinations. When participants either saw or heard (depending on the modality of storage in that block) the color coded element, they were to respond by pressing the appropriately labeled key on a second keyboard (keys 'q' through 'p' were labeled with the warning portion of each combination) with their right hand or verbally respond via the microphone (e.g., in the pictorial response condition, response was to be made by pressing the key which had the appropriate warning pictorial on it).

2.5. MOUT Ghost Recon® suspect detection task

The operational military task was developed through use of Tom Clancy's Ghost Recon® which is a commercially available, first-person shooter video game produced by Redstorm Entertainment. 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 2-min mission in Ghost Recon while simultaneously responding to the WCCOM task.

The Ghost Recon suspect detection task took place inside a building. The objective in this task was to identify all suspects in the building and participants were informed that anyone in the building was an enemy (as shown in Fig. 2). In addition, participants were told not to leave the building for any reason. The enemies were strategically placed throughout the building and the number of enemies in any one building ranged from five to seven. Task difficulty did not vary from building to building. Participants were not aware of the number of possible enemies in the specific building and did not get feedback as to how many enemies they had identified compared to the total number of enemies in the building. Performance was measured by calculating the number of enemies that the participant identified divided by the number of enemies that appeared in the trial. Ten different missions were designed so that participants would not become over familiar with any one single mission. Missions were randomly assigned to the three blocks of trials.

2.6. Navigation Ghost Recon® task

The navigation task was developed within the Ghost Recon® environment. It took place in an open area of a sparsely wooded forest. The objective of the task was to navigate between four sequential waypoints. As shown in Fig. 3, a military tank marked each waypoint. There were four waypoints in each mission. Participants were to begin at the base denoted as waypoint 0 which was represented by the presence of a tank. The direction in which

Fig. 2. The Ghost Recon urban environment used in conditions 1–3.

Fig. 3. The Ghost Recon open area environment used in conditions 4–6.

the tank's gun pointed was the direction the participant was to navigate to in order to reach the next waypoint. No other navigational aids were provided. Once the participant was out of view of the waypoint, he or she had to use their spatial working memory to navigate to the next waypoint. When participants reached the fourth waypoint, the task was considered successfully completed. Seven different navigational missions were designed so that participants would again not become overly familiar with any one mission. The tanks were strategically placed throughout the rural terrain. Task difficulty did not vary from mission to mission. Three navigation missions were randomly assigned to the three blocks of trials.

Navigation performance was assessed using the time (in %) it took a participant to reach all four waypoints ((actual time-goal time)/goal time $\times$ 100). The goal time for a mission was the fastest possible time that the participant could have navigated through all four waypoints. In few instances, when participants skipped one of the waypoints on the way, 200% was added to the equation to avoid generating a time advantage to those who omitted any waypoint.

2.7. Individual differences – WM spans

The WM capacity theory proposes that individual differences in WM capacity are inherent in the individual and do not change as they acquire expertise. Individual differences in WM capacity are often measured with span tasks performed in a dual-task situation (e.g., Daneman and Carpenter, 1980; Shah and Miyake, 1996). Such tasks require participants to keep track of presented information while simultaneously doing other mental operations. The capacity measures, based primarily on the retention of the information over time, have been shown to predict performance in a variety of cognitive tasks, but to the best of our current understanding this has not been accomplished in warning-coding in the applied-military domain context examined here. Four working memory tasks were used in order to predict the processing and storage capacity of working memory of each individual. This set is a full replication of the four tests used by Shah and Miyake (1996), to test for; a) spatial working memory capacity; b) spatial span; c) verbal working memory capacities and d) reading span.

2.7.1. Spatial span

The spatial span task (Shah and 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. 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 ms in one of seven possible orientations in 45° increments, excluding the upright position. 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 (see Shah and 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.

2.7.2. Verification arrow

The verification arrow task (Shah and Miyake, 1996) consisted of reading short sentences (sentences ranged from three to six words) 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". One example of a short sentence used 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 ms in one of seven possible orientations in 45° increments, excluding the upright position. 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. 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.

2.7.3. Reading span

The reading span task (Daneman and Carpenter, 1980) was analogous to the spatial 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.

2.7.4. Verification word task

The verification word task (Shah and Miyake, 1996) was analogous to the verification arrow task. Again, participants were to decide if a sentence was true or false 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 ms. 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 were 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 Francis and Kučera (1982). Of the 275 most frequently used words, 70 two-syllable nouns (excluding proper nouns) were selected and each one was only used once in the task.

2.7.5. Subjective workload measures

Questionnaires were administered via the experimental software program Inquisit Version 1.32 (Millisecond Software, 2002) on a desktop computer.

2.7.6. Rating scale mental effort

The RSME is a one-dimensional scale that measures the amount of invested effort exerted during a task (Zijlstra and Van Doorn, 1985). The scale's range is from 0 to 150 mm and a hash mark is placed at every 10 mm. Anchor points were identified at several locations on the scale, describing the mental effort invested, such as ‘almost no effort’ or ‘extreme effort’. Mental effort is measured by the number that is identified as the invested mental effort for a given task from 0 to 150. The higher the score, the more subjective mental effort was exerted. RSME was used between trials in order to determine differences between sizes of memory set (task demand).

2.7.7. NASA Task Load Index

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 (Hart and Staveland, 1988).

2.8. Experimental procedure

A mixed between-within-participants design was used. Each participant was assigned to one experimental condition. Each condition consisted of three blocks (one for each modality of storage and retention: verbal, written, and pictorial) and each block consisted of 15 WCCOM task trials; 5 repetitions for each level of demand; coding and retention of 2, 4, and 8 warnings, respectively. The order of the modality conditions was counterbalanced between blocks. Additionally, the order of trials within each block was randomized.

Testing occurred in two sessions on different days during one single week. During the first session, participants were asked to complete the informed consent, a demographic questionnaire, and the working memory span tasks. After a short break they were briefed about the WCCOM compliance task and the operational task. Participants then completed a practice session, which consisted of three trials of the operational task alone, the WCCOM alone, and both tasks simultaneously (dual-task setting). Following the practice sessions, they completed the first experimental block. The second day session consisted of the remaining two blocks. A 5-min break was scheduled between blocks. During both sessions, the RSME and the card-sorting task were administered following each trial and the NASA-TLX following each block. Finally, following the experimentation, participants were debriefed via a verbal and written statement.

3. Results

Where applicable, data was aggregated across conditions for each operational mission separately (i.e., conditions 1–3 and 4–6, respectively) and by WCCOM response format; pictorial response format (1 and 4), written (2 and 5), and verbal (3 and 6), respectively. A common approach for missing data was taken, substituting means for absent data prior to data analysis (Tabachnick and Fidell, 2001). Rate of missing cells was negligible (recall that altogether there were 45 trials in each experimental condition). All post-hoc tests were LSD.

3.1. WCCOM compliance task

A three-way between-within-participants repeated measure ANOVA was conducted with presentation format (verbal, written and pictorial) × response format (verbal, written, pictorial) × task demand (2, 4, and 8 items). The dependent variable was the WCCOM compliance score. Table 3 summarizes the main effects and interactions among these factors, Table 4 provides the descriptive statistics.

There were significant main effects for response format, presentation format, and task demand, as well as significant interactions between response and presentation, response and task demand, and presentation and task demand. The three-way interaction was also significant. Post hoc tests for response format showed that participants were significantly more likely to comply when the response was verbal ($M = .79$, $SD = .03$), followed by pictorial ($M = .68$, $SD = .03$), and then written format ($M = .59$, $SD = .02$). Thus, verbal response proved to be the superior response mode. Planned comparison tests for presentation format showed that participants were significantly more likely to comply when the storage format was written ($M = .75$, $SD = .02$) or pictorial ($M = .72$, $SD = .02$) compared to verbal ($M = .60$, $SD = .01$). No significant differences were found between written and pictorial storage formats and were superior to the verbal warning storage, as shown in Fig. 4.

Comparisons for task demand revealed that participants were significantly more likely to comply at level two ($M = .83$, $SD = .01$), than at level four ($M = .71$, $SD = .02$) and least likely to comply at eight ($M = .52$, $SD = .02$), all pairwise comparisons being significantly different from one another. The interaction between presentation format and task demand is shown in Fig. 5. Post hoc analyses also revealed differences among presentation formats at the same level of task demand, as shown in Fig. 6.

Table 3
Main effects and interactions for the WCCOM task for each experimental condition separately.

Condition	Response mode	Task demand	Warning format	Format by task demand
1	Pictorial	$F(2, 20) = 32.6$, $p = .0005$	$F(2, 20) = 11.7$, $p = .0005$	$F(4, 40) = 2.7$, $p = .04$
2	Written	$F(2, 20) = 87.2$, $p = .0005$	$F(2, 20) = 10.6$, $p = .001$	NS
3	Verbal	$F(2, 22) = 36.5$, $p = .0005$	NS	NS
4	Pictorial	$F(2, 22) = 94.7$ $p = .0005$	$F(2, 22) = 43$, $p = .0005$	$F(4, 44) = 5.2$, $p = .002$
5	Written	$F(2, 22) = 35.4$ $p = .0005$	$F(2, 22) = 33.3$ $p = .0005$	$F(2, 22) = 3$ $p = .03$
6	Verbal	$F(2, 22) = 33.8$ $p = .0005$	NS	NS

α is significant at .05. Main Effect is abbreviated as ME. NS-no significant difference.

3.2. Suspect detection mission

In order to determine if the format of the warning presentation or task demand affected performance on the operational mission, a three-way 3 (format) × 3 (task demand) × 5 (trial) within-participants repeated measure analysis of variance (ANOVA) was conducted on the number of enemies identified divided by enemies present in the building. The independent variables included WCCOM storage format: verbal, pictorial, and written, task demand: storages of two, four, and eight, and trial: three blocks of 15 trials.

There was a significant main effect for task demand, Wilk's $\Lambda = .71$, $F(2, 66) = 6.3$, $p = .003$, partial $\eta^2 = .16$. No significant

Table 4
ANOVA Table for WCCOM compliance across all experimental conditions.

Effect	df	MS	F	p
Response format	2,66	2.07	15.3	.001
Storage format	2132	1.30	48.4	.001
Response by Storage	4132	.30	11.1	.001
Task demand	2132	4.96	233.4	.0001
Task demand by Response	4132	.12	5.8	.001
Task demand by Storage	4264	.04	3.3	.05
Storage by Task demand by Response	8264	.025	2.1	.05

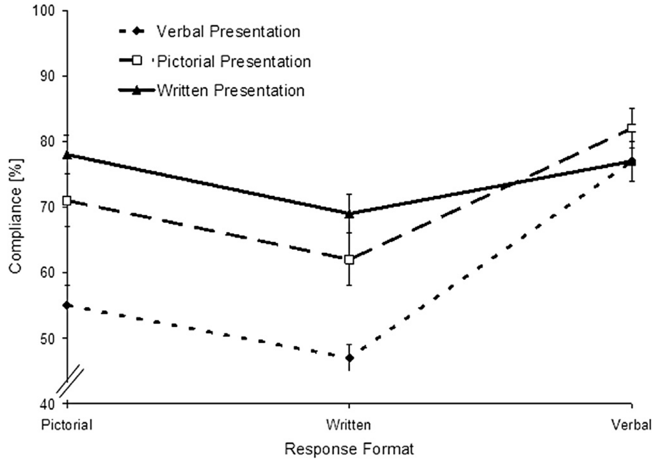

Fig. 4. Presentation format by response format interaction for the WCCOM task.

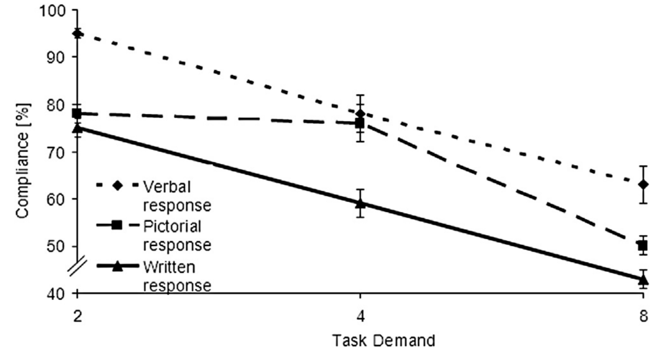

Fig. 6. Response format by task demand interaction for the WCCOM task.

effects were found for Presentation format or Trial and no significant interactions. Post hoc tests for task demand revealed that participants were significantly more likely to perform better at levels of two ($M = .965$, $SD = .006$) and four ($M = .958$, $SD = .006$) than at level eight ($M = .939$, $SD = .009$). No significant differences between level two ($M = .965$, $SD = .006$) and level four ($M = .958$, $SD = .006$) were evident.

3.3. Navigation mission

A similar analysis to the suspect detection mission, was conducted on the relative time (in %) it took to reach all 4 waypoints. There was a significant main effect of task demand, Wilk's $\Lambda = .51$, $F(2, 70) = 13.8$, $p = .0005$, partial $\eta^2 = .99$. No significant effects were found for storage format or trial and no interactions. Post hoc tests for task demand revealed that participants took significantly longer at level eight ($M = .508$, $SD = .05$) than at level four ($M = .398$, $SD = .05$). Participants took significantly longer at level eight ($M = .508$, $SD = .05$) than at level 2 ($M = .311$, $SD = .04$). No significant difference between level two and four was found.

3.4. Rating scale mental effort (RSME)

In order to determine perceived difference in effort among trials across all experimental conditions, a four-way storage format (3) X task demand (3) X trial frequency (5) within-participants repeated measure analysis of variance (ANOVA) with response format (3) as a between-participant variable was conducted. The dependent variable was the RSME measure derived for each trial. There was no significant main effect for trial or interaction with trial, thus trial was omitted as a factor from the analysis. The subsequent three-way interaction yielded a main effect for task demand ($F(2, 44) = 203.7$, $p < .0001$). Post hoc tests for task demand revealed that participants evaluated their load as significantly lower of mental effort at level of two ($M = .271$, $SD = .156$) than either four ($M = .431$, $SD = .165$) or eight ($M = .625$, $SD = .168$) warnings. Thus, as the rate of task demand increased, subjective effort estimates ratings increased. No effects were found for response format or presentation format, respectively.

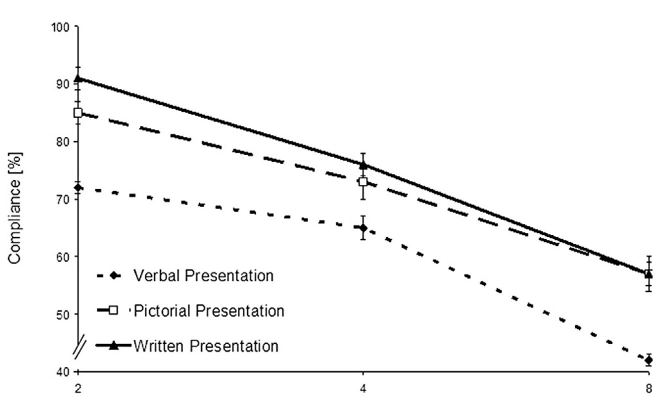

Fig. 5. Presentation format by task demand interaction for the WCCOM task.

3.5. Individual differences – WM related span tasks

Table 5 provides the summary of descriptive statistics for the four memory span measures across all participants and these values are compared to the ones reported in Shah and Miyake's (1996) study. Spearman rank order correlations were conducted among the four memory span measures and WCCOM task compliance scores for the three storage formats, across all participants, as shown in Table 6. The Spatial span was the only predictive measure that correlated with the WCCOM compliance scores but only when the storage format was pictorial or written. Hence, participants who scored higher on the spatial span were also likely to comply more when storage format was written or pictorial.

4. Discussion

Our first two hypotheses addressed differences between warning storage formats. Our first hypotheses predicted that participants would have a higher rate of compliance when coded warnings were presented in verbal compared to either written or pictorial form. This prediction was made based on a synthesis of

Table 5
Summary of Descriptive Statistics for the Individual difference Memory span measures ($N = 69$) in the current study and as reported in Shah and Miyake (1996).

Measure	M	SD	Min	Max
Current study experimental data				
Spatial span	2.05	1.40	1.0	5.0
Reading span	2.96	.58	2.0	5.0
Arrow verification	4.14	.93	1.0	5.0
Word verification	3.43	.91	1.0	5.0
Shah and Miyake, 1996 Experiment 2				
Spatial span	2.67	1.05	1.0	5.0
Reading span	3.71	.54	3.0	4.5
Arrow verification	3.04	.92	1.0	5.0
Word verification	3.14	.77	1.5	4.0

Table 6
Spearman rank order correlations for the WCCOM compliance task presentation format and the span measures (marked correlations in bold are significant at $p < .05$, $N = 69$).

	Pictorial	Written	Verbal
Pictorial	–	**.408**	.092
Written		–	.052
Spatial span	**.372**	**.373**	.220
Reading span	−.052	.171	.085
Arrow verification	.028	.148	−.059
Word verification	.030	.101	.032

existing findings already reported in the literature concerning the memory and warnings domain. There is evidence in the memory literature which supports the general notion that the verbal format results in superior recall for working memory tasks, auditory modes of communication having been generally found to be superior to written information (Penney, 1975; Murdock, 1968; Craik, 1969; Watkins and Watkins, 1980). In addition, the literature on warnings has also found greater behavioral compliance when warnings are presented in a verbal format in various environmental settings (Jaynes and Boles, 1990; Wogalter et al., 1991; Wogalter and Young, 1991). However, as shown by the results in Table 4, only one of the experimental conditions, the verbal–verbal combination of storage and response format, yielded superior compliance when the verbal format was used as the mode of warning storage. The storage format that resulted in the highest rate of behavioral compliance was the written ("printed") format.

The summarized results of previous studies were derived from studies which only examined response to explicit warnings. However, warnings in the present study were presented implicitly in a dual task situation in which the primary mission was of high demand. This situation replicates real working conditions of military operators. With respect of such conditions, Broadbent et al. (1978) and Gardiner et al. (1974) had previously shown that the 'distracters' that fill the interval of time between storage and recall serve to influence the amount of information that is remembered. Hence, if the interval between storage and recall was silent or if non-verbal distractions were present, auditory information was recalled at a greater rate than visual information. The recall of verbal cues was reduced only when an auditory distracter was presented in the time interval between storage and response. In contrast, both forms of distraction, verbal or visual, reduced the rate of recall on visually presented words. In the present study, the time interval between the warning storage component and warning recall was filled by events related operational tasks, which were visual-spatial tasks in nature. Wickens et al. (1983) who previously examined the relationship between the mode of storage of the warning and the operational task, found that cross-modal timesharing was superior to intra-modal sharing. Our second hypotheses therefore, predicted that the verbal storage format would be the superior mode of warning storage since the operational task in this study, that filled the time interval between warning storage and recall, was a visual/spatial task. However, contrary to this prediction and the studies that supported them (Broadbent et al., 1978; Gardiner et al., 1974; Wickens et al., 1983) it was found that the auditory warning mode was the least effective on average across all six experimental conditions. Inconsistent with the argued predictions, verbal warning storages did not result in the superior format. Furthermore, the existence of a dominant operational task, did not degrade performance on the visual warning presentations (pictorials and written) as compared to verbal presentation. In addition, the format of warning presentation did not affect performance on either one of the operational tasks. Thus, the visual/spatial operational task, regardless of its complexity did not interfere in timesharing with intra-modal warning storages.

Response mode (i.e., acknowledgment of compliance) played an integral role in the experimental design. The way compliance was handled here in this study is not in the common way compliance is dealt with (e.g., CHIP model, Conzola and Wogalter, 2001). The focus here was merely on the acknowledgment component of the warning while generally in the workplace operator do not need to acknowledge the receipt of the warning. For presentation format, written and pictorial warnings were the superior forms (Table 4). Yet, with regard to response format, participants were significantly more likely to comply when the response format was verbal compared to pictorial or written, regardless of storage format. Therefore, we suggest that it is the combination of storage and response format together that must be considered when evaluating possible formats of warnings delivery and acknowledgment in real world, multi-task circumstances.

4.1. WM demand

Our subsequent hypotheses examined were directed toward understanding the optimal amount of warning information that could be presented to participants before the demands of that warning affected their primary task performance. We predicted that when two or four warnings were presented, performance on neither the WCCOM task nor the operational mission would be affected; yet when the storage increased to eight warnings, performance on both the WCCOM task and the operational tasks would degrade. This assumption was based on the well-researched limits to memory storage capacity (Miller, 1956; but see also Ericson and Staszewski, 1989).

As predicted, WM load affected compliance scores for the WCCOM task across all six experimental conditions. Consistent with the Hancock and Warm (1989) model, as task demand increased, performance on the WCCOM task decreased. When the task-load was at level two (retention of two warnings) participants complied at a rate of 83%. Similarly, when the task-load was at level four (four warnings) compliance remained relatively high at a rate of 71%. Compliance scores dropped dramatically when the level of demand was eight (eight warnings). Under those conditions a low rate of 52% was observed.

In addition, WM demand affected performance on the operational tasks. For the suspect identification mission (experimental conditions 1–3) there were no significant performance differences between task-load of two or four warnings, but performance at level eight significantly differed from the lower levels. Performance at all three levels ranged from 97 to 94% indicating that the overall performance even for the highest level of demand (eight warnings) was only 6% less than perfect. This suggests a form of ceiling effect in those circumstances. For the navigation mission (experimental conditions 4–6), there was a significant effect for task demand as well, i.e., as the demand increased the time it took to complete the navigation task increased. When the task-load was at four, it took on average 30% longer to reach all four waypoints as compared to the time at level two. When participants were presented with eight warnings (level 8) it took 70% longer than it did at level two. These degrees of degradation are well characterized in the Hancock and Warm (1989) model.

The magnitude of change in performance due to WM demand seems greater on the navigation mission than on the suspect identification mission, in which participants had to visualize or construct representations of the environment in their working memory to accomplish the navigation task (see Tversky, 2003). In the suspect detection mission, as it was structured here, participants had more limited need to construct a representation of the environment. This further emphasizes how the nature of the operational task is important to consider. However, it also emphasizes the importance of the measures used to assess mission performance (e.g., time based versus accuracy based). Specifically here for the navigation mission, it was the time-based performance assessment measure that was sensitive to WM demand.

4.2. Subjective workload ratings

Our following hypotheses were predictions about the effects of task demand on subjective workload. We hypothesized that subjective workload measures for two warning storages would be lower compared to conditions with four warning storages and than eight warnings. We expected subjective workload and task demand to be associated when the number of warning storages was two or four. With eight warnings, we expected a workload dissociation or insensitivities to occur since participants may exceed their available resources (Hancock, 1996).

O'Donnell and Eggemeir (1986) suggested that performance and workload were not associated when the task demand exceeded the resources available, otherwise workload and performance were directly associated. In the Hancock and Warm (1989) model, when task demand is low or at moderate levels, the individual performing the task can adapt to the task demand and thus, performance and workload would be directly associated. Yet, when task demand is high, the individual performing the task can no longer adapt to the task demand and insensitivities will occur. RSME scores were taken after each trial in order to observe the variations in task demand. We found that task demand did in fact affect mental effort ratings; as task demand increased, subjective ratings increased. RSME ranges from 0 to 150 (150 being the highest score), the scores for levels two, four and eight were 27, 43, and 63 respectively. Unlike our expectation, RSME for the eight warning conditions did not exceed the mid-range of the scale implying perhaps that resources were not depleted, and why workload and performance associations were found for all three levels of demand. This may suggest that even greater levels of demand can be explored in subsequent experiments.

4.3. Individual differences in WM

This study also made a first attempt at examining whether individual differences in WM affect the way warnings should be coded. To do so we have replicated the Shah and Miyake (1996) spatial and verbal working memory resources span tests. Neither one of the verbal tests was found to correlate with WCCOM presentation format. Surprisingly, higher scores on the spatial span test correlated with both the pictorial storage format (an expected outcome since they are both spatial in nature) and the written ("printed") storage format, however this correlation was moderate. Future studies should aim to further examine individual difference in WM in order to further understand how those may affect compliance to warnings, particularly now, as presentation format can be more flexibly changes via personal devices.

5. Summary and conclusions

This study dealt with an uncommon way of warning compliance, when warnings are present implicitly and not directly to the operator. An experimental task, the WCCOM task was specifically designed to assess compliance under such circumstances. Compliance scores for the WCCOM task revealed the combinatorial importance of both the storage and response formats. Contrary to traditional studies in warning format under the operational conditions examined in this study, pictorials and written presentation formats resulted in higher behavioral compliance. Verbal response format was superior to pictorial and written formats, at all times.

Not surprisingly, demand affects warning compliance; as it increases, compliance decreases. In tasks that are more complex, it is expected to find effects on performance at an earlier rate. As the number of retained warnings increased from 2 to 4 items, degradations were found in the navigation task, but for the suspect identification task such changes were observed only with 8 warning items. Complex tasks can be affected by mental demand, and it is important to limit the number of potential warnings to the minimum necessary. Finally, although coding of warnings generates implicit, indirect and prone to error messages, participants were able to retain a small number of warnings without harming their performance on the main operational task. Thus, in small numbers, coding of critical warnings is feasible even in more demanding operational contexts.

Acknowledgments

The view expressed in thesis work are those of the authors and do not necessarily reflect official Army policy. This work was supported in part under Grant DAAD19-01-0621 by the Defense Advanced Research Projects Agency (DARPA) via the Department of Interior, P.A. Hancock, Principal Investigator. The Grant Monitors were Dr. Dylan Schmorrow, Dr. Russ Shilling, and Dr. Jim Patrey. The authors wish to acknowledge the contribution of Dr Paul Ward to the design of the WCCOM compliance task and Dr Kelly A. Burke for her general assistance.

References

Baddeley, A.D., 1986. Working Memory. Oxford University Press, Oxford.
Broadbent, D., Vines, R., Broadbent, M., 1978. Recency effects in memory, as a function of modality of intervening events. Psychol. Res. 40 (1), 5–13.
Cairney, P., Sless, D., 1982. Communication effectiveness of symbolic safety signs with different user groups. Appl. Ergon. 13 (2), 91–97.
Conzola, V.C., Wogalter, M.S., 1998. Consumer product warnings:effects of injury statistics on recall subjective evaluations. In: Proceedings of the Human Factors Society 42nd Annual Meeting. Human Factors Society, Santa Monica, pp. 559–563.
Conzola, V.C., Wogalter, M.S., 2001. A communication-human information processing (C-HIP) approach to warning effectiveness in the workplace. J. Risk Res. 4 (4), 309–322.
Craik, F.I., 1969. Modality effects in short-term storage. J. Verbal Learn. Verbal Behav. 8 (5), 658–664.
Daneman, M., Carpenter, P.A., 1980. Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior 19, 450–466.
Duarte, M.E.C., Rebelo, F., 2007. Virtual reality in the study of warnings effectiveness. In: Dainoff, M.J. (Ed.), Ergonomics and Health Aspects of Work with Computers. HCII 2007, Lecture Notes in Computer Science, 2007, Vol. 4566/2007. Springer-Verlag Berlin, Heidelberg, pp. 189–198.
Driskell, J.E., Olmstead, B., 1989. Psychology and the military: research applications and trends. Am. Psychol. 44 (1), 43–54.
Engle, R.W., 2002. Working memory capacity as executive attention. Curr. Dir. Psychol. Sci. 11, 19–23.
Engle, R.W., Tuholski, S.W., Laughlin, J.E., Conway, A.R., A, 1999. Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. J. Exp. Psychol. Gen. 128, 309–331.
Ericsson, K.A., Staszewski, J.J., 1989. Skilled memory and expertise:Mechanisms of exceptional performance. In: Klahr, D.K., Kotovski, K. (Eds.), Complex Information Processing: The Impact of Herbert A. Simon. Erlbaum, Hillsdale, NJ.
Familant, M.E., Detweiler, M.C., 1993. Iconic reference: evolving perspectives and an organizing framework. Int. J. Man Mach. Stud. 39, 705–728.
Francis, W., Kučera, H., 1982. Frequency analysis of English usage: Lexicon and grammar. Houghton Mifflin, Boston.
Gardiner, J.M., Thompson, C.P., Maskarinec, A.S., 1974. Negative recency in initial free recall. J. Exp. Psychol. 103 (1), 71–78.
Hart, S.G., Staveland, L.E., 1988. Development of a multi-dimensional workload rating scale: results of empirical and theoretical research. In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Elsevier, Amsterdam, The Netherlands, pp. 139–183.
Hancock, P.A., Warm, J.S., 1989. A dynamic model of stress and sustained attention. Hum. Factors 31, 519–537.
Hancock, P.A., 1996. Effect of control order, augmented feedback, input device and practice on tracking performance and perceived workload. Ergonomics 39, 1146–1162.

Jaynes, L.S., Boles, D.B., 1990. The effect of symbols on warning compliance. In: Proceedings of the Human Factors Society 34th Annual Meeting. Human Factors Society, Santa Monica, CA, pp. 984–987.

Kanki, B.G., 1996. Stress and aircrew performance: a team-level perspective. In: Driskell, J.E., Salas, E. (Eds.), Stress and Human Performance. Erlbaum, Mahwah, NJ, pp. 127–162.

Keller, P., Stevens, C., 2004. Meaning from environmental sounds: types of signal–referent relations and their effect on recognizing auditory icons. J. Exp. Psychol. Appl. 10 (1), 3–12.

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

Murdock, B.B., 1968. Modality effects in short-term memory: storage or retrieval? J. Exp. Psychol. 77 (1), 79–86.

Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63, 81–97.

O'Donnell, R.D., Eggemeir, F.T., 1986. Workload assessment methodology. In: Boff, K.R., Kaufman, L., Thomas, J.P. (Eds.), Handbook of Perception and Human Performance. Wiley, New York, pp. 4/1–42/49.

Penney, C.G., 1975. Modality effects in short-term verbal memory. Psychol. Bull. 82 (1), 68–84.

Rogers, W.A., Lamson, N., Rousseau, G.K., 2000. Warning research: an integrative perspective. Hum. Factors 42 (1), 102–139.

Shah, P., Miyake, A., 1996. The separability of working memory resources for spatial thinking and language processing: An individual differences approach. Journal of Experimental Psychology: General 125, 4–27.

Stephan, L.K., Smith, E.S., Martin, L.R., Parker, P.A.S., Mcanally, I.K., 2006. Learning and retention of associations between auditory icons and denotative Referents: Implications for the design of auditory warnings human factors. J. Hum. Factors Ergon. Soc. 48 (2), 288–299.

Tabachnick, B.G., Fidell, L.S., 2001. Using Multivariate Statistics, fourth ed. Allyn & Bacon, Needham Heights, MA.

Tversky, B., 2003. Structures of mental spaces: how people think about space. Environ. Behav. 35, 66–80.

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

Watkins, O.C., Watkins, M.J., 1980. The modality effect and echoic persistence. J. Exp. Psychol. Gen. 109, 251–278.

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

Wogalter, M., 1994. Factors Influencing the Effectiveness of Warnings. Public Graphics, Lunteren, The Netherlands.

Wogalter, M.S., Fontenelle, G.A., Laugherty, K.R., 1992. Effects of warning signal words on consumer products hazard perceptions. In: Proceedings of the Human Factors Society 37th Annual Meeting. Human Factors Society, Santa Monica, CA, pp. 935–939.

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. Human Factors Society, Santa Monica, CA, pp. 525–529.

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. Human Factors Society, Santa Monica, CA, pp. 718–722.

Wogalter, M.S., Young, S.L., 1991. Behavioral compliance to voice and print warnings. Ergonomics 34 (1), 79–89.

Woods, D.D., 1995. The alarm problem and directed attention in dynamic fault management. Ergonomics 38 (11), 2371–2393.

Zijlstra, F.R.H., Van Doorn, L., 1985. The Construction of a Scale to Measure Perceived Effort. Department of Philosophy and Social Science, Delf University of Technology, Delf, The Netherlands.