Technical Report 1223

Automated Feedback and Situation Awareness in Net-Centric C3

John S. Barnett
U.S. Army Research Institute

Jennifer M. Ross
University of Central Florida
Consoritum Research Fellows Program

February 2008

20080320058

United States Army Research Institute for the Behavioral and Social Sciences

U.S Army Research Institute
for the Behavioral and Social Sciences

A Directorate of the Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

Barbara A. Black

BARBARA A. BLACK, Ph.D.
Research Program Manager
Training and Leader Development

Michelle Sams

MICHELLE SAMS, PhD.
Director

Technical reviews by:

Jenifer S. Tucker, U.S. Army Research Institute
Daniel Horn, U.S. Army Research Institute

NOTICES

DISTRIBUTION: Primary distribution of this Technical Report has been made by ARI. Please address correspondence concerning distribution of reports to: U.S. Army Research Institute for the Behavioral and Social Sciences, Attn: DAPE-ARI-MS, 2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926.

FINAL DISPOSITION: This Technical Report may be destroyed when it is no longer needed. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The findings in this Technical Report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) February 2008	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) March 2006 – June 2007

4. TITLE AND SUBTITLE Automated Feedback and Situation Awareness in Net-Centric C3	5a. CONTRACT OR GRANT NUMBER
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) John S. Barnett (U.S. Army Research Institute) Jennifer M. Ross (University of Central Florida)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER
	5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: DAPE-ARI-IF 2511 Jefferson Davis Highway Arlington, Virginia 22202-3926	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 2511 Jefferson Davis Highway Arlington, Virginia 22202-3926	10. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Tehcnical Report 1223

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Contracting Officer's Representative and Subject Matter POC: John Barnett

14. ABSTRACT *(Maximum 200 words)*:

The goal of net-centric warfare (NCW) is to give soldiers an information advantage that leads to a war-fighting advantage. However, NCW systems are quite complex and dynamic, characteristics which can lead to impaired situation awareness (SA) and increased mental workload. It has been suggested that an automated alerting system would help Soldiers focus their attention on mission critical events. This series of experiments investigated how automated audio-visual alerts affect user SA and perceived workload. Two similar experiments were conducted. In each experiment, participants viewed a simulation of a net-centric system, the Force XXI Battle Command Brigade and Below (FBCB2), which included an automated alerting system. SA and workload were measured both with the alerting system enabled and disabled. In the second study, the difficulty of the monitoring task was increased and the automated alerts included a pop-up pictorial representation of the critical event. Results indicate that automated alerting systems do not improve user SA, but they also do not impair user SA. However, mental workload was significantly lower when alerts were enabled. These results can be used to aid decisions about whether or not to include automated alerts in NCW systems.

15. SUBJECT TERMS

Net-Centric Warfare; Digitization: Alerts; Workload; Situation Awareness

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	30	Ellen Kinzer Technical Publication Specialist 703-602-8047

Technical Report 1223

Automated Feedback and Situation Awareness in Net-Centric C3

John S. Barnett
U.S. Army Research Institute

Jennifer M. Ross
University of Central Florida
Consortium Research Fellows Program

Technology-Based Training Research Unit
Stephen L. Goldberg, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
2511 Jefferson Davis Highway, Arlington, Virginia 22202-3926

February 2008

Army Project Number
622785A790

Personnel Performance and Training Technology

AUTOMATED FEEDBACK AND SITUATION AWARENESS IN NET-CENTRIC C3

EXECUTIVE SUMMARY

Research Requirement:

The goal of net-centric warfare (NCW) system is to give Soldiers an information advantage that leads to a war-fighting advantage. However, the complex and dynamic nature of NCW systems leads to impaired situation awareness (SA) and heightened levels of mental workload for the human operator. In an effort to increase SA, it has been recommended to include an automated alerting system that would provide visual and auditory alerts when critical events occur, thus aiding users' ability to focus their attention on mission critical events. The following research investigated the moderating effects of automated audio-visual alerts on SA and perceived workload while using a NCW system.

Procedure:

Two experiments were conducted in which the effects of automated alerts on SA and WL in monitoring NCW systems were examined. In the first experiment, participants (N=18) viewed a NCW system, the Force XXI Battle Command Brigade and Below (FBCB2), both with and without the automated alerting system (System to Help Implement and Empower Leader Decisions; SHIELD). The alerts appeared as auditory dings and pop-up text boxes. In the second research, participants' (N=26) SA and workload were examined for FBCB2 again with and without SHIELD. However, in this research, the difficulty of the monitoring task was increased and the automated alerts included pop-up pictorial representations of the critical events.

Findings:

For the first experiment, we did not find significant effects for overall SA or WL. However, we found a significant interactive effect for military experience. For those without military experience, there was a significant increase in SA in the automated alerting condition as compared to the no automated alerts condition. However, we did not find this effect in the second experiment. We also did not find significant effects for overall SA. On the other hand, the results from Experiment 2 indicated that WL was significantly lower for the condition with the automated alerts compared to the condition without the automated alerts.

Utilization and Disseminationof Findings:

The findings from the first experiment indicate that while automated alerts do not affect the SA of experienced users they do benefit novice users. Thus, automated alerts may potentially be used to train novice users to recognize critical events for themselves, but as users gain experience the automated alerts become less beneficial.

In the second experiment, it was found that automated alerts decreased WL. This finding indicates that while the addition of an auditory and visual alerting system increases the physical complexity of the system, it also has the unanticipated effect of decreasing the perceived cognitive complexity of the system. A post-hoc explanation of these results indicates that by providing this alerting system to cognitively cue participants to critical events, users may feel that they can rely on SHIELD to alert them to unsafe situations (i.e., they have a teammate) and thus their personal feelings of responsibility and individual mental demand may be reduced.

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

AUTOMATED FEEDBACK AND SITUATION AWARENESS IN NET-CENTRIC C3

Sharing information over a computer network is becoming a common means of communication in many fields. The U.S. Army has taken advantage of this technology to improve command, control, and communication (C3) for its forces. Field commanders can now watch events unfold on a computer screen.

The use of networked computers to enhance C3 has opened up new possibilities. Data can be analyzed by computer and results displayed to leaders, providing them with better information on which to base decisions. Leaders can be provided with immediate feedback about their performance. One promising technology is for computer systems to monitor the data stream and provide alerts when critical events occur to ensure they are not missed by the operator.

As the U.S. Army gains more experience with networked digital C3 systems, more features are likely to be added to the systems. Features such as the automated alerting system described above can help direct the user's attention to important events and increase situation awareness (SA). However, experience with automated alerting systems in other areas has shown that automation can sometimes become intrusive and capture the user's attention at inappropriate times.

This raises the question of whether immediate feedback or automated alerting systems enhance SA or interfere with SA. Research suggests that it depends on the situation and environmental variables (Billings, 1997). Under certain conditions alerts improve SA while under other conditions they interfere with SA. If immediate feedback and alerts are built into future networked C3 systems, it would be important to know how alerts affect SA, and under what conditions.

To answer these questions, we conducted two experiments using simulations of a networked C3 system with an automated alerting application. This paper presents the experiments and our findings. The first section of the paper discusses some background about networked C3 and concerns about how automated alerts may affect SA, as well as some related issues. Following that, we will discuss the method, results, and a discussion of findings for the experiments.

BACKGROUND

Digital C3

The U.S. Army employs net-centric C3 systems in a process referred to as "digitization." These systems employ computer automation to help leaders and Soldiers perform many of the C3 functions previously accomplished manually, such as planning missions, distributing orders and reports, and creating and distributing map overlays containing battlefield graphics such as obstacle areas, unit boundaries, and phase lines.

Information on the tactical situation can be distributed over the network from upper-echelon command centers down to the lowest-level combat formations, and vice-versa. Digitization not only increases combat capabilities, but also improves safety by reducing the chances of fratricide or "blue on blue" incidents. In addition, combat units who use digital systems are expected to maintain better SA and to plan and execute operations more quickly than non-digital units (Barnett, Meliza, & McCluskey, 2001).

Thus, digital C3 serves as a decision-support system for commanders. It helps them visualize the battle space and presents needed information in a format that fosters the commander's SA. Digitization also provides analytical tools, such as terrain analysis tools and automated warnings that can further enhance SA.

There are a number of different digital systems. Many of the systems are specific to certain warfighting functions such as Intelligence, Maneuver, Fires, and Sustainment. These systems were designed to fulfill C3 related to the warfighting functions and are typically located in tactical operations centers (TOC) at higher echelons. Although these systems were originally designed to operate within the warfighting function, they are able to share most information with other systems on the network. Operators of these systems spend part of their time observing SA displays and part of their time using analytic tools, such as a line-of-sight calculator, and preparing products, such as operations or obstacle overlays.

The Force XXI Battle Command, Brigade and Below (FBCB2; see Figure 1) is a digital C3 system designed for lower-echelon maneuver units and is typically located in vehicles such as tanks and infantry fighting vehicles. FBCB2 allows lower echelon units to plan missions and routes, develop battlefield graphics such as obstacle overlays, and share the information over the network.

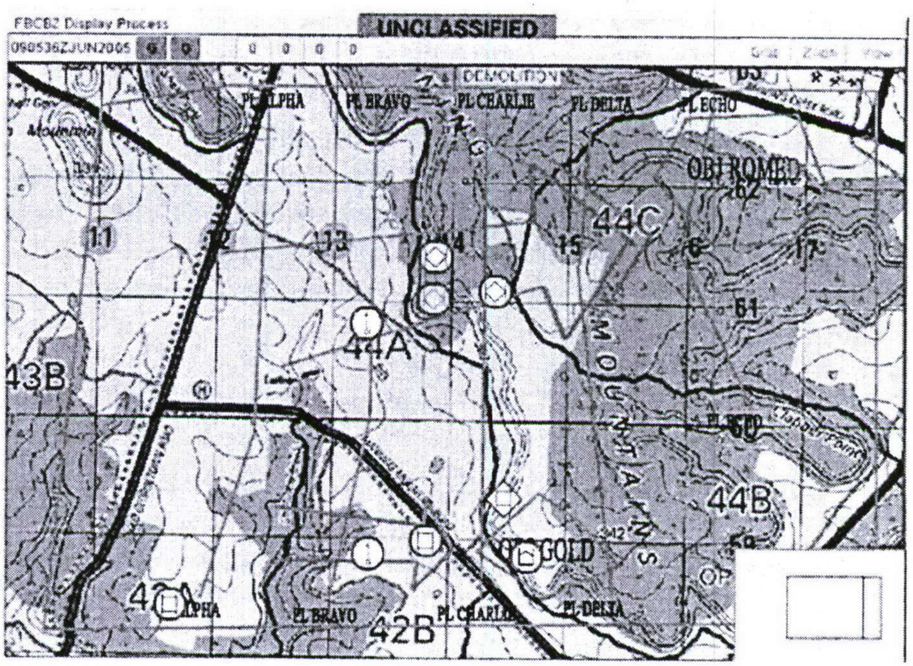


Figure 1. Example of an FBCB2 display.

FBCB2 presents a dynamic view of the battlespace. Once the mission is executed, leaders and Soldiers can follow the progress of the mission on the FBCB2 display. Vehicle-mounted FBCB2s can automatically update the vehicle's position using a Global Positioning System

(GPS). The vehicle's position is periodically updated and transmitted to the network, so that the FBCB2 display shows the vehicle's own position as well as the positions of other vehicles in the unit. This allows leaders and Soldiers to develop sound awareness of the friendly situation. In addition, when enemy units are reported on the network through SPOT reports (i.e., report providing information about the location of enemy forces), their positions are displayed as well. In the field, as vehicle crews approach expected threat areas, crews are likely to spend less time viewing FBCB2 SA displays and more time visually scanning for threats.

Automation's Affect on Situation Awareness

Research suggests that including automation into a system introduces potentially detrimental consequences (Bowers, Deaton, Oser, Prince, & Kolb, 1995), including complacency (Morgan, Herschler, Weiner & Salas, 1993), increased monitoring requirements (Kantowitz & Campbell, 1996), and a loss of SA (Bowers, Oser, Salas & Cannon-Bowers, 1996).

A number of different factors may contribute to a reduction in SA. Bowers et al. (1996) suggested that relying too much on the automation leads to complacency, which in turn leads to reduced SA. Automation may reduce workload during low-workload periods and increase workload during high workload periods (Parasuraman, Mouloua, Molloy & Hilburn, 1996). Therefore, while automated alerts may be helpful during low workload periods and help increase SA, during high stress, high workload periods, alerts may become intrusive and responding to them may not only increase workload, but shift the operator's attention from more important tasks, thus reducing SA.

Change Blindness

Another consideration for introducing automated alerting into digital C3 systems is that sometimes operators who view visual displays fail to detect changes that occur on those displays, a phenomenon called change blindness. Change blindness tends to occur in concurrence with various types of visual transients such as icon movement, screen flashes, or eye blinks. In addition, operators may fail to detect changes if they are performing other tasks. Durlach and Chen (2003) found that individuals tended to detect changes in icon appearance, disappearance, and color changes but had more difficulty detecting changes in icon type and movement, particularly if the icon was in the periphery of the screen and the movement was small. The concept of change blindness relates to automated alerts in two ways. First, alerts may bring critical events to the attention of system operators when SA displays are not being closely monitored. Second, alerts may hinder the operator by capturing the operator's attention at inopportune times, thus causing the operator to miss a change on the display they would have noticed otherwise. Either case affects the operator's SA.

System to Help Implement and Empower Leader Decisions (SHIELD)

SHIELD is a research tool that simulates an automated alerting system for net-centric C3 systems (for information regarding the development of SHIELD see Aiken, Green, Arntz, & Meliza, 2005). SHIELD has numerous features designed to reduce intrusiveness and support the conduct of after action reviews (AARs). SHIELD was designed to run as a stand alone system or as an application on any network system. As an application on FBCB2 and the Command and Control PC (C2PC), this feature allows the work of monitoring alerts to be distributed among nodes within a network. SHIELD allows the user to decide whether alerts are triggered from a unit or vehicle perspective. For example, using SHIELD within the TOC may provide an alert when any platform within the unit approaches a minefield. In addition, using SHIELD on a vehicle may alert the vehicle commander when their vehicle approaches a minefield). SHIELD provides information in several different formats, including textual and graphic displays, as illustrated in Figure 2. SHIELD allows the user to temporarily dismiss an alert or cancel the alert for the rest of a mission. SHIELD maintains an AAR log file that enables the user to call up alerts, data on user responses to alerts, and other information relevant to AARs. Users can dismiss alerts during high workload periods and then call up the AAR log file during subsequent lower workload periods to see if any of the alerts are still relevant, that is, the situations prompting the alert are still a concern.

Current Research

The current research seeks to answer questions about how immediate feedback from automated alerting systems affects SA under varying workload conditions. Based on prior research, under conditions of low workload alerts should direct the user's attention to important events and thus improve SA.

Hypothesis 1. Under low workload SA is greater with alerts enabled than with alerts disabled.

On the other hand, under high workload conditions alerts may divert attention from more important tasks and consequently may interfere with maintaining SA.

Hypothesis 2. Under high workload conditions, SA is lower with alerts enabled than with it disabled.

EXPERIMENT 1

Method

Participants. Eighteen people participated in experiment 1. Participants were six U.S. Army Soldiers and 12 university students. For the Soldiers, the mean age was 26 (SD = 8), while the mean age for the students was 27 (*SD* = 5). The Soldiers averaged seven (SD = 5) years of military experience, while no university students reported any military experience. Students were compensated with course credit, while Soldiers received no compensation. For Soldiers, ranks ranged from Specialist to Sergeant, and included several United States Military Academy (USMA) cadets. Military Occupational Specialties included 21B and 11B, as well as the USMA cadets. Three of the Soldiers had prior experience with FBCB2, and three did not.

Apparatus. A simulation of a networked C3 system was presented on a laptop computer (Pentium M) using a 14" graphics monitor operating under 1400 by 1050 pixel resolution. The system was integrated with headphones that the participants wore during all experimental and practice trials. The C3 system simulated was the FBCB2. Like FBCB2, the simulation presented a map display showing locations of friendly units, enemy units (if known) and battlefield graphics such as phase lines, unit boundaries, obstacle belts, etc. For the purposes of this experiment, three Army training scenarios were programmed, a practice scenario and two full-length experimental scenarios, each based on different topographical maps and orders of events.

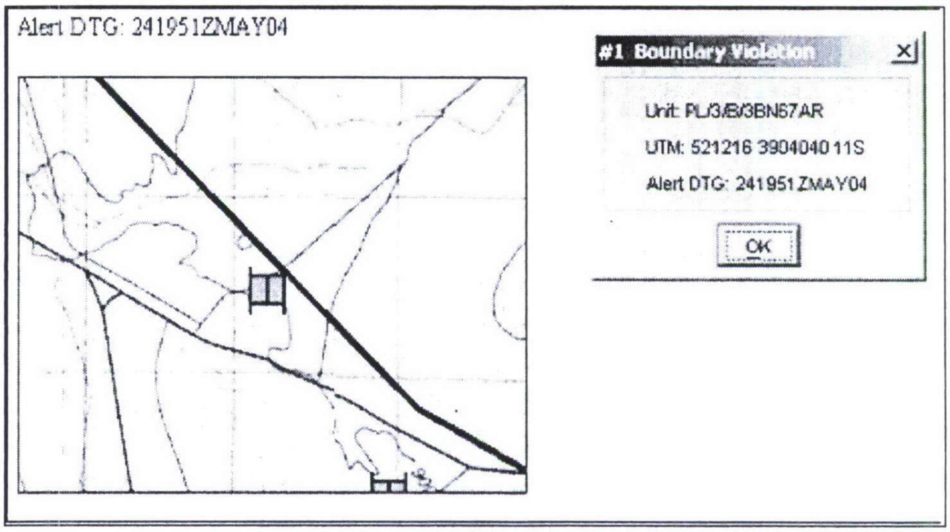


Figure 2. Examples of a SHIELD graphic alert (Left) and text alert (Right).

A simulation of SHIELD ran concurrently with the FBCB2 task during experimental trials. The current experiment employed SHIELD alerts triggered by five significant events: unit approaching a minefield, receipt of an enemy SPOT report, a new friendly unit appearing on the display, a friendly unit violating battlefield boundaries, or a unit approaching a nuclear-biological-chemical (NBC) contaminated area. The SHIELD system has a number of other features which were not used for this research as they would have added unnecessary complexity to the experiment. This experiment employed SHIELD textual alerts, but it did not include graphical alerts or other SHIELD capabilities. Further, participants were not required to interact with the alerts in any way (i.e., subjects did not have to take actions to remove alerts from the screen), nor could they control how long the alert was displayed. The only SHIELD features used were the visual and auditory alerts.

Workload was varied by manipulating task difficulty. In each FBCB2 scenario, task difficulty, defined as the number of significant events (i.e., minefield, SPOT report, etc.) per time interval, was increased every 5 minutes. The exception was the practice scenario, which was conducted at a very low difficulty condition of one event roughly every 40 seconds (10 total events) for a total of 6 minutes. In each of the 15 minute experimental scenarios, the first 5 minute interval represented a low difficulty condition and had on average 1 event every 30 seconds (10 total events). The second 5 minute interval represented the moderate difficulty condition and had on average 1 event every 20 seconds (15 total events). The third 5 minute interval represented the high difficulty condition and had on average 1 event every 10 seconds (30 total events). The order of significant events was randomized for each scenario with the constraint that each type of event appeared equally as often as the other events during each of the five-minute blocks.

Participants' SA was measured using the Situation Awareness Global Assessment Technique (SAGAT). The SAGAT questionnaires were developed following Endsley's (2000) guidance. Example queries included recalling the approximate number of friendly units currently on the display and drawing conclusions about which objective the commander's unit (icon) is heading towards. For an example of a SAGAT questionnaire please see Appendix A.

Subjective workload was measured using a pen and paper version of the NASA Task Load Index (TLX; Hart & Staveland, 1988). The NASA-TLX uses six dimensions to assess mental workload: mental demand, physical demand, temporal demand, effort, performance, and frustration. Each dimension is first rated by the participant on a scale from 0 to 100 with higher numbers reflecting more workload. Next, paired comparisons are conducted, which require the participant to choose among two dimensions, for which was the most important contributor to workload (e.g., Temporal Demand vs. Mental Demand). These forced-choice comparisons were conducted across all pairs of the six dimensions (15 total paired comparisons).

At the end of the experiment participants completed a six-item exit questionnaire, which asked participants their opinion of how the alerts affected their performance or how they may affect the performance of others when using FBCB2 (See Appendix B). Responses were made on an agreement scale (1 = Strongly Disagree; 4 = Strongly Agree).

Procedure. After completing an informed consent and demographics questionnaire, participants then completed a training session, which informed them about their task during the experiment and introduced them to the basic information they would need on FBCB2 and automated alerts. Following the training the participants were given a training evaluation; this evaluation ensured that they understood the task and the basic information on FBCB2 and SHIELD needed to complete the task. After successfully completing the training, participants were given a six-minute practice scenario followed by the SAGAT and NASA-TLX questionnaires to familiarize them with how the experiment would proceed.

Participants then completed two experimental sessions of approximately 15 minutes each. During the sessions they were asked to view the FBCB2 display and monitor it for significant events. For one session, alerts were enabled, while for the other session alerts were disabled. The conditions (alerts enabled or disabled) and scenarios (which of the programmed FBCB2 scenarios they observed) were counterbalanced, creating four possible orders.

At pre-programmed intervals of every five minutes, the FBCB2 simulation was halted and the display replaced by a blank black screen. Immediately, after the simulation was stopped the SAGAT and NASA-TLX questionnaires were administered to the participants. After the participants completed the questionnaires, the participants continued monitoring the FBCB2 display from the point at which it was stopped. These stops were repeated three times for each 15 minute scenario; this number was chosen because Endsley (2000) reported no decrements to performance with three stops within a fifteen minute period. In all, there were three SAGAT and three NASA-TLX measures for each FBCB2 session, for a total of six of each measure for the two sessions. Following the two sessions, participants were asked to complete the exit questionnaire.

Results

First, we examined the equivalence of the two FBCB2 scenarios by collapsing over alert condition. A paired comparisons t-test was conducted examining overall performance on the two programmed FBCB2 scenarios. The results indicated that the two scenarios did not significantly differ in SAGAT performance (Scenario 1: $M = 88.70$, SE = 0.01; Scenario 2: M = 88.90, SE = 0.01), $t(17) = -0.19$, $p = .85$, *ns*. In addition, the two scenarios did not differ in overall workload levels (Scenario 1: M = 41.73, SE = 4.59; Scenario 2: M = 46.82, SE = 4.27; $t(17) = -1.51$, $p = .15$).

To examine SAGAT performance, we conducted a repeated measures ANOVA, with a 2 (Alert Condition) by 3 (Task Difficulty) design with all within-subject measures (See Table 1). There was a significant effect for task difficulty. To compare the means of the task difficulty treatment groups, a Fischer least significant difference pairwise analysis was conducted. Results indicated that the low difficulty condition had significantly higher SAGAT scores than either the moderate difficulty or high difficulty sessions, which did not significantly differ from one another ($p > .05$; See Figure 3). However, the main effect for alerts and the interaction between alerts and task difficulty were not significant ($p > .05$ in both cases; See Table 1).

Table 1. *Analysis of Variance for SAGAT Performance*

Source	*df*	F	η	*p*
Alerts (A)	1	1.90	.32	.19
Task Difficulty (D)	2	8.04*	.57	.001
A * D	2	0.88	.22	.42
Within-group error	34	(.03)		

Note. Values enclosed in parentheses represent mean square errors.*$p<.01$.

In regards to our first hypothesis, that under low workload SA will be greater with alerts enabled than with alerts disabled, our results using a paired comparison t-test were not significant, albeit in the right direction, $t(18) = 0.31$, $p = .186$ (see Table 1). In regards to our second hypothesis concerning the high workload condition, in which we hypothesized that alerts may divert attention from more important tasks and consequently interfere with maintaining SA, we also had nonsignificant results, $t(17) = -0.27$, $p = .793$). There was no significant difference between the alerts and no alerts conditions in the moderate difficulty level ($p > .05$).

Table 2. *SAGAT performance as a function of feedback and difficulty condition*

Feedback Condition	Difficulty Level	
	Low Difficulty	High Difficulty
Alerts Enabled	87.33 (13.67)	69.89 (21.45)
Alerts Disabled	86.00 (16.25)	71.30 (14.55)

Note: standard errors reported in parenthesis.

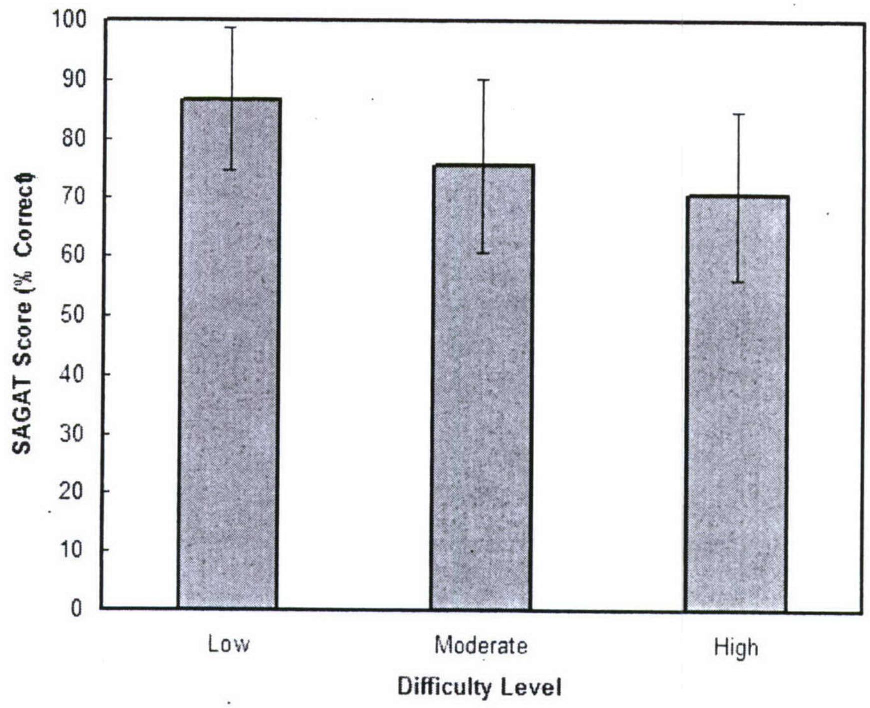


Figure 3. SAGAT performance across task difficulty conditions.

Note: Bars represent standard deviation.

The findings suggest that there may be an intervening variable which affected our results. An overall analysis indicated that the alert condition by military experience approached significance (p=.11; See Table 3). Examination of the data indicated a difference in SA scores between the Soldiers and students. We analyzed the effect of alerts separately for Soldiers and students using paired samples t-test. These analyses found no significant difference between Soldiers' SA scores with and without alerts (t(5) = -0.44, p = .677), but a significant difference between non-Soldier SA scores with and without alerts (t(11) = 2.49, p = .03). This suggests those without military experience demonstrated significantly higher SA with alerts enabled than without alerts enabled (see Table 4).

Table 3. *Analysis of Variance for SAGAT Performance with Military Performance*

Source	*df*	F	η	*p*
Between Subjects				
Military Experience (M)	1	0.01	>.00	.94
Within Subjects				
Alerts (A)	1	0.64	.20	.43
A * M	1	2.88	.39	.11
Task Difficulty (D)	2	5.60	.51	.008
D * M	2	1.09	.25	.35
A * D	2	0.93	.23	.41
A * D * M	2	0.11	.08	.90
Within-group error	34	(.03)		

Note. Values enclosed in parentheses represent mean square errors.*p<.01.

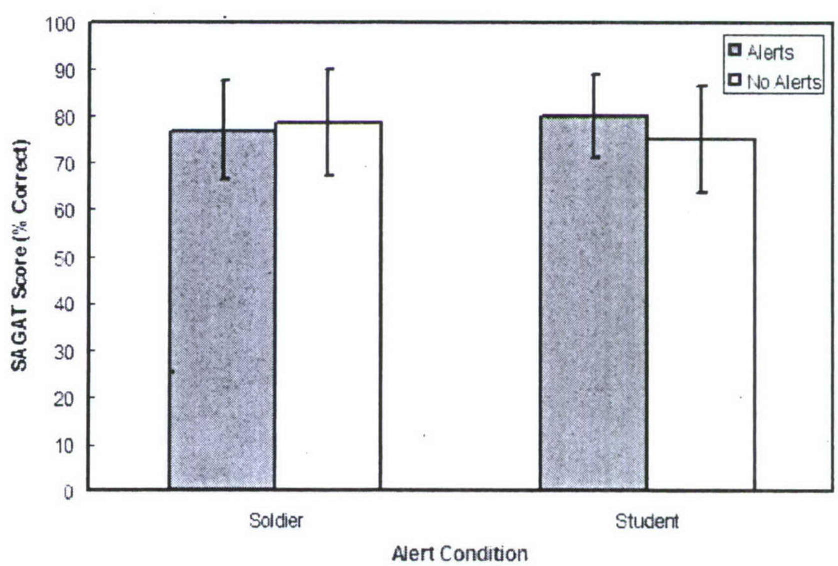


Figure 4. SAGAT performance as a function of military experience.

Note: Bars represent standard deviation.

To examine workload we used a repeated measures ANOVA, with a 2 (alerts condition) by 3 (task difficulty) design examining NASA-TLX overall weighted workload scores with all within-subject measures (See Table 4). There was a significant effect for task difficulty. To compare the means of the task difficulty treatment groups, a Fischer least significant difference pairwise analysis was conducted. Results indicated that all conditions were significantly different from one another in the expected direction (low < moderate < high; $p < .01$ in all cases; See Figure 5). However, the effect of alerts and the interaction between alerts and task difficulty did not reach significance ($p > .05$ in both cases; See Table 4).

Table 4. *Analysis of Variance for NASA-TLX Workload*

Source	*df*	F	η	*p*
Alerts (A)	1	2.26	.34	.15
Task Difficulty (D)	2	30.19*	.80	<.005
A * D	2	1.03	.24	.22
Within-group error	34	(43.82)		

Note. Values enclosed in parentheses represent mean square errors.*p<.01.

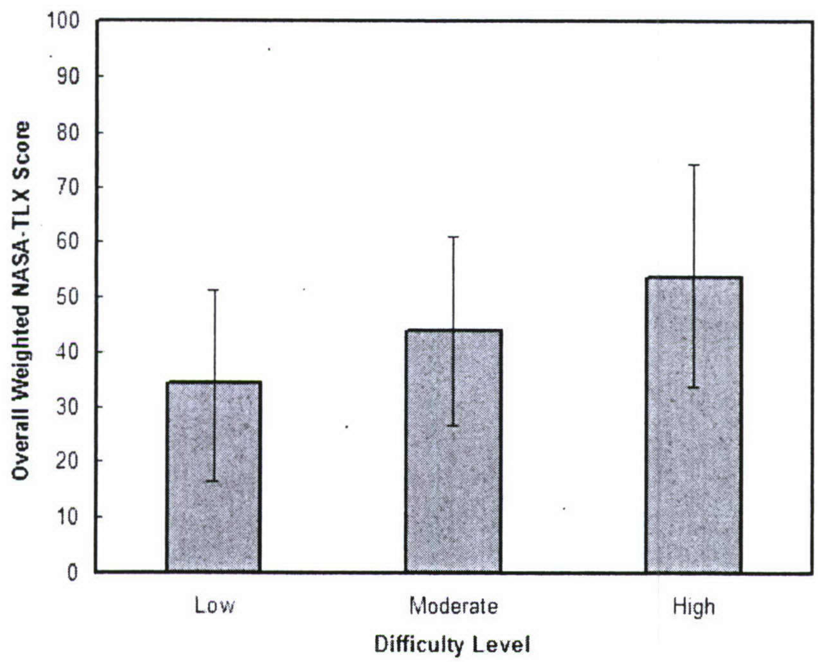


Figure 5. Workload across task difficulty conditions.

Note: Bars represent standard deviation.

While not significant at the overall analysis an exploratory more in-depth level of analysis was performed due to the theoretical nature of this analysis to our hypothesis. Namely that level of difficulty should interact with the alert condition to produce a noticeable effect on user workload experience. Using Fisher's least significant differences paired samples t-test for alert condition at each level of difficulty, a significant effect was found for the low difficulty level. A t-test revealed that there was a significant difference between the alerts enabled condition and the alerts disabled condition, $t(17) = -2.07$, $p = .05$. The results demonstrated that the alerts condition had a significantly lower workload ($M = 31.15$, $SE = 3.69$) than the no alerts condition ($M = 37.00$, $SE = 4.94$) in the low difficulty level. The moderate and high difficulty session however failed to reach significance in paired comparisons t-tests between the alerts and no alerts conditions for workload ($p > .05$ in all cases).

In regards to the exit questionnaire, all mean responses were between disagree and agree. Thus, responses were relatively neutral for this experiment (see Table 3).

Table 5. *Descriptive Statistics for Exit Questionnaire.*

Exit Questionnaire Item	Mean	Standard Deviation
1	2.50	1.04
2	2.61	0.92
3	2.72	0.90
4	2.78	0.81
5	2.83	0.86
6	2.50	0.71

Discussion

Alerts are important in bringing critical tactical situations to the attention of operator and users of networked C3 systems, however, attention must be given to the relationship between workload and the ability of alerts to enhance SA. In the current experiment, task difficulty was operationally defined in terms of the frequency of occurrence of events presented to the participants via the display.

We found that perceived workload increased as task difficulty was intensified. Workload also had an affect on SA, as participants had greater SA scores at the low difficulty level.

The results indicated that alerts did not affect SA or perceived workload for the group of participants as a whole. Indeed, though there were no differences in perceived workload, mean workload was actually lower for the FBCB2 tasks with alerts enabled ($M = 41.57$; $SE = 4.01$) than for the FBCB2 task alone ($M = 46.98$; $SE = 4.81$).

However, we found differences between the Soldiers and students such that the SA scores for the students differed depending on whether the alerts were enabled. Since the Soldiers were more familiar with the task and environment simulated by FBCB2 than the non-Soldiers, the alert conditions may have had less of an effect on the Soldiers' SA and a greater effect on students' SA. This suggests that expertise plays a role in how automated alerts affect SA, in that including automated alerting systems similar to SHIELD in net-centric C3 systems may be more beneficial for novices. On the other hand, experts may not benefit from it as much.

Additionally, SA was significantly better and workload was significantly lower in the low difficulty condition than at moderate or high difficulty. Taken together, these findings reiterate the connection between SA and mental workload. Fracker (1989) suggested both SA and mental workload require the same cognitive resources (attention), so that increased mental workload may reduce the individual's ability to maintain SA.

After examining the results of Experiment 1, a question arose as to whether the levels of workload were sufficient to show a statistically significant difference in SA with SHIELD enabled or disabled. We conjectured that there may have been a ceiling effect that prevented the difference in SA scores from reaching significance. Consequently, we designed a second experiment with higher levels of workload to answer this question.

EXPERIMENT 2

Method

Participants. Twenty-six people participated in Experiment 2. Participants included 11 U.S. Army Soldiers and 15 university students. For the Soldiers, the mean age was 31 (SD = 5), while the mean age for students was 19 (SD = 1). The Soldiers averaged 10 years (SD = 7) of military experience, while university students reported no prior military experience. No participants from Experiment 1 participated in Experiment 2.

Apparatus. A net-centric simulation based off the FBCB2 simulation from experiment 1 was presented to participants in experiment 2. The FBCB2 simulation was made more difficult than the one in experiment 1 by adding additional unit types (friendly support units, friendly civilians, unknown civilians). The simulation was presented on a laptop computer. To control for extraneous noise and to convey the auditory cues, participants wore headphones during all experimental and practice trials. For the purposes of this experiment, three Army training scenarios were programmed, a practice scenario and two full-length experimental scenarios, each based on different topographical maps and order of events.

A simulation of SHIELD ran concurrently with the FBCB2 task during the experimental trials. The current experiment employed SHIELD alerts triggered by six significant events; these events were the five reported in experiment 1 and a fire-plan mismatch (alerts when pre-planned target reference points do not match enemy icons). Additionally, SHIELD was altered in that it employed the textual and auditory alerts from experiment 1 as well as additional pop-up pictorial alerts (See Figure 3). Alerts were presented to participants for 5 seconds.

In each FBCB2 scenario, task difficulty, again defined as the number of significant events (i.e., minefield, fire-plan mismatch, etc.) per time interval, was increased every 5 minutes. The practice condition was conducted at a low difficulty condition of 1 event roughly every 25 seconds (12 total events for a 5 minute practice session). In the experimental sessions, the first 5 minute interval represented a low difficulty condition and had on average 1 event every 25 seconds (12 total events). The second 5 minute interval represented the moderate difficulty condition and had on average 1 event every 17 seconds (18 total events). The third 5 minute interval represented the high difficulty condition and had on average 1 event every 8 seconds (36 total events). The order of significant events was randomized for each scenario with the constraint that each type of event appeared equally as often as the other events during each of the five-minute blocks. It is important to mention that the labels "low", "moderate", and "high" are relative to one another, and are all harder than their counterparts in Experiment 1.

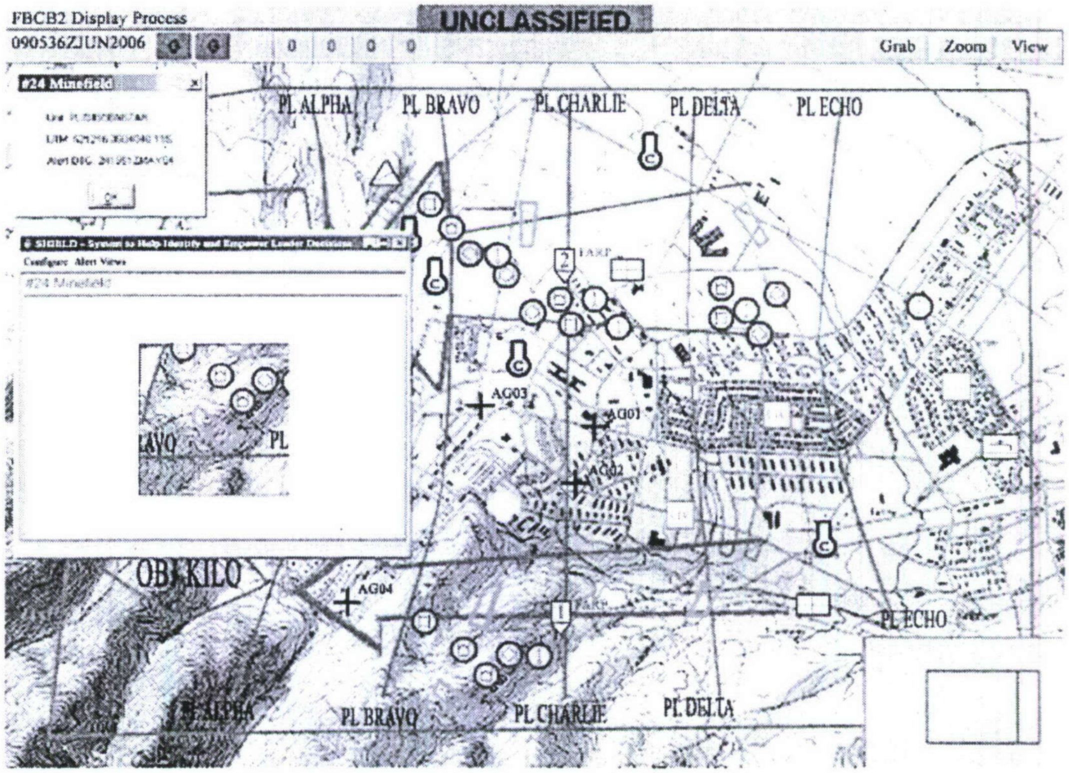


Figure 6. Example pictorial SHIELD alerts.

Participants' SA was measured using SAGAT similar to experiment 1. However, in experiment 2, the SAGAT questions reflecting the more complex environment in experiment 2. In addition, in experiment 2, as well as measuring overall SA, the questions were classified as to what level of SA they measured, based on the three levels of SA suggested by Endsley (1995). Subject matter experts (SMEs) were consulted in order to create levels for the SA questions. Specifically, Level 1 SA consisted of questions reflecting direct observation of visible events (e.g., how many friendly units are on the screen). Level 2 SA questions reflected an understanding of the meaning of events (e.g., recognizing that enemy units appear to be guarding a particular area). Level 3 SA questions reflected being able to predict future events (e.g., which group of units based on past speed will reach a specified objective first). These levels were combined for analysis to provide an overall rating of user performance. The SAGAT tests consisted of 12 questions each and were administered at the end of each 5 minute block. The screen was blacked out while participants completed the questionnaires, forcing the participants to rely solely on recall to complete the surveys. For an example of a SAGAT questionnaire please see Appendix A.

Subjective workload was again measured using a pen and paper version of the NASA-TLX. NASA-TLX questionnaires were administered after the SAGAT questionnaires between the task blocks. As in experiment 1, the exit questionnaire was administered at the end of the experiment (See Appendix B).

Procedure. The procedure was the same as Experiment 1. The only difference was the alteration of the simulation to increase difficulty and the adjustment of the SAGAT questionnaires to address all levels of SA.

Results

First, we examined the equivalence of the two scenarios that were used in the FBCB2 simulation. As a manipulation check participant performance on the two programmed scenarios was examined using a paired samples t-test. The two programmed scenarios did not significantly differ in terms of SAGAT performance (Scenario A: M = 58.13, SD = 8.26; Scenario B: M = 54.97, SD = 8.17 respectively), t(25) = 1.71, p = .10. However, we found a nearly significant difference concerning perceived workload of the two scenarios (Scenario A: M = 54.68, SD = 19.94; Scenario B: M = 50.57, SD = 18.15), t[25] = 2.04, p = .052). However, since the magnitude of the workload difference was relatively small and the presentation of scenarios was counterbalanced, we concluded that this would not affect the results.

To examine SA, we conducted a repeated measures ANOVA, with a 2 (Alert Condition) by 3 (Task Difficulty) design with all within-subject measures. There was a significant effect for task difficulty (See Table 6). Using Fisher LSD post-hoc comparison, we discovered that the results were in the predicted direction and indicated that as task difficulty increased SAGAT scores decreased (See Figure 7). The high difficulty condition had significantly lower SAGAT scores then either the low or moderate difficulty sessions; which did not significantly differ from one another, $p>.05$. Contrary to our hypotheses there were no significant differences in SA scores across the automated alerting conditions (alerts: M = 55.23, SD = 8.46; no alerts: M = 56.54, SD = 8.17; $p>.05$). Nor was there a significant interaction between alert condition and difficulty level ($p>.05$).

Table 6. *Analysis of Variance for SAGAT Performance*

Source	*df*	F	η	*p*
Alerts (A)	1	0.48	.14	.50
Task Difficulty (D)	2	9.92*	.53	<.0005
A * D	2	0.65	.16	.53
Within-group error	34	(.03)		

Note. Values enclosed in parentheses represent mean square errors.*p<.01.

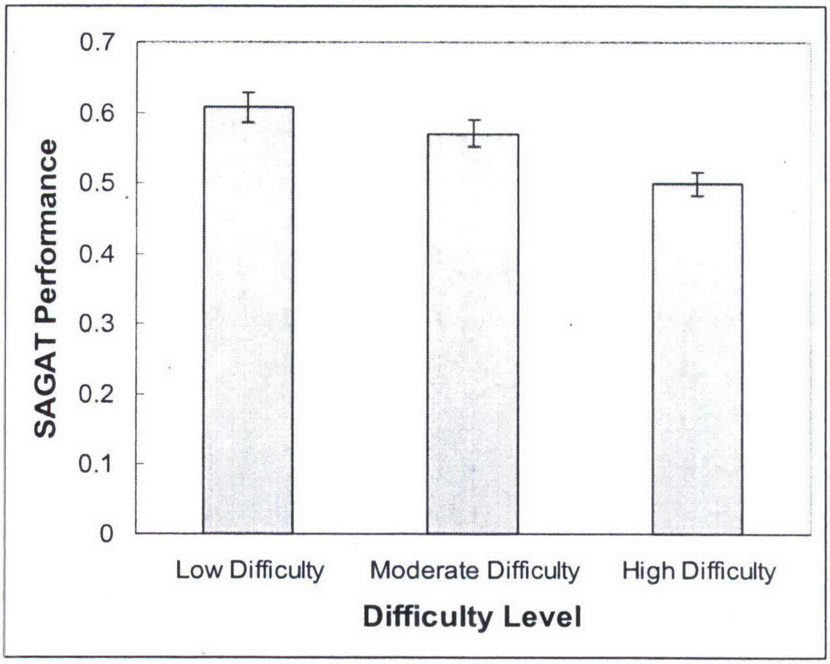


Figure 7. SAGAT scores as a function of task difficulty.

Note: Error bars represent standard deviation.

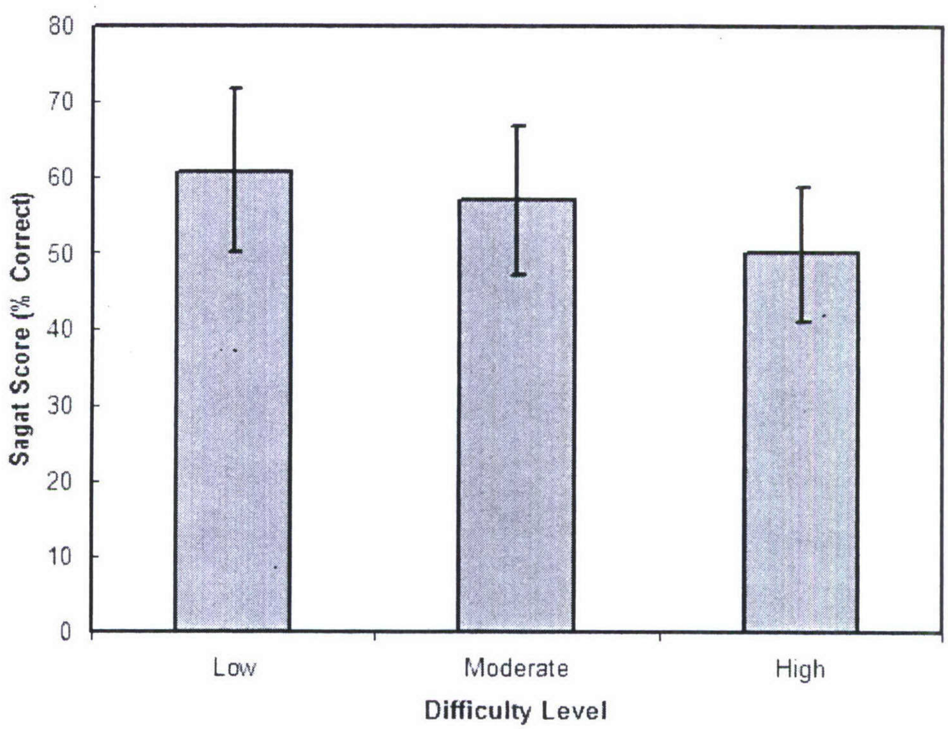


Figure 8. SAGAT scores as a function of task difficulty.

To examine workload performance, we conducted a repeated measures ANOVA, with a 2 (Alert Condition) by 3 (Task Difficulty) design with all within-subject measures. There was a significant main effect for alert condition (See Table 7). Using Fisher's LSD post-hoc comparisons we found that the results were not in the predicted direction. The condition with alerts had significantly lower perceived workload (M = 50.28, SD = 18.47) than the condition without alerts (M = 54.97, SD = 19.57). As expected, perceived workload was related to task difficulty level. Results indicated that as the task increased in difficulty, perceived workload increased (See Figure 8). However, the alert condition by difficulty level interaction was not significant ($p > .05$; See Figure 9). Overall these effects indicate that SHIELD alerts reduce mental workload associated with monitoring an FBCB2 task and that the mental demand associated with monitoring an FBCB2 task (with or without SHIELD) increases with increased task difficulty.

Table 7. *Analysis of Variance for NASA-TLX Workload*

Source	*df*	F	η	*p*
Alerts (A)	1	5.72*	.34	.03
Task Difficulty (D)	2	10.76***	.80	<.005
A * D	2	2.02	.24	.14
Within-group error	50	53.16		

Note. Values enclosed in parentheses represent mean square errors.*p<.05, **p<.01, ***p<.001.

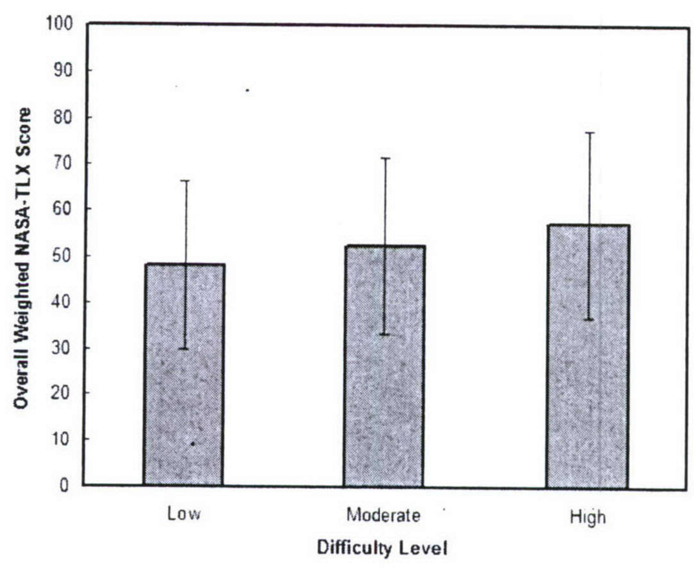


Figure 9. Workload as a function of task difficulty.

Since Experiment 1 demonstrated differences in performance for participants with and without military experience, we conducted an ANOVA with military experience as a between-subject factor for SAGAT performance and mental workload. Analysis including military experience did not indicate any effect of military experience on SAGAT performance (See Table 8). However, there was a trend (p=.07) for military experience to influence perceived workload (See Table 9). Examination of means indicated that there was a trend for students (M = 58.24, SD = 17.33) to demonstrate higher levels of workload compared to Soldiers (M = 44.96; SD = 17.62).

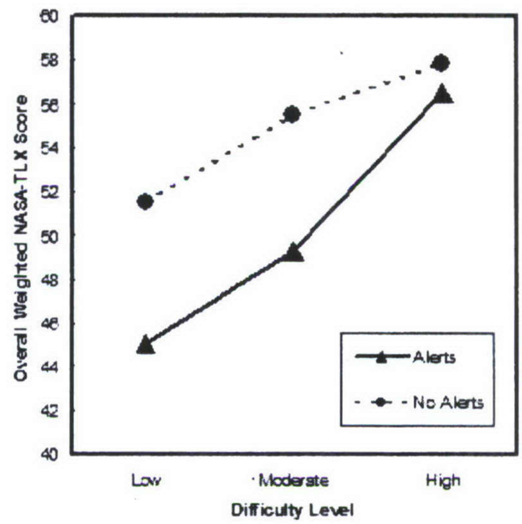


Figure 10. Workload as a function of task difficulty and alert conditions.

In regards to the main effect for task difficulty condition, the results for students and Soldiers were in the expected direction (based on Experiment 1 results) and significant (See Figure 10). However, notice that Soldiers had lower perceived mental workload scores across all difficulty levels, which may be due to their increased familiarity with the NCW task. To further investigate the effect of experience we conducted a 2 (alerts) by 3 (task difficulty) ANOVA separately for students and Soldiers. Results indicated that the effect for alerts did not reach significance for students (alerts: M = 56.78, SD = 16.97, no alerts: M = 59.70, SD =19.16) but was significant for Soldiers (alerts: M = 41.41, SD = 17.31, no alerts: M = 48.50, SD = 19.09). However, the high workloads perceived by the students may have masked the alerting condition effect on perceived workload by decreasing the sensitivity of the measure.

On the exit questionnaire, most of the mean responses were between disagree and agree (see Table 4), thus, the responses were relatively neutral. Scores were on the agree side of the scale for the following statement, "The best use for SHIELD is to reduce the Soldier's workload. With SHIELD active, Soldiers don't have to spend time looking for unsafe situations, they can let SHIELD do that." This illustrates that SHIELD not only affects perceived workload as measured by the NASA-TLX but was also reflected by participant opinion of how SHIELD would affect others.

Table 8. *Analysis of Variance for SAGAT Performance with Military Performance*

Source	*df*	F	H	*p*
		Between Subjects		
Military Experience (M)	1	1.61	.25	.22
		Within Subjects		
Alerts (A)	1	0.91	.19	.35
A * M	1	2.59	.31	.12
Task Difficulty (D)	2	9.01***	.52	<.0005
D * M	2	0.34	.12	.72
A * D	2	0.80	.18	.46
A * D * M	2	0.51	.15	.60
Within-group error	34	(.03)		

Note. Values enclosed in parentheses represent mean square errors. *p<.05, **p<.01, ***p<.001.

Table 9. *Analysis of Variance for Workload with Military Performance*

Source	*df*	F	H	*p*
		Between Subjects		
Military Experience (M)	1	3.68	.36	.07
		Within Subjects		
Alerts (A)	1	6.40*	.46	.018
A * M	1	1.11	.21	.30
Task Difficulty (D)	2	9.67***	.54	<.0005
D * M	2	0.38	.13	.68
A * D	2	2.19	.29	.12
A * D * M	2	0.37	.12	.69
Within-group error	34	(54.53)		

Note. Values enclosed in parentheses represent mean square errors. *p<.05, **p<.01, ***p<.001.

Discussion

Some researchers have expressed concerns that automation may reduce SA by detracting cognitive resources from the task at hand (e.g., Bowers et al., 1996). Our null hypothesis was that alerts had no effect on SA, and in this case we could not reject the null hypothesis. The present research presented no evidence that, given the conditions in the simulated environment, alerts have any affect on SA. Suggestions that alerts can become intrusive and reduce SA were not supported by the results of this experiment. One suggestion that has been raised about this result is that in our experimental environment, participants were able to focus their attention on the primary task of monitoring the display, and that if there had been a secondary task as a distractor, there may have been a measurable affect on SA.

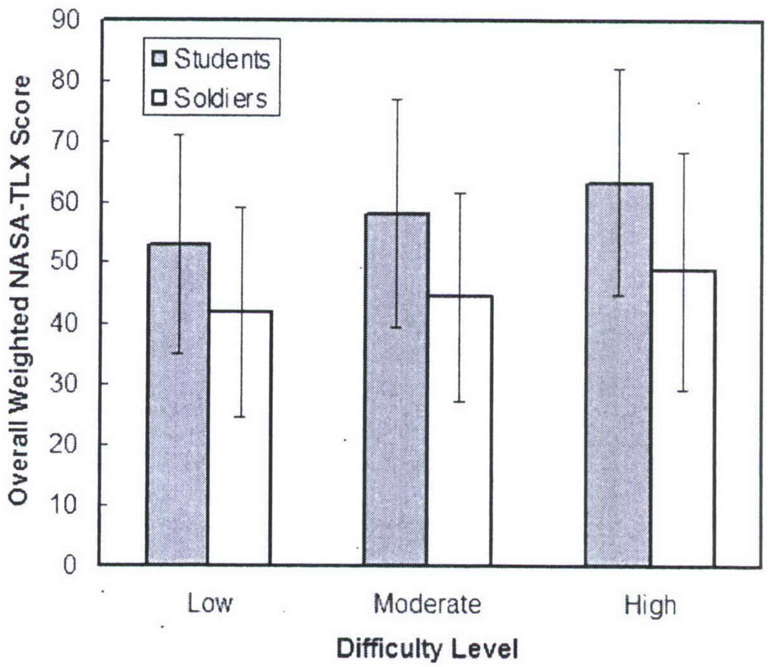


Figure 11. Weighted NASA-TLX workload scores as a function of difficulty level and military experience.

Note: Error bars represent standard deviation.

Another noteworthy finding of the research is that while automated audio/visual alerts did not improve SA performance, these alerts reduced perceived user mental workload. This finding indicates that while the addition of an auditory and visual alerting system increases the physical complexity of the system, it also has the unanticipated effect of decreasing the perceived cognitive complexity of the system.

A possible explanation of these results is that by providing this alerting system to cognitively cue participants to critical events, users may feel that they can rely on SHIELD (as they would a human teammate) to alert them to unsafe situations and thus their personal feelings of responsibility and individual mental demand may be reduced. Thus, they were able to distribute their cognitive workload to another element of the "socio-technical system" (Hutchins, n.d.).

Table 10. *Descriptive Statistics for Exit Questionnaire*

Exit Questionnaire Item	Mean	Standard Deviation
1	3.08	0.80
2	2.54	0.86
3	2.77	0.86
4	2.62	0.90
5	2.46	0.90
6	2.50	0.86

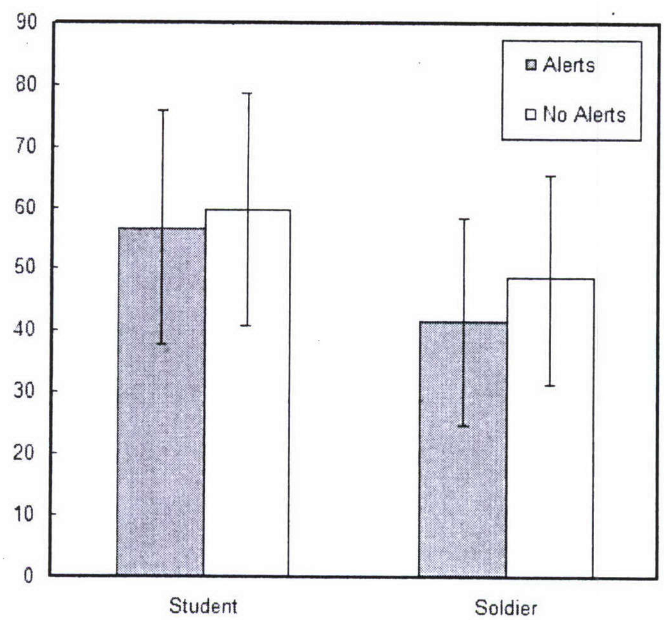


Figure 12. Weighted NASA-TLX workload scores as a function of alert condition and military experience.

Note: Error that error bars represent standard error deviation.

However, it is important to emphasize that the SAGAT questions were not directly related to the automated alerts (e.g., while the system may alert the operator to a new friendly unit, a SAGAT question might query the user on how many friendly units have reached the main objective). Therefore, individuals may subjectively perceive that they have less mental workload but they still maintain the same level of situational awareness.

Originally, we had categorized the SAGAT questions in experiment 2 as to whether they measured level 1, 2, or 3 SA. We wanted to see if automated alerts had an affect on a particular level of SA. However, in view of the fact that alerts seem to have no effect on SA, it is not surprising there was no pattern to how alerts affected levels of SA. Had the results shown a relationship between levels of SA and alerts, it would have been possible to categorize the questions from experiment 1 and analyze those data the same way, however, such an analysis was determined to be unnecessary.

A recommended extension of this research would be to examine the validity of these findings to the field. Of critical importance is to examine the impact of automated alerts on SA when the NCW monitoring task is treated as a concurrent or secondary task (i.e., monitoring the NCW system while concurrently completing another mental and/or physical task). While our current research allowed users to concentrate solely on the NCW task, soldiers in the field are often unable to devote their full attention to the FBCB2 system. Thus, while the current research found no benefit in SA from the alerting system (SHIELD), it may benefit user performance significantly in a distracted environment in which participants are unable to devote their full attention to the task and attentional cueing may be more constructive. This impact may be especially noticeable when the concurrent task relies on the same limited mental resources (i.e., mental demand).

CONCLUSION

The results of both of these experiments suggest that automated alerting systems, such as SHIELD, do not detract from SA even at relatively high levels of workload. On the other hand, it is rather surprising that SHIELD did not facilitate SA. We expected that the automated alerting system would increase SA by directing the operator's attention to situations in the environment that were important to the operator, thus increasing his SA. However, it may be that the types of events the automated alerting system provided alerts for are not events that have an effect on operator SA as we measured it.

The first experiment showed that SHIELD did not have either positive or negative effects on SA at several levels of workload. This may have been due to a ceiling effect, such that workload levels were not intense enough to show a difference. We also considered that our measure of SA was not sensitive enough to measure increases in SA facilitated by SHIELD. For these reasons, a second experiment with higher difficulty levels was conducted, and the SAGAT questions were designed to be more sensitive to the operator's SA.

However, the second experiment showed results similar to the first, even at higher workload levels and with a more sensitive SAGAT. These findings suggest that automated alerting systems like SHIELD do not negatively affect SA, however they do not facilitate SA either.

Even so, the second experiment showed a decrease in mental workload with SHIELD enabled. This suggests that the benefits of automated alerting systems may be realized more in reducing mental workload rather than influencing SA. Considering the complexity of operating a net-centric C3 system while dealing with the situation in a combat environment, a reduction of mental effort may be a real benefit.

REFERENCES

Aiken, D.S., Green, G.E., Arntz, S.J., & Meliza, L.L. (2005). *Real time decision alert, aid and after action review system for combat and training* (ARI Technical Report 1165). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA 437006).

Barnett, J. S., Meliza, L. L. & McCluskey, M. R. (2001). *Defining digital proficiency measurement targets for U.S. Army units* (ARI Technical Report 1117). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (DTIC No. ADA396899).

Billings, C. E. (1997). *Aviation automation: The search for a human-centered approach.* Mahwah, NJ: Lawrence Erlbaum.

Bowers, C. A., Deaton, J., Oser, R. L., Prince, C. & Kolb, M. (1995). Impact of automation on aircrew communication and decision-making performance. *International Journal of Aviation Psychology 5*, 145-167.

Bowers, C. A., Oser, R. L., Salas, E. & Cannon-Bowers, J. A. (1996). Team performance in automated systems. In R. Parasuraman and M. Mouloua (Eds.), *Automation and human performance: Theory and applications.* (pp. 243-263). Mahwah, NJ: Lawrence Erlbaum Associates.

Durlach, P. J. & Chen, J. Y. C. (2003). *Visual change detection in digital military displays.* Paper presented at the Interservice/Industry Training, Simulation, and Education Conference (I/ITSEC), December 1-4, 2003.

Endsley, M. (1995). *Theoretical underpinnings of situation awareness: A critical review.* Proceedings of the International Conference on Experimental Analysis and Measurement of Situation Awareness, Daytona Beach, FL November 1-3 1995. Daytona Beach: Embry-Riddle Aeronautical University.

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley and D. J. Garland (Eds.), *Situation Awareness Analysis and Measurement.* (pp. 147-173) Mahwah, NJ: Lawrence Erlbaum Associates

Fracker, M. L. (1989). Attention allocation in situation awareness. *Proceedings of the Human Factors Society 33rd annual meeting*, 1396-1400.

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

Hutchins, E. (n.d.). *How a cockpit remembers its speeds.* Retrieved 20 September 2007 from the University of California, San Diego Distributed Cognition and HCI Laboratory web site: http://hci.ucsd.edu/lab/hci_papers/ EH1995-3.pdf.

Kantowitz, B. H. & Campbell, J. L. (1996). Pilot workload and flightdeck automation. In R. Parasuraman and M. Mouloua (Eds.), *Automation and human performance: Theory and applications.* (pp. 117-136). Mahwah, NJ: Lawrence Erlbaum Associates.

Morgan, B. B. Jr., Herschler, D. A, Weiner, E. L. & Salas, E. (1993). Implications of automation technology for aircrew coordination and performance. *Human/Technology Interaction in Complex Systems, 6*, 105-136.

Parasuraman, R,. Mouloua, M., Molloy R. & Hilburn, B. (1996). Monitoring automated systems. In R. Parasuraman and M Mouloua (Eds.), *Automation and human performance: Theory and applications*. (pp. 117-136). Mahwah, NJ: Lawrence Erlbaum Associates.

APPENDIX A: EXAMPLE SAGAT QUESTIONNAIRE FROM EXPERIMENT 2

Please answer each of the following twelve questions.
Note that some questions may have multiple answers.

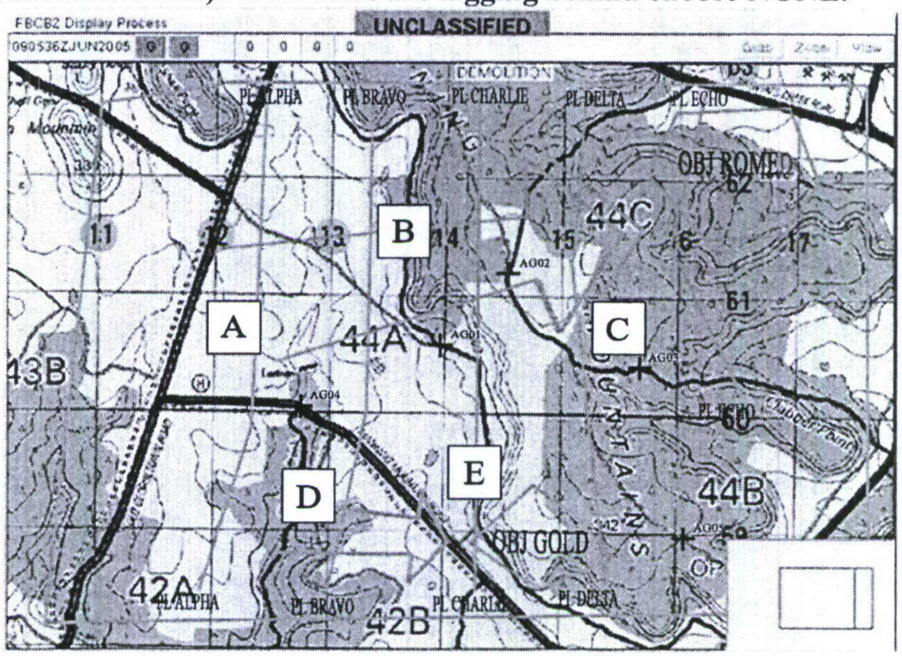
1. **Does it appear that all units are moving cohesively?**

YES NO

2. **How many units appear to be heading on the supporting axis of attack? These are units that have not yet reached the objective.**

0 1-3 4-6 7-9 10+

3. **Do any BLUE units appear to be lagging behind (at least three unit distance behind the next nearest unit)? If no units are lagging behind choose NONE.**



A B C D E NONE

4. **Which <u>objective</u> (OBJ) is currently closer to a Nuclear, Biological, or Chemical (NBC) Contamination?**

OBJ ROMEO OBJ GOLD ABOUT EQUAL

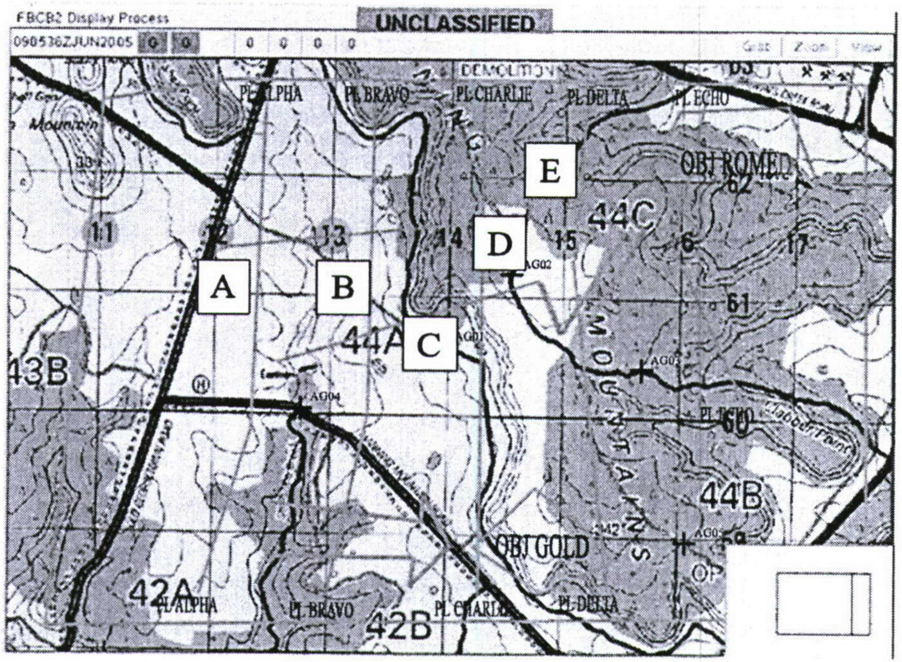
5. **Are any BLUE units within one unit distance of an enemy unit?**

YES NO

6. **How many MINEFIELDS are currently on the display?**

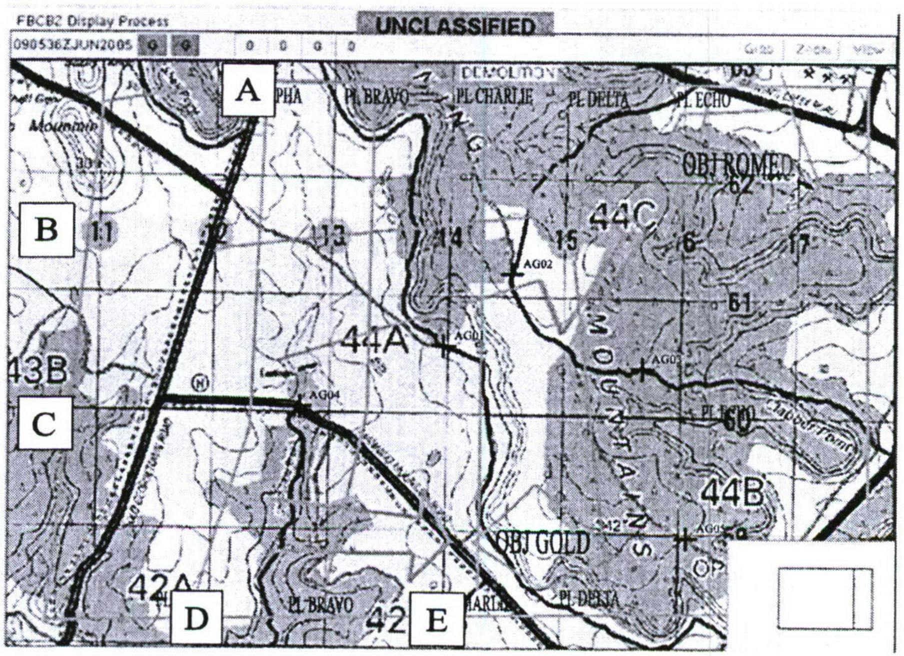
0 1 2 3 4 5 6 7+

7. **Where might you expect the new enemy units to move to immediately block the main axis of advance?**



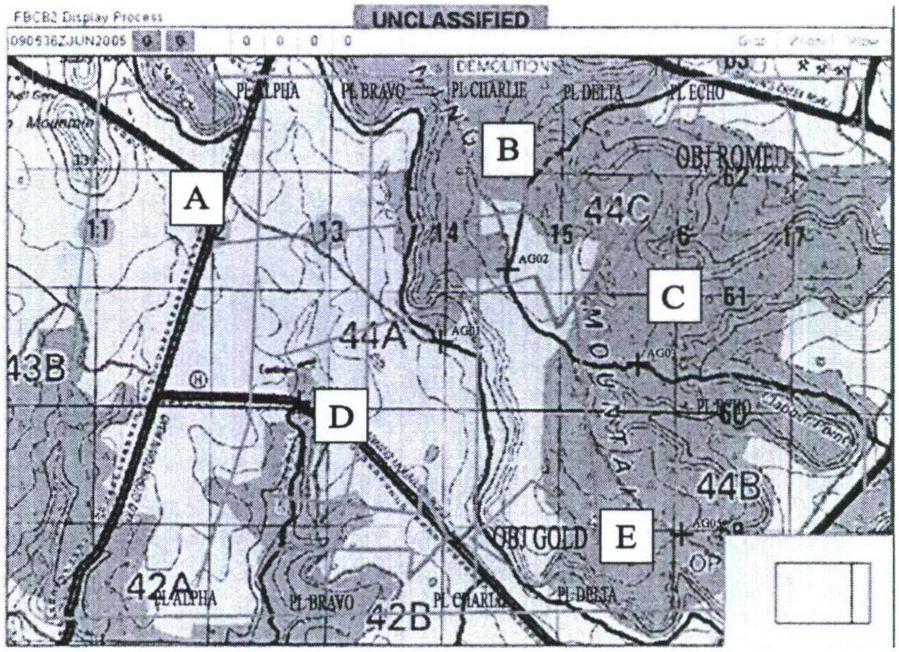
A B C D E

8. **On the map above, which letter is closest to the most recent Boundary Violation (if no Boundary Violation has occurred circle NONE)?**



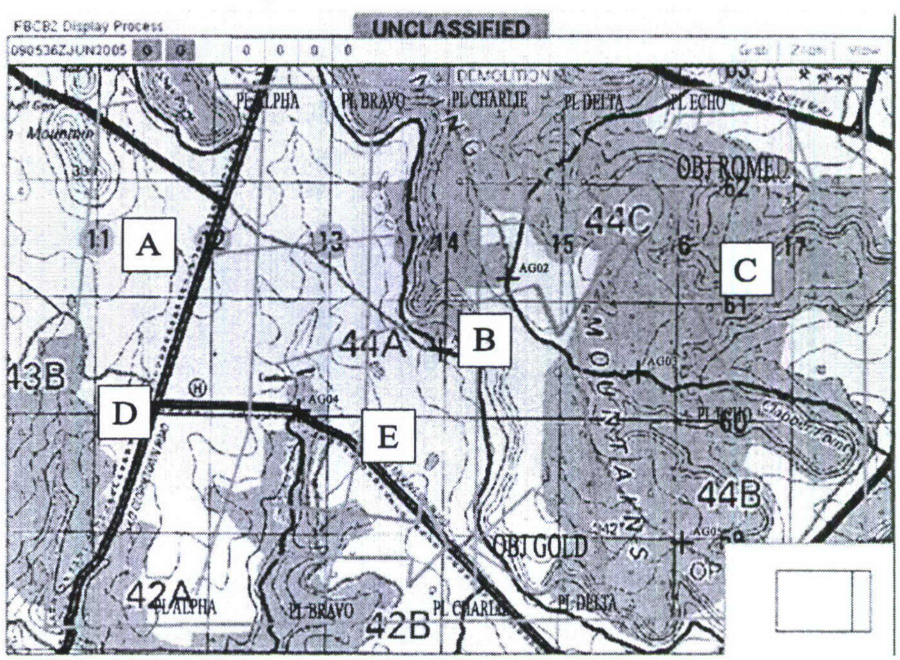
A B C D E NONE

9. **Where might you expect the new enemy units to appear?**



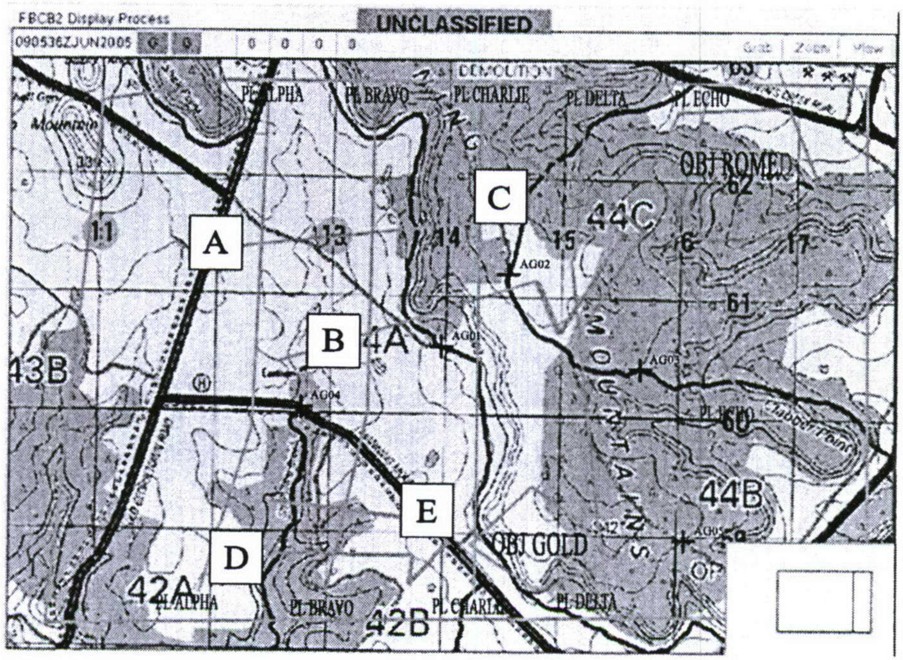
A B C D E

10. **Where might you expect the enemy to place artillery forward observers? (Choose a letter from the map below).**



A B C D E

11. On the map above, which letter is closest to the most recent <u>NEW MINEFIELD</u> (if no minefields have appeared circle NONE).



A B C D E NONE

12. Are any BLUE units in danger of encountering, within one unit distance of, a NBC contamination area?

YES NO

APPENDIX B: EXIT QUESTIONNAIRE

Exit Questionnaire

For each of the following statements, circle whether you **STRONGLY AGREE**, **AGREE**, **DISAGREE**, or **STRONGLY DISAGREE** with the statements.

1. "The best use for SHIELD is to reduce the Soldier's workload. With SHIELD active, Soldiers don't have to spend time looking for unsafe situations, they can let SHIELD do that"

Strongly Disagree	Disagree	Agree	Strongly Agree

2. "The best use for SHIELD is to train Soldiers to notice unsafe situations themselves"

Strongly Disagree	Disagree	Agree	Strongly Agree

3. "Soldiers should learn to pay attention to what's going on and not rely on SHIELD alerts"

Strongly Disagree	Disagree	Agree	Strongly Agree

4. "The more SHIELD alerted me to unsafe situations, the more I learned to spot those situations myself"

Strongly Disagree	Disagree	Agree	Strongly Agree

5. "Since I knew SHIELD would alert me to certain unsafe situations, I did not look for those situations myself and let it do it's job"

Strongly Disagree	Disagree	Agree	Strongly Agree

6. "If Soldiers get use to having SHIELD, they will stop looking for unsafe situations themselves"

Strongly Disagree	Disagree	Agree	Strongly Agree