ARI Research Note 97-25

Examining the Effects of Communication Training and Team Composition on the Decision Making of Patriot Air Defense Teams

Leonard Adelman
George Mason University

Matthew Christian
System Planning Corporation

James Gualtieri
Enzian Technology, Inc.

Terry A. Bresnick
Innovative Decision Analysis

Research and Advanced Concepts Office
Michael Drillings, Chief

August 1997

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Director

Research accomplished under contract
for the Department of the Army

George Mason University

Technical review by

Joseph Psotka

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED (from. . . to)
1997, August	Final	June 1995-January 1996

4. TITLE AND SUBTITLE

Examining the Effect of Communication Training and Team Composition on the Decision Making of Patriot Air Defense Teams

5a. CONTRACT OR GRANT NUMBER
MDA903-92-K-0134

5b. PROGRAM ELEMENT NUMBER
0601102A

6. AUTHOR(S)

Leonard Adelman (George Mason University), Matthew Christian (System Planning Corporation), James Gualtieri (Enzian Technology, Inc.), and Terry Bresnick (Innovative Decision Analysis)

5c. PROJECT NUMBER
B74F

5d. TASK NUMBER
2901

5e. WORK UNIT NUMBER
C01

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Dr. Leonard Adelman
c/o Department of Operations Research
George Mason University
Fairfax, VA 22030-4444

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
ATTN: PERI-BR
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. MONITOR ACRONYM
ARI

11. MONITOR REPORT NUMBER
Research Note 97-25

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

13. SUPPLEMENTARY NOTES
COR: Michael Drillings

14. ABSTRACT *(Maximum 200 words)*:

An experiment investigating the effect of communication training and four group composition variables was performed with Patriot air defense teams for two different types of aircraft identification tasks. It was predicted that communication training would significantly enhance communication quantity and quality and, in turn, team performance for both tasks. Although the training did sometimes improve team communications processes, it did not improve team performance. The variable that had the biggest positive effect on communication quality and team performance was the number of hours a team had worked together. This effect was only found, however, for the type of task for which Patriot teams routinely train. It did not transfer to the less frequent and more cognitively stressing task where there is conflicting information about unknown aircraft, as in the U.S.S. Vincennes tragedy.

15. SUBJECT TERMS

Judgment heuristics	Biases	Communications training
Group composition	Army air defense	

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT	17. ABSTRACT	18. THIS PAGE			
Unclassified	Unclassified	Unclassified	Unlimited	59	

EXAMINING THE EFFECTS OF COMMUNICATION TRAINING AND TEAM COMPOSITION ON THE DECISION MAKING OF PATRIOT AIR DEFENSE TEAMS

CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Examining the Effects of Communication Training and Team Composition on the Decision Making of Patriot Air Defense Teams

INTRODUCTION

Previous research with Army air defense operators found significant order effects for difficult identification tasks. In particular, order effects were found for tasks with conflicting information about an incoming aircraft (e.g., Adelman & Bresnick, 1992; Adelman, Bresnick, Black, Marvin, & Sak, in press; Adelman, Tolcott, & Bresnick, 1993). Conflict occurs when some information or "cues" indicate the aircraft is friendly and other information indicates it is hostile. An order effect occurs when different identification judgments or different engagement decisions are made depending on the sequence in which the same information or cues are presented to the operator.

It is possible that one ordered sequence is more indicative of a hostile (or friendly) aircraft than another. If true, the order effect would be justified. It would not represent a cognitive bias. Unfortunately, such empirical data do not exist. The best one can do is examine training protocols and the target classification algorithm in the Patriot air defense system. According to both sources, operators should combine information using an additive combination rule; consequently, the order of the information should not affect their judgment. The fact that it did suggests a performance problem for tasks where there is conflicting information about an unknown aircraft.

Adelman et al. (1995) found significant order effects with Patriot air defense teams, as well as with individual operators. A Patriot team consists of one Tactical Control Officer (TCO) and one Tactical Control Assistant (TCA). The TCO is responsible for all identification judgments and engagement decisions. The TCA is responsible for launching missiles at the tracks that are to be engaged by Patriot. In addition, the TCA aids the TCO by supplying information about incoming tracks.

A series of multiple regression analyses found that the size of the team's order effect was related to the team's communication processes (Adelman et al. 1995). For example larger order effects were related to the TCO disagreeing with the TCA significantly more, and having significantly more opinions when making engagement decisions. In contrast, TCAs gave significantly less information for the engagement decision.

Although the general result of greater communication by the officer (TCO) and less by the enlisted soldier (TCA) was obtained for all three dependent measures, disagreement by the TCO was the only process measure that was significant for all three dependent measures (i.e., probability judgments, identification judgments, and engagement decisions). Moreover, the R^2 was smaller than desired for all three equations, accounting at best for only 18% of the variance in the engagement decisions. Consequently, we thought that future research needed to rely on experiments that manipulated communication processes to better assess their impact on the decision making performance of Patriot teams.

Previous research with experienced personnel in other domains has found a relationship between a team's communication processes and their performance. For example, in their review, Orasanu and Salas (1993) report a positive relationship between the amount (and type) of communication and the performance of commercial flight crews. In their review of studies with operational Navy teams, McIntyre, Morgan, Salas, and Glickman (1988) found that good teams exhibited greater incidence of communication, cooperation, and coordination than poor teams. And research by Leedom and Simon (1995) found a positive relationship between communication training and the performance of Army aircrews. Consequently, we hypothesized that communication training would also improve the communication processes and performance of Patriot teams.

However, there were a number of issues regarding (a) the composition of the teams, and (b) the tasks they perform that we thought were important to consider. First, when there were few tracks on the Patriot screen, but conflicting information about them, we predicted that the effect of communication training would depend on the prior accuracy of the TCO vs. TCA. The "few track/cue conflict" task does not occur frequently, and training on it is not routine. However, this task does represent realistic situations. For example, when the U.S.S. Vincennes shot down an Iranian airliner, there were only two aircraft tracks on its display. During the Gulf War, there was often only one track on the Patriot display at any time. Although these tracks were high-speed tactical ballistic missiles automatically engaged by Patriot, many aspects of the identification and decision making process were similar to those used in the "few track/cue conflict" task of the experiment.

The TCO is reponsible for the Patriot team's identification judgments and engagement decisions. If TCOs combine information using an additive rule with the correct weights, then their identification judgments and engagement decisions will match those of the Patriot system in a fully automated mode, and there will be no order effects. [Note that ground truth data do not exist for the "few track/cue conflict" scenarios we developed for examining order effects using the Patriot air defense simulator.]

If the TCO is more accurate than the TCA, the TCO does not need the TCA's help for the "few track/cue conflict" task, for there is ample time to make all decisions and take all actions. Moreover, communication will not improve performance because the TCO is already using the correct judgment and decision rules. In fact, it might even hurt performance because the more accurate TCO is listening to, and perhaps may be persuaded by, the less accurate TCA.

However, in cases where the TCA is more accurate than the TCO, we predicted that communication training would improve team performance. Previous research found a considerable range in the accuracy scores of both TCOs and TCAs for tracks with conflicting information. We predicted that performance would improve after the communication training because training would open communication channels between the TCO and the TCA, allowing the TCA's expertise to impact the decision. This prediction rests on three assumptions: (1) that, prior to training, there would be room for improving communication processes and team performance; (2) that communication training would improve communication processes; and (3) that such improvement would translate into better team performance.

We also predicted that communication training would improve the communications processes and level of team performance for the "many track/no conflict" task. In this task, there are many aircraft (e.g., ten or more) on the Patriot display at the same time. The tracks have no conflicting information; consequently, performance does not depend on the TCO's or TCA's accuracy in combining information. Instead, performance depends on the flow of information and coordination of activities under high time pressure. We predicted that communication training would improve the information flow and coordination between the TCO and the TCA and, thus, team performance for the "many track/no conflict" task, regardless of the TCO's and TCA's accuracy for the "few track/cue conflict" task.

The second team composition variable was the number of hours team members had worked together in the past. Having set teams, so that personnel routinely work together, is referred to as "battle rostering" by the Army. It is believed that (a) the more team members work together, the better their communication and coordination and, in turn, (b) team performance. This position is supported by empirical research reported in McIntyre et al. (1988) and Leedom and Simon (1995).

Therefore, we predicted that teams that had worked together longer would perform better, but only for the "many track/no conflict" task. This is the type of task that Patriot operators primarily train on; they seldom train on the "few track/cue conflict" task. Therefore, even members of teams that have been together longer may not (1) know how their teammate processes conflicting information, or (2) have a good procedure for communicating this information. Consequently, we did not predict that teams that had worked together longer would affect team performance on the "few track/cue conflict" task.

We predicted that communication training would improve performance for both tasks. However, for the "many track/no conflict" task, we predicted a larger positive effect for the teams that had less time working together. Again, this is the task on which Patriot teams routinely train. Consequently, we predicted that teams that had less time together would have more to gain from communication training.

In contrast, we predicted that the positive effect from communication training would be the same, regardless of the amount of time that teams had worked together, for the "few track/cue conflict" task. Again, this is the task for which Patriot teams seldom train. Consequently, the amount of time the team had worked together was predicted to be irrelevant to their performance on this task. The hypothesized benefit to be derived from clarifying how teammates do (and should) combine conflicting information to make their identification judgments and engagement decisions was predicted to be the same because even teams that had been together a long time would not have developed the "shared mental models" (Cannon-Bowers, Salas, & Converse, 1993) necessary for performing this task well.

The third team composition variable was good versus poor communicating teams. In the above discussion, it was assumed that there would be a positive relationship between the amount of time that a team had worked together and the quantity and quality of its communications. This may not be true. Consequently, another way to categorize Patriot teams was in terms of the quantity and quality of their communications prior to communication training.

The predictions for this third composition variable were analogous to those made for the second. Specifically, we predicted that better communicating teams, both in terms of communication quantity and quality, would have better team performance for both tasks. In addition, we predicted that communication training would have a positive effect for both tasks. However, we predicted that this effect would be bigger for the poorer communicating teams for the "many track/no conflict" task because there would be much more room for improvement.

In contrast, we predicted that communication training would result in equivalent levels of improvement in team performance for the "few track/cue conflict" task because participants get less training on this type of task. An implicit assumption in this last hypothesis is that even the better communicating teams do not communicate well when performing the "few track/cue conflict" task; consequently, there is ample room for improving team communication processes and, in turn, team performance.

In summary, the experiment described herein tests the general hypothesis that communication training will improve the communication processes and, in turn, the performance of Patriot teams. Laboratory support for this hypothesis can be found in a number of sources (e.g., Adelman, Zirk, Lehner, Moffett, & Hall, 1986; Gualtieri, 1994; Hackman & Morris, 1975). Support also can be found in actual environments outside the laboratory (e.g., Foushee & Helmreich, 1988; Leedom & Simon, 1995; McIntyre et al., 1988). However, we predict that the generality of this hypothesis depends on (1) the type of task ("few track/cue conflict" or "many track/no conflict"), and (2) team composition factors (TCO vs. TCA accuracy, how long the team has worked together, and the quantity and quality of team communications).

METHOD

This part of the paper is divided into five sections: (1) participants and research site; (2) procedural overview, (3) tasks, (4) independent variables; and (5) dependent variables.

Research Site and Participants

The experiment was conducted in two sessions using the Patriot training simulators in the Directorate of Training Development at Ft. Bliss, Texas. The sessions were conducted on October 12-13, 1994, and January 11-12, 1995.

Twenty-eight teams, consisting of one tactical control officer (TCO) and one tactical control assistant (TCA) participated in the experiment. Only 32 teams were possible for data collection; therefore, the 28 teams represent a high percentage of the total teams available.

Mean experience with the Patriot air defense system was 22.39 months for TCOs and 47.21 months for TCAs. The ranges were 0 to 78 months and 1 to 96 months, and the standard deviations were 15.59 and 27.91, respectively. Seven team members were not "TCO/TCA Qualified" at the time of the experiment, but all had completed the basic course.

Procedural Overview

Each experimental session was constrained to be four hours in length in order to obtain participation of Patriot personnel. And, due to the number of available Patriot simulators, sessions were limited to at most four TCOs and four TCAs. Teams were formed from the TCOs and TCAs attending a session.

Each session began with introductory remarks. Then the participants independently evaluated aircraft tracks presented using overhead transparencies. This was done to form teams varying in TCO vs. TCA accuracy. Then, each team evaluated four scenarios using the Patriot simulator. In particular, they performed one "few track/cue conflict" scenario and one "many track/no conflict" scenario before communication training, and then another scenario for each type of task after training. The session ended with a debriefing and question-and-answer period.

Procedurally, it is important to note that a post Gulf War context was presented to provide an operating context for the session. However, the displays used during the session provided no true geographic information. The situation description detailed the mission of the operators, the rules of engagement, and the parameters of the friendly safe passage corridors, including speed and altitude.

The situation description further indicated that teams were to work in a decentralized, fully autonomous mode; that is, without communications with headquarters or other units. By being in autonomous mode the teams were solely responsible for their identification and engagement decisions. Finally, the general situation description noted that there were equivalent chances of encountering friendly aircraft as hostile aircraft to make these prior probabilities equivalent. [Note: The same situation description was used in Adelman et al., 1995.]

Tasks

There were two types of air defense tasks: a "few track/cue conflict" task and a "many tracks/no conflict" task. In addition, there were two scenarios for each type of task so that each team could perform each type of task before and after training. The scenarios for each type of task were similar and counter-balanced so that half of the teams performed one scenario before training and the other half performed the other scenario. In addition, we counter-balanced the number of times each scenario was performed before and after the briefing, and the order in which the scenarios were performed. No team did a task scenario twice.

"Few Track/Cue Conflict" Tasks. In this task, there were one to at most three incoming aircraft tracks on the Patriot screen at a time. The sequence was always the same. First, a track would appear over enemy territory, cross into friendly territory, and continue in friendly territory until it left the screen. Then, after a track was about half-way into friendly territory, the next track would appear on the screen. This sequence ensured that the Patriot teams were busy, but still had plenty of time to examine each track before making their engagement decisions.

There were 19 tracks in one scenario and 15 in the other. Most of these tracks had conflicting information; that is, some of the information suggested that the aircraft was friendly and some of it suggested that it was hostile. The analyses presented herein only focus on the teams' responses to the 8 test tracks that were created by systematically manipulating the order in which information appeared to the Patriot teams. The remaining tracks were essentially "fillers" so that participants couldn't determine which tracks were the test tracks.

Three independent variables were used to create the characteristics of the 8 test tracks. The first variable was whether the initial piece of information about an aircraft indicated that it was friendly or hostile. The second variable was whether the aircraft originated from the left-hand or right-hand side of the display. And the third variable was the order in which confirming and disconfirming information appeared late in a track's history.

By "late in a track's history" we mean the last two pieces of information about a track before the teams had to make their decisions. For Late Order #1, teams received information that confirmed the initial piece of information and, then, information that disconfirmed it. This was the confirm-disconfirm (CD) sequence. For Late Order #2, the opposite ordered sequence was used; that is, disconfirming then confirming information (DC). In order to ensure that all teams received all information about a track, the teams were asked not to actually engage the track by launching missiles. Instead, the TCO was asked to write-down the team's identification and engagement decision at a specific "decision point" for that track.

It is important to note that the initial hypothesis about the aircraft is what is being confirmed or disconfirmed. For example, if the first piece of information, or cue, about an aircraft suggests that it was a friend, then for Late Order #1 (CD), the next-to-last cue was friendly (i.e., indicative of a friendly aircraft) and the last cue was hostile. In the case where the first piece of information was indicative of hostile, the two cues were reversed for Late Order #1, so that a hostile cue was followed by a friendly cue in the CD sequence. Since we crossed the three independent variables to create the eight test tracks, the two tracks that (1) had the same initial information and (2) originated at the same place on the screen, were identical (i.e., same information and same flight path) except for the order in which the last two pieces of information were presented to the Patriot teams.

<u>"Many track/no conflict" task</u>. Both scenarios for this task

had been created previously for training purposes and represented "school house training scenarios." They were selected through discussions with Army air defense training personnel at Ft. Bliss, and represented the kind of scenario for which Patriot teams routinely train.

In both scenarios, there were multiple waves of enemy aircraft moving toward the Patriot battery or assets it was protecting. The enemy aircraft were flying at different altitudes and speeds, and in various formations, depending on the tactics being implemented in the scenario. In addition, there were friendly aircraft moving to engage the incoming enemy aircraft. Consequently, there were seldom less than ten aircraft on the screen at any given time, and often more than twenty or thirty. The teams were permitted to launch missiles to engage aircraft at any time they wanted for these scenarios.

In all cases, the aircraft exhibited no conflicting information. All friendly and all enemy aircraft were doing exactly what an air defense operator would expect them to do. What makes this kind of task difficult, however, is its sheer scope; that is, one Patriot battery monitoring and responding to a situation where there are so many enemy aircraft implementing coordinated tactics designed to pierce friendly defenses. This kind of task is representative of a Soviet-like air threat, and the kind of situation for which Patriot was originally designed.

<u>Independent Variables</u>

There were five independent variables: (1) communication training, (2) TCO vs TCA accuracy, (3) number of hours teams had worked together, (4) amount of communication prior to communication training, and (5) quality of communication prior to communication training. Each is considered, in turn.

Communication Training. This independent variable was operationalized via a briefing emphasizing the importance of communication processes for good team performance. The briefing began by mentioning the positive findings for previous research with Army air crews and Patriot teams, but did not discuss order effects.

General principles of good communication were discussed, and then specific process behavior was recommended for each type of task, consistent with the performance needs of that task. The behavior for the "few track/cue conflict task" was directed toward use of a global, additive processing rule to deal with the conflicting information about the incoming aircraft track. It included the following recommendations:

* Since time is available, describe how you are making your judgments.
* Consider the entire track history when making judgments.
* Talk about each piece of information about the track and why it is important or not.
* Consider the weight set that the Patriot system is using.

In contrast, there was no conflicting information about the tracks in the "many track/no conflict" task; there were just a lot of tracks and limited time to deal with them. Consequently, the behavior recommended for this task emphasized highly coordinated actions and information exchanges.

Time constraints and the small number of available Patriot teams limited the duration of the briefing to ten minutes and eliminated the opportunity to give the teams practice sessions. As previously mentioned, we had only four hours to complete all work with a team. In addition, because of the small number of teams, we wanted each team to perform a scenario for each of the two types of tasks before and after the briefing so that we could use a within-subject ANOVA to increase our statistical power for team composition variables that could not be manipulated easily. Although we knew that we had a weak training manipulation, we considered it strong enough to affect the teams' communication processes based on the findings in Adelman et al. (1995).

TCO vs. TCA Accuracy. This independent variable was based on participants' accuracy for 34 aircraft tracks presented using overhead transparencies at the beginning of the four-hour session. The 34 tracks were the "test tracks" in the two "few track/cue conflict" scenarios that would be presented later using the Patriot simulator. Overhead transparencies were used instead of the simulator to save time and increase experimenter control by presenting only one track at a time.

Participants worked independently when evaluating the tracks. At the decision point for each track, participants indicated whether they would (1) identify the aircraft as friendly, hostile, or unknown, and (2) engage the aircraft or not. The correct answer was the Patriot system's automatic engagement response for the track; that is, whether the system recommended that the track be engaged or not.

Since we created the tracks, the system's response was the most readily available measure of accuracy. The only alternative would have been to have Army air defense evaluation personnel evaluate the tracks, and that was not feasible. Higher scores meant that participants were using the Patriot's additive weight set to resolve the conflicting information about the track. Since this is what Patriot personnel should normally do, we considered it to be an acceptable measure of decision accuracy.

TCO vs. TCA accuracy, like all other group composition variables, had to be created from the four TCOs and four TCAs that attended each four-hour session. We had no control over the four TCOs and four TCAs that were scheduled to participate in each four-hour session. In addition, we were not permitted to evaluate the participants' accuracy for test cases prior to the four-hour session. We simply had to do the best we could at creating teams that varied on the group composition variables when they arrived at each session.

These conditions constrained our ability to create teams where there was a large difference in the TCO's versus the TCA's accuracy. The statistical characteristics of this independent variable are presented in Table 1. The mean difference for teams where the TCO was more accurate was 4.47. The mean difference for teams where the TCA was more accurate was 2.85. These differences were much smaller than desired. However, in all cases, the TCO was more accurate in the TCO More Accurate condition, and the TCA was more accurate in the TCA More Accurate condition. Examination of Table 1 indicates that mean accuracy level for the two levels was only 63% (i.e., [(22.267 + 20.462)/2]/34), leaving ample room for improvement in performance.

- - - - - - - - - - - - - - - -

Insert Table 1 about here

- - - - - - - - - - - - - - - -

Number of Hours Worked Together. Due to battle rostering, TCOs and TCA that routinely worked together often arrived together at the experimental session. When forming the teams we tried to keep arrival information in mind. However, we did not ask team members how long they had worked together until a convenient point after completion of the first "few track/cue conflict" task.

The number of hours team members had worked together ranged from 0 to 1500 hours, with a mean of 229 hours. We created two levels on this variable. The seventeen (17) teams that had worked together less than 200 hours comprised the first level; the eleven (11) teams that had worked together more than 200 hours comprised the second level. Table 2 presents the statistical properties of the "Number of Hours Worked Together" variable.

- - - - - - - - - - - - - - - -

Insert Table 2 about here

- - - - - - - - - - - - - - - -

Communication Quantity. This independent variable was created by calculating the total number of communications for each team, for each task, before the interaction briefing. Communications were coded by listening to audio tapes of the teams' communications made during each of the simulator sessions.

A given communication was noted under one of the following seven communication categories: (1) Give Information, (2) Ask for Information, (3) Give Opinion, (4) Ask for Opinion, (5) Give an Order (only for the TCO), (6) Ask for an Order (only for the TCA), and (7) Express Disagreement. These represented a slightly more refined list of categories than those used in Adelman et al. (1995), which had been based on Bales' Interaction Process Analysis (1950) and Nieva, Fleishman, and Rieck's (1978) Communication Process Topology.

The tapes for two of the 28 teams could not be coded due to background noise in the simulator room; consequently, only 26 teams were coded on this variable. The communications for the "few track/cue conflict" task were coded separately from the communications for the "many track/no conflict" task. Each track in the "few track/cue conflict" task was coded individually. In addition, for the "many track/no cue conflict" task, the number of communications were normalized based on the time length of the scenario.

To create two levels on this variable, we took a median split on the total number of communications for each team before the briefing, across all of the communication categories. Again, this was done for each task. Table 3 presents the statistical properties of the amount of communication variable.

- - - - - - - - - - - - - - - - -

Insert Table 3 about here

- - - - - - - - - - - - - - - - -

In general there was high agreement in the two median splits. Only six of the 26 teams had differing levels for the two tasks. For the 26 teams, the correlation between the total number of communications for the two tasks, before the briefing, was 0.39. This was significantly different than zero at the $p = 0.048$ level. These results suggests that the teams' relative amount of communication did not vary much as a function of the task prior to the briefing. Nevertheless, we chose to use the median splits for each task because they were the best measures of the amount of communication, prior to training, for that task.

It is important to note that the correlation between the number of hours team members had worked together and the total number of communications was 0.40 for the "few track/cue conflict" task, and 0.16 for the "many track/no cue conflict" task. The first correlation is significantly different from zero ($p = 0.046$); the second is not. This suggests that the variables were related, but different.

Communication Quality. Army air defense evaluation personnel were not available to make judgments about the quality of each team's communication processes. Consequently, we had to develop measures of communication quality. We addressed this issue by using a median split on the sufficiency scores for each task.

"Sufficiency scores" were originally developed by Gualtieri (1994) to measure the quality of a team's interaction on different functional processes. For each functional process, a team's received a sufficiency score equal to 1.0 if that team scored one standard deviation above the mean on that process; otherwise, they scored 0 on that process. A cumulative sufficiency score was developed for each team by summing the sufficiency scores over all the processes. Gualtieri found that teams with higher cumulative sufficiency scores had, in turn, high performance scores on a decision making task.

We used the same basic approach. Specifically, sufficiency scores were calculated for each task prior to the briefing using a four-step process. First, we calculated the mean and the standard deviation of the communications separately for the TCO and the TCA for each of the seven general categories. This resulted in 12 mean values because only the TCO gives orders and only the TCA asks for orders. Second, we added the standard deviation to the mean for each of the twelve categories to create the sufficiency score standard for that category. Third, we compared the individual team values to the sufficiency score standards. A team received a point for each category for which its number of communications exceeded the one standard deviation standard. And, fourth, the points were summed to calculate the cumulative sufficiency score for each team. This score could be between 0 and 12, inclusive.

A median split of the sufficiency scores was taken for each task to define high and low quality communicating teams prior to the briefing. The statistical properties of this variable are shown in Table 4 for each task. The correlation for the 26 teams' sufficiency scores was 0.43 for the two tasks. Although this was significantly different than zero at the $p = 0.028$ level, it clearly did not approach a correlation of 1.0. Therefore, we chose to use the separate median splits for each task.

Insert Table 4 about here

The quantity measure (i.e., number of communication) and quality measure (i.e., sufficiency score) were related, but not identical. The correlation was 0.53 ($p = 0.006$) for the "few track/cue conflict" task, and 0.37 ($p = 0.063$) for the "many track/no cue conflict" task. Again, the correlations did not approach 1.0. Moreover, as will be shown below, they result in somewhat different findings. Consequently, we kept both variables, calling one communication quantity and the other communication quality.

Correlations between the number of hours worked together as a team and the sufficiency scores for the "few track/cue conflict" task and the "many track/no cue conflict" task were 0.24 and 0.42 respectively. The first correlation is not significantly different from zero; the second is ($p = 0.03$). As in the case with the amount of communication variable, these correlations suggest that the variables are related, but different. Consequently, we did separate analyses for each one.

Dependent Variables

There were two types of dependent variables in the experiment: 1) communication process variables, and 2) task performance variables. The general hypothesis was that the independent variables would positively affect the communication process variables and, in turn, the performance variables. Each type of dependent variable is considered briefly, in turn.

Communication Process Variables. There were two communication process variables for each of the two tasks: (1) the number of communications (i.e., communication quantity), and (2) the sufficiency score (i.e., communication quality). As described in the Results section, these process measures were used as dependent measures for all independent variables, even communication quantity and quality, in order to assess significant interactions involving comunication training.

Performance Variables. There were two performance variables (or measures), one for each type of task. Specifically, agreement with the Patriot system's recommendation was used to measure performance for the "few track/cue conflict" task. As we discussed in the section on TCO vs. TCA Accuracy, agreement between a team's and the system's recommendation was both a feasible and acceptable measure of performance because that is what Army air defense personnel should do in most situations. Performance was only measured for the eight test tracks in each "few track/cue conflict" scenario. Higher scores meant that the teams agreed with a higher percentage of the system's recommendations for the eight tracks.

The Patriot simulator calculated a "area defense score" for the "many track/no conflict" scenarios. Again, these scenarios were developed by training personnel at Ft. Bliss, and the simulators were programmed to calculate this score. The area defense score was the percentage of hostile aircraft that penetrated friendly territory that were shot down by the Patriot team, corrected for the number of missiles that were launched and failed to hit their target. Since the scenarios were designed so that there was a small, randomly generated probability that a missile would miss its target, it was often appropriate to fire multiple missiles to ensure that hostile aircraft are destroyed. Consequently, one would not expect high area defense scores.

RESULTS

A series of 2 x 2 Analyses of Variance (ANOVAs) were performed to test whether the independent variables significantly affected the dependent variables. Communication training (i.e., the briefing) was a within-subject variable, and always included in these ANOVAs. The other four independent variables were between-subject variables. Each ANOVA included only one between-subject variable. We decided not to include more than one between-subject variable in an ANOVA because of the small and unequal sample sizes in many of the cells. Consequently, we were unable to examine interactions between the between-subject variables. The ANOVAs were performed separately for each type of task because the tasks used different performance variables.

We first present the results of the ANOVAs for the Communication Training and TCO vs. TCA Accuracy independent variables. Then, we present the results of the ANOVAs for the other independent variables. A table summarizing the results that approached or were below the traditional 0.05 alpha level is presented toward the end of the Results section.

We begin each section by restating the hypotheses presented in the INTRODUCTION. Then, we present the ANOVA results for the two communication process measures: (1) a team's total number of communications (communication quantity), and (2) it's sufficiency score (communication quality). Finally, we present the ANOVA results for the task performance measures. The general hypothesis is that the independent variables will affect the communication process variables and, turn, the performance variables.

<u>Communication Training and TCO vs. TCA Accuracy</u>

<u>Predictions</u>. First, we predicted a main effect for communication training for both tasks; that is, that the briefing would improve communication processes and performance in both cases. This briefing main effect was predicted for each 2 x 2 ANOVA; consequently, it will not be mentioned each time below.

Second, we did not predict a main effect for TCO vs. TCA Accuracy for either task. We did, however, predict a briefing x TCO vs. TCA accuracy interaction, but only for the "few track/cue conflict" task. In particular, we predicted that when the TCA was more accurate, the briefing would improve communication processes and performance. However, when the TCO was more accurate, we expected an improvement in the process measure after the briefing, but a decrement in performance because the TCO might defer to the less accurate TCA.

Results for Communication Quantity. There were no significant results for the "few track/cue conflict" task. Contrary to our predictions, the briefing did not significantly increase the total number of communications for this task.

However, as predicted, there was a significant briefing main effect for the "many track/no cue conflict task; [$F(1,24) = 42.34$, $MSe = 20.56$, $p = .0001$]. The briefing increased the number of communications by over 50% (mean before = 231; mean after = 349) for this task. No other effects were significant for this dependent variable.

Results for Communication Quality. Contrary to our predictions, the briefing failed to improve teams' communication quality for either task. Nor was the predicted briefing x accuracy briefing significant.

Results for Task Performance. The briefing main effect approached significance for the "few track/cue conflict" task; [$F(1,26) = 3.17$, $MSe = 0.018$, $p = 0.087$]. However, contrary to our predictions, performance was worse, not better after the briefing (mean before = 77.5%; mean after = 71%).

Examination of Figure 1 shows that when the TCO was more accurate, performance decreased after the briefing for the "few track/cue conflict" task, as predicted. However, contrary to our predictions, performance when the TCA was more accurate also decreased after the briefing.

- - - - - - - - - - - - - - - - - - -

Insert Figure 1 about here

- - - - - - - - - - - - - - - - - - -

Contrary to our predictions, team performance was approximately the same before the briefing, regardless if the TCA or TCO was more accurate. Since the communication quality and quantity measures were about the same for the TCO More Accurate and the TCA More Accurate conditions before the briefing, the result indicate that, contrary to our hypothesis, the communication channels were not closed in the TCA More Accurate condition before the briefing.

There were no significant performance effects for the "many track/no cue conflict" task.

Hours Worked Together as a Team

Predictions. We predicted a main effect for this variable, for both process measures and the performance measure, but only for the "many track/no conflict" task because this is the type of task for which Patriot personnel routinely train. In addition, we predicted a briefing x hours interaction for this task because teams that had worked together less had more to gain from communication training. We did not predict an interaction for the "few track/cue conflict" task because teams seldom perform this task and, therefore, need communication training regardless of the number of hours they had worked together. (Note: As mentioned above, a briefing main effect was predicted for each 2 x 2 ANOVA, for each dependent variable, for each task.)

Results for Communication Quantity. There were no significant effects involving the number of hours that teams had worked together. Although this was consistent with our predictions for the "few track/cue conflict" task, it was inconsistent with our predictions for the "many tracks/no conflict" task.

Results for Communication Quality. As predicted, there were no significant effects involving the number of hours that teams had worked together for the "few track/cue conflict" task. This supports the prediction that the time a team has worked together as a team has no effect on the quality of their communications for this task.

And consistent with our predictions, there was a significant main effect for the number of hours worked together [$F(1,24) = 20.26$, $MSe = 1.41$, $p = 0.0001$] for the "many track/no conflict" task. As predicted, the number of hours teams had worked together increased the quality of the teams' communications for this task. The mean sufficiency score for teams that had worked together less than 200 hours was 0.94; it was 2.50 for teams that had worked together for more than 200 hours. However, the predicted hours x briefing interaction was not significant for this task.

Results for Task Performance. Consistent with our predictions, there were no significant effects involving the number of hours teams had worked together for the "few tracks/cue conflict" task. However, it should be noted that the F-value for the briefing main effect reached the traditional $p = 0.05$ significance level for the 2 (Briefing) x 2 (Hours) ANOVA for this task; $F(1,28) = 4.16$, $MSe = 0.018$, $p = 0.052$. A larger F-value was obtained for this ANOVA than the TCO vs. TCA Accuracy ANOVA because of differences in the way the sum of squares were partitioned. The means for the briefing main effect (77.5% accuracy before and 71% after the briefing) remained unchanged.

Consistent with our predictions, there was a significant hours main effect for "many track/no conflict" task; $F(1,21) = 6.79$, $MSe = 109.23$, $p = 0.017$. Teams that had worked together longer had better performance on this task; a 54.6% vs. 46.1% area defense score. However, contrary to predictions, neither the briefing x hours interaction or the briefing main effect was significant for this task.

Teams That Communicate A lot vs A little

Predictions. We predicted a main effect for communication quantity for both communication process measures, and for team performance, for both tasks. In addition, we predicted a team quantity x briefing interaction. We predicted that the briefing would have its biggest, positive effect for the teams that communicated less.

Results for Communication Quantity. There was a significant main effect for the communication quantity independent variable on the communication quantity dependent variable for the "few track/cue conflict" task; F(1,21) = 23.61, MSe = 8.20, p = 0.0001. Although predicted, this was not particularly interesting because the variable was created by taking a median split on the number of communications.

What was more interesting, was that the predicted team quantity x briefing interaction approached significance; F(1,21) = 3.96, MSe = 11.05, p = 0.06. Low communicating teams increased their communications by 61%, from 6.17 before the briefing to 9.95 after it. There was essentially no change for the high communicating teams, going from 12.23 before to 12.10 after the briefing. This was contrary to our prediction because we predicted that the briefing would have a positive, albeit smaller impact for this group.

There was also a significant main effect for the communication quantity independent variable on the communication quantity dependent variable for the "many track/no cue conflict" task; F(1,24) = 27.33, MSe = 5473.7, p = 0.001. However, the interaction did not approach significance for this task.

Results for Communication Quality. When the sufficiency score was the dependent variable, only the team quantity main effect approached significance for the "few track/cue conflict" task; F(1,24) = 3.68, MSe = 2.09, p = 0.067. The mean sufficiency score for the High Quantity Teams was 1.92; it was 1.15 for the Low Quantity Teams.

The team quantity main effect was significant for the "many track/no conflict" task; F(1,24) = 8.38, MSe = 1.93, p = 0.008.

The mean sufficiency scores for the High Quantity Teams was 2.04; it was 0.92 for the Low Quantity Teams. The predicted briefing main effect and the team quantity x briefing interaction were not significant.

Results for Task Performance. Contrary to predictions, there were no significant effects for either task. Communicating more did not translate into improved performance for either task.

[Note: There were two fewer teams when running the 2 (Quantity) x 2 (Briefing) ANOVA. One consequence of this was that the briefing main effect was no longer significant for the "few track/cue conflict" task; $F(1,24) = 2.53$, $MSe = 0.02$, $p = 0.124$. Of course, the briefing still had a negative effect: mean agreement before = 77.4%, and after = 71.2%.]

Teams That Have High vs. Low Quality Communications

Predictions. Consistent with our predictions for team communication quantity, we predicted a main effect for team communication quality for both communication process measures, and for team performance, for both tasks. In addition, we predicted a team quality x briefing interaction, with the briefing having a large, positive effect for the teams that had the lowest quality communications before the briefing.

Although we predicted a small increase in the number of communications exhibited by High Quality Teams, we actually predicted a small decrease in their quality (i.e., sufficiency) scores. This prediction was simply a function of how the quality scores were calculated. A team received a score of 1 point for each communication category where its total number of communications was more than one standard deviation above the mean for that category. Therefore, if there was a large increase in the total number of communications for the Low Quality Teams, then their quality scores had to increase somewhat and the High Quality Teams' quality scores had to decrease somewhat.

Results for Communication Quantity. The team quality main effect approached significance for the "few track/cue conflict" task; $F(1,21) = 4.06$, $MSe = 19,625$, $p = 0.057$. The High Quality Communicating Teams generated approximately 25% more communications than the Low Quality Teams; the means were 91.5 communications and 72.9 communications, respectively.

More interestingly, the predicted quality x briefing interaction was significant for the "few track/cue conflict" task; $F(1,21) = 6.42$, $MS = 643.6$, $p = 0.019$. As predicted, the Low Quality Teams generated more communications after the briefing. In particular, there was a 53% increase, from a mean of 57.6 communications before to 88.1 communications after the briefing. Contrary to our predictions, there was a 10% decrease in the mean number of communications for High Quality Teams; from 95.7 to 87.3.

In contrast to the findings for the "few track/cue conflict" task, the main effect for Team Quality was not significant for the "many track/no conflict" task. However, the quality x briefing interaction was; $F(1,24) = 6.76$, $Mse = 2063.3$, $p = 0.016$. As predicted, the Low Quality Teams had more communications after the briefing. In particular, there was a 19% increase, from a mean of 207.4 to 247.1. Like the "few track/cue conflict task," the High Quality Teams exhibited a 10% decrease in the mean number of communications: from 284 to 257.

<u>Results for Communication Quality</u>. As expected, there was a significant main effect for team quality on the communication quality dependent variable for the"few track/cue conflict" task; $F(1,24) = 13.21$, $MSe = 1.56$, $p = 0.001$. However this effect is not particularly interesting because this variable was created by taking a median-split on the sufficiency scores.

More interestly, we also obtained the predicted team quality x briefing interaction for the "few track/cue conflict" task; $F(1,24) = 4.83$, $MSe = 1.63$, $p = 0.038$. For the Low Quality Teams, the mean sufficiency scores increased by almost 100%; from 0.67 before the briefing to 1.33 after it. This was consistent with our hypothesis. The briefing led to 33% decrease in the mean sufficiency scores for High Quality Teams; from 2.272 to 1.82.

There was also a team quality main effect for the sufficiency score dependent variable for the "many track/no cue conflict" task; $F(1,24) = 14.88$, $MSe = 1.61$, $p = 0.001$. Again, and more interesting, we obtained a significant team quality x briefing interaction; $F(1,24) = 9.75$, $MSe = 1.19$, $p = 0.005$. For Low Quality Teams, the mean sufficiency scores increased by 350%; from .40 before the briefing to 1.40 after it. This was consistent with our hypothesis. And just as in the "few track/cue conflict" task, the briefing led to a 33% decrease in the mean sufficiency scores for High Quality Teams; from 2.72 to 1.82.

Results for Task Performance. Contrary to our prediction, the team quality main effect was not significant for the performance dependent variable for the "few track/cue conflict" task. However, the predicted team quality x briefing interaction did approach significance; $F(1,24) = 3.97$, $MSe = 0.017$, $p = 0.058$.

For the Low Quality Teams, the level of performance remained unchanged at 75%. Surprisingly, significant increases in the mean number and quality of the Low Quality Teams' communications did not translate into improved performance for the "few track/cue conflict" task. For the High Quality Teams, performance actually decreased by almost 20% after the briefing, from a mean of 81% to a mean of 66%. For this group, the briefing's negative effect on the number and quality of their communications did result in a decrease in performance for this task. [Note: The briefing main effect approached significance for the 2 (Quality) x 2 (Briefing) ANOVA; $F(1,24) = 3.97$, $MSe = 0.017$, $p = 0.058$.]

There were no significant effects on performance for the "many track/no conflict" task. Just as in the case of the "few track/cue conflict task, the briefing's positive effect on the number and quality of the communications for the Low Quality Teams did not translate into a performance increment. In contrast, the briefing's negative effect on the number and quality of the communications for the High Quality Teams did not translate into a performance decrement. Taken in total, these results suggest that a one-to-one communication process to task performance linkage does not exist for these tasks, at least not according to how we were able to operationalize them.

Summary of Results

Table 5 presents a summary of all results reported thusfar that approached or were below the traditional 0.05 alpha level. The key points are as follows:

* Communication training (via the briefing) had a negative effective on performance for the "few track/cue conflict" task. Although it significantly increased the amount of communication for the "many track/no conflict" task, this did not translate into better performance.
* TCO vs. TCA Accuracy had no effect.
* Teams that had worked together longer had higher quality communications and better performance, but only for the "many track/no conflict" task.
* Teams that had more communications also tended to have higher quality communications for both tasks. However, this did not translate into better performance for either task. The briefing increased the number of communications for Low Quantity Teams, but only for the "few track/cue conflict" task.
* Teams that had higher quality communications had more communications, but only for the "few track/cue conflict" tasks. However, for both tasks, the briefing increased the amount and quality of communications for the Low Quality

Teams. In contrast, it decreased the amount and quality of the communications for the High Quality Teams. This resulted in significantly poorer performance for the "few track/cue conflict" task.

Additional Analyses

We did two additional types of analyses in an effort to better understand why the briefing had a negative effect on team performance for the "few track/cue conflict" task. First, we performed a 2 (team quality) x 2 (briefing) x 2 (late order) x 2 (initial information) x 2 (side) ANOVA for each of the two layouts (i.e., scenarios) for the "few track/cue conflict" task. Task performance (i.e., agreement with the Patriot system's recommendation) was the dependent variable. These ANOVAs were performed separately for each layout because Adelman et al. (1995) found significant differences in the order effects depending on the layout. Because the ANOVAs were performed separately for each layout, both briefing and team quality were between-subject factors; late order, initial information, and side were within-subject factors.

For layout 1, the team quality x briefing x late order x initial information x side interaction approached significance; $F(1,22) = 2.99$, $MSe = 0.07$, $p = 0.098$. For layout 2, the team quality x briefing x late order x initial information was significant; $F(1,22) = 5.2$, $MSe = 0.11$, $p = 0.033$. For both layouts, the briefing had a pronounced negative effect on the teams' performance for the two aircraft whose initial information suggested it was hostile (H), and who had the disconfirm/confirm (DC) late order sequence. This was particularly the case for the High Quality Teams. This is shown in Figure 2, which illustrates the significant four-way interaction for layout 2.

- - - - - - - - - - - - - - - - - -

Insert Figure 2 about here

- - - - - - - - - - - - - - - - - -

At the same time, the High Quality Teams achieved a perfect performance after the briefing for the two aircraft whose initial information suggested a friend (F), and whose late order sequence was confirm/disconfirm (CD), for layout 2. Taken together, these results suggest that, after the briefing, the High Communication Quality Teams were engaging these two types of aircraft (i.e., initial information equal to friend and hostile) significantly less, with the result being better performance for the F, CD aircraft and much worse performance for the H, DC aircraft.

This led us to perform a second analysis, which was another 5-way ANOVA for each layout, but this time using whether or not the teams engaged (i.e., shot at) the aircraft as the dependent variable. For layout 1, the teams quality x briefing x late order x side interaction approached significance; $F(1,22) = 2.99$, $MSe = 1.55$, $p = 0.098$. For layout 2, there was a significant team quality x briefing x late order interaction; $F(1,22) = 5.20$, $MSe = 0.11$, $P = 0.033$. This occurred because, for the Low Quality Teams, engagement behavior was lowest only for the DC late order sequence after the briefing (mean after = 22%; mean before =

50%). For the High Quality Teams, engagements were substantially lower for both the CD and DC orders after the briefing (means were 33% and 29% after versus 60% and 50% before, respectively).

DISCUSSION

Communications training, as implemented via the briefing, did not have the predicted, positive effect on the performance of Patriot teams. Although it significantly increased the number of communications, it did not increase the quality of those communications or the teams' overall performance for the high workload/low cognitive stress task (i.e., the "many track/no cue conflict" task). The briefing actually hurt performance for the low workload/cognitively stressing task (i.e. the "few track/cue conflict" task). This negative effect was localized in the more experienced (in terms of hours worked together) and better communicating teams (in terms of communication quality).

Our results are contrary to those presented elsewhere in the literature. Our post-hoc hypothesis for explaining them is that our communication training intervention (i.e., the briefing) was too short. The training was only ten minutes long versus, for example, being days in length as in the case of the behavioral training course reported in Leedom and Simon (1995). We knew that ten minutes was an extremely short time period for a communications intervention before doing the study. However, we thought that it might be enough time to emphasize the importance of communication principles and behaviors, particularly the importance of using an additive processing strategy and, thereby, improve performance. At a minimum, we thought the briefing would have no effect. We never anticipated a negative effect.

The briefing's negative effect was localized to the "few track/cue conflict" task for the more experienced and better quality communicating teams. The results indicate that the briefing disrupted their natural communication processes. For example, High Quality Teams had lower communication quantity and quality scores after the briefing. In addition, they had much worst performance after the briefing for tracks whose initial information was hostile. This was caused by a decrease in the teams' engagement behavior. Simply specifying good communication principles, even when illustrated with specific examples, was not an adequate communication training intervention in general, and certainly not for a task with substantial cue conflict.

We think that a longer training period, and substantial practice to make good communication principles second-nature, would be a better communication training intervention. This communication training must be tailored specifically for the type of task. We predict that the type of behavioral "crew coordination" training illustrated in Leedom and Simon (1995) will be effective for high workload/low cognitive stress tasks, like the "many track/no cue conflict" task, but not necessarily for cognitive stress/low workload tasks, like the "few track/cue conflict" task. This type of training emphasizes critical behaviors for performing a task effectively and efficiently. It does not focus on how to analyze differences in how people are thinking about a problem when there is conflicting information and substantial uncertainty. We think the latter will require a different type of communication training.

Our hypothesis is that it will take analysis of team members' judgment processes, and comparison with the processes deemed appropriate by training personnel, to improve performance in the "few track/cue conflict" task. This is referred to as cognitive feedback in the Social Judgment Theory literature (e.g., see Hammond, Stewart, Brehmer, & Steinmann, 1975). A review by Balzer, Doherty, and O'Connor (1989) indicates that cognitive feedback has successfully reduced cognitive differences, and improved performance, in a variety of domains.

The key idea is that team members need to learn how to communicate how they are reaching their decisions, assess where there are critical differences, and learn how to resolve them. Not only does such training take more time than we had in the current study, it requires a cognitive, not behavioral (i.e., coordination) focus. In contrast we believe a behavioral, not cognitive, focus is required for high workload tasks, like the "many track/no cue conflict" task. This predicted training x task interaction is an issue for future research.

We made a number of hypothesis regarding group composition variables, some of which were confirmed, while others were not. With regard to TCO vs. TCA Accuracy, we predicted that communication training (i.e., the briefing) would enhance performance in the "few track/cue conflict" task when the TCA was more accurate, but hurt performance when the TCO was more accurate. These predictions rested on two assumptions. The first assumption was that, on the average, communication channels would not be fully open before the briefing; consequently the TCA's more accurate knowledge would not be able to impact the decision as much as it should. The second assumption was that, after the briefing, the TCO would listen more to the less accurate TCA and, as a result, that would be a decrease in performance.

There was some support for the second assumption; performance was worse after the briefing when the TCO was more accurate. However, there was no support for the former assumption. When the TCA was more accurate, performance was the same before and after the briefing. In fact, before the briefing, performance was the same regardless of whether the TCO or TCA was more accurate on the pencil-and-paper pretest. Although the two conditions were not as different as we would have liked, they were still different. Therefore, the results suggest that communication channels were not closed more in one condition than another before the briefing. We were not able to discover this before the study because, as stated earlier, we were limited to one participation session with each Patriot TCO and TCA, and all aspects of the research effort with them had to be completed at that time. It should be noted that, as predicted, TCO vs. TCA Accuracy on the "few tracks/cue conflict" task had no effect on the "many track/no cue conflict" task.

The results supported two hypotheses with respect to the

"Number of Hours Worked Together" group composition variable. First, as predicted, teams that had worked together longer performed better, but only on their primary task (i.e., the "many track/no cue conflict" task). Better performance was accompanied by better communication quality, not a greater number of communications.

Second, as predicted, the number of hours a team had worked together had no effect on communications quality or performance for the "few track/cue conflict" task. Patriot teams seldom train on this kind of task because it seldom occurs. However, anecdotal stories of incidents during Desert Shield, and the U.S.S. Vincennes tragedy, indicates that this type of air defense task can occur. Patriot operators need to be trained on it because the results suggest that mental models and communication skills will not transfer from one kind of task (i.e., the "many track/no cue conflict") to another (i.e., the "few track/cue conflict" task).

Contrary to our hypothesis, better communicating teams did not perform better on either task. It did not matter whether one defined "good communication" in terms of communication quantity or quality. Consistent with our prediction, the briefing did improve the communication processes for poor communicating teams. Teams which had a low number of communications before the briefing increased their communications 100% for the "few track/cue conflict" task and 350% for the "many track/no cue conflict" task after the briefing. However, these process improvements did not lead to performance improvements for either task. And, contrary to our predictions, the briefing reduced the quality scores for High Quality Teams by 33% for both tasks. This process decrement resulted in a 20% performance decrement for the "few track/cue conflict" task.

We have assumed that, for both tasks, performance would be mediated by communication process measures. There is substantial support for this position in the literature (e.g., Adelman et al., 1986; Gualtieri, 1994; Hackman & Morris, 1975). However, Duffy (1993, p. 258) notes that the evidence for communication effectiveness is "piecemeal and conflicting" for computer-supported decision making. This is what we found in the current study with Patriot teams using the Patriot simulator.

Of the two communication process measures we used, the quality measure had a greater impact on performance. In fact, there was no case where an increase in the number of communications (i.e., quantity) resulted in an increase in performance. For example, the briefing increased the number of communications by the Low Quantity and Low Quality Teams for both tasks without increasing performance.

On the other hand, there were some relationships between communication quality and team performance. For example, teams that had been together longer had significantly higher quality (not quantity) scores for the "many track/no conflict" tasks, and significantly higher performance scores for that task. And the briefing's negative effect on the quality scores of the High Quality Teams resulted, in turn, in lower performance scores for the "few track/cue conflict" task.

However, there was not a one-to-one relationship between communication quality and task performance. For example, the Higher Quality Teams did not perform better on either task. The large quality score improvement for the Low Quality Teams for both tasks did not translate into performance improvements on either task. Nevertheless, these findings are important because many of the papers cited above list both quantity and quality improvements as the value of communication training. The research reported herein suggests that, at least according to how we operationalized them for Patriot air defense tasks, communication quality, not quantity, is the more important determinant of team performance.

REFERENCES

Adelman, L., Zirk, D.A., Lehner, P.E., Moffett, R.J., & Hall, R. (1986). Distributed tactical decisionmaking: Conceptual framework and empirical results. IEEE Transactions on Systems, Man, and Cybernetics, SMC-16, 794-805.

Adelman, L., and Bresnick, T.A. (1992). Examining the effect of information sequence on Patriot air defense officers' judgements. Organizational Behavior and Human Decision Processes, 53, 204-228.

Adelman, L., Bresnick, T., Black, P., Freeman, F.F., & Sak, S.G. (in press). Research with Patriot air defense officers: Examinining information order effects. Human Factors.

Adelman, L., Bresnick, T., Christian, M., Gualtieri, J., & Minionis, D. (1995). Demonstrating the effect of context on order effects. Fairfax, VA: George Mason University.

Adelman, L., Tolcott, M. A., & Bresnick, T.A. (1993). Examining the effect of information order on expert judgement. Organizational Behavior and Human Decision Processes, 56, 348-369.

Bales, R.F. (1950). Interaction process analysis: A method for studying small groups. Cambridge, MA: Addison-Wesley.

Balzer, W.K., Doherty, M.E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. Psychological Bulletin, 106, 410-433.

Cannon-Bowers, J.A., Salas, E., & Converse, S. (1993). Shared decision models in expert team decision making. In N.J. Castellan Jr. (ed.), Individual and group decision making: Current issues. Hillsdale, NJ: Lawrence Erlbaum.

Duffy, L. (1993). Team decision making and technology. In N.J. Castellan Jr. (ed.), Individual and group decision making: Current issues. Hillsdale, NJ: Lawrence Erlbaum.

Gualtieri, J. (1994). Group decision making for a complex dynamic task: An examination of the effects of individual and group characteristics on decision processes and performance. (Unpublished doctoral dissertation.) Fairfax, VA: George Mason University.

Hackman, J.R., & Morris, C.G. (1975). Group tasks, group interaction process, group performance effectiveness: A review and proposed integration. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (vol. 8). New York: Academic Press.

Hammond, K.R., Stewart, T.R., Brehmer, B., Steinmann, D.O. (1975). Social judgment theory. In M.F. Kaplan & S. Schwartz (Eds.), Human judgment and decision processes. New York: Academic Press.

Leedom, D., & Simon, R. (1995). Improving team coordination: A case for behavior-based training. Military Psychology, 7, 109-122.

McIntyre, R.H., Morgan, B.B. Jr., Salas, E., & Glickman, A.S. (1989). Teamwork from team training: New evidence for the development of teamwork skills during operational training. Orlando, FL: Naval Training Systems Command (Code 712).

Nieva, V.F., Fleishman, E.A., & Rieck, A.M. (1978). Team dimensions: Their identity, their measurement, and their relationships. Washington, D.C.:ARRO.

Orasanu, J., & Salas, E. (1993). Team decision making in complex environments. In G.A. Klein, J. Orasanu, R. Calderwood, & C.E. Zsambok (eds.), Decision making in action: Models and methods. Norwood, NJ: Ablex.

FOOTNOTES

[1] The research described herein was supported by contracts from the Research & Advanced Concepts Office of the U.S. Army Research Institute (ARI), Contract No. MDA903-92-K-0134, and the AASERT Program sponored by the U.S. Army Research Office (ARO), DAAH04-93-G-0286 (32239-RT-AAS), to George Mason University (GMU). We would like to acknowledge the effort of many people who contributed to this endeavor, and identify where they were located at the time of the study. In particular, we would like to thank Dr. Michael Drillings at ARI Headquarters; Dr. John Lockhart at the ARI Field Unit at Ft. Bliss, TX; Mr. Andy Washko, Mr. Jerry Crisp, and all personnel involved in scenario development and simulator operations, respectively, at Ft. Bliss; and Major Mark Bahr and all participating TCOs and TCAs of the 11th Air Defense Brigade at Ft. Bliss. We also want to thank Ms. Mary Anne Flood, Ms. Karen Johnson, and Ms. Laura Bowling, at GMU, for their assistance. The views, opinions, and findings herein are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other official documentation.

[2] Reprint requests should be addressed to Leonard Adelman, Dept. of Operations Research and Engineering, School of Information Technology and Engineering, George Mason University, Fairfax, VA 22030-4444.

Table 1

Statistical Characteristics of the "TCO vs. TCA Accuracy" Independent Variable

	TCO More Accurate (n=15)				TCA More Accurate (n=13)			
	Mean	Min	Max	SD	Mean	Min	Max	SD
TCO	22.267	19	32	3.218	17.615	14	20	1.609
TCA	17.800	15	19	1.207	20.462	15	23	2.332
DIFF	4.4670	1	14	3.399	2.846	1	7	1.951

Table 2

Statistical Characteristics of the "Numbers of Hours Worked Together" Independent Variable

	N	Mean	Min	Max	SD
<200 Hours	17	10.29	0	100	24.70
>200 Hours	11	859.09	280	1500	438.17

Table 3

Statistical Characteristics of the "Communication Quantity" Independent Variable

	"Few Track/Cue Conflict"					"Many Track/No Cue Conflict"				
	N	Mean	Min	Max	SD	N	Mean	Min	Max	SD
Low	13	45.7	2	71	19.3	13	181.2	114	234	37.3
High	13	95.2	74	147	23.6	13	298.4	239	432	58.3

Table 4

Statistical Characteristics of "Communication Quality" Independent Variable

	"Few Track/Cue Conflict"					"Many Track/No Cue Conflict"				
	N	Mean	Min	Max	SD	N	Mean	Min	Max	SD
Low	15	0.67	0	1	0.49	15	0.40	0	1	0.51
High	11	2.91	2	5	0.95	11	2.73	2	6	1.27

Table 5

Summary of Significant Results

Independent Variables	"Few Track/Conflict" Comm. Quan.	"Few Track/Conflict" Comm. Qual.	"Few Track/Conflict" Task Perf.	"Many Track/No Conflict" Comm. Quan.	"Many Track/No Conflict" Comm. Qual.	"Many Track/No Conflict" Task Perf.
Comm. Briefing			-	+		
TCO vs. TCA Acc.						
Hours Together					+	+
Comm. Quantity	+ L+B=+ H+B=NE	+		+	+	
Comm. Quality	+ L+B=+ H+B=-	+ L+B=+ H+B=-	L+B=NE H+B=-	L+B=+ H+B=-	+ L+B=+ H+B=-	

Notation:

+	indep. var. had a positive main effect on dep. var.;
-	indep. var. had a negative main effect on dep. var.;
L+B=+	means low level and briefing had a positive effect;
L+B=NE	means low level and briefing had no effect;
H+B=NE	means high level and briefing had no effect; and
H+B=-	means high level and briefing had a negative effect.

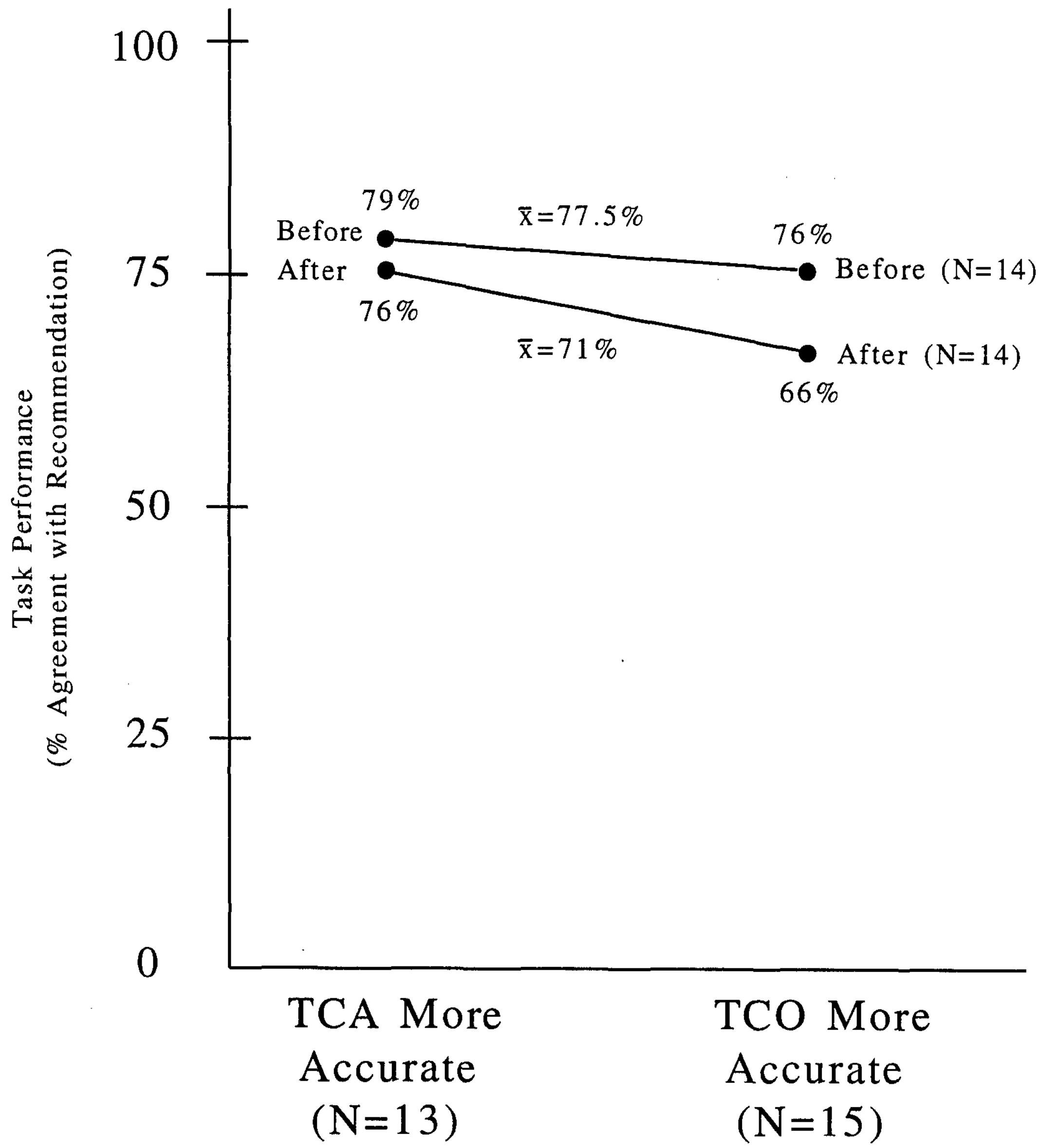

Figure 1. Results of 2 (Briefing) x 2 (TCO vs. TCA Accuracy) ANOVA for "few track/cue conflict" task.

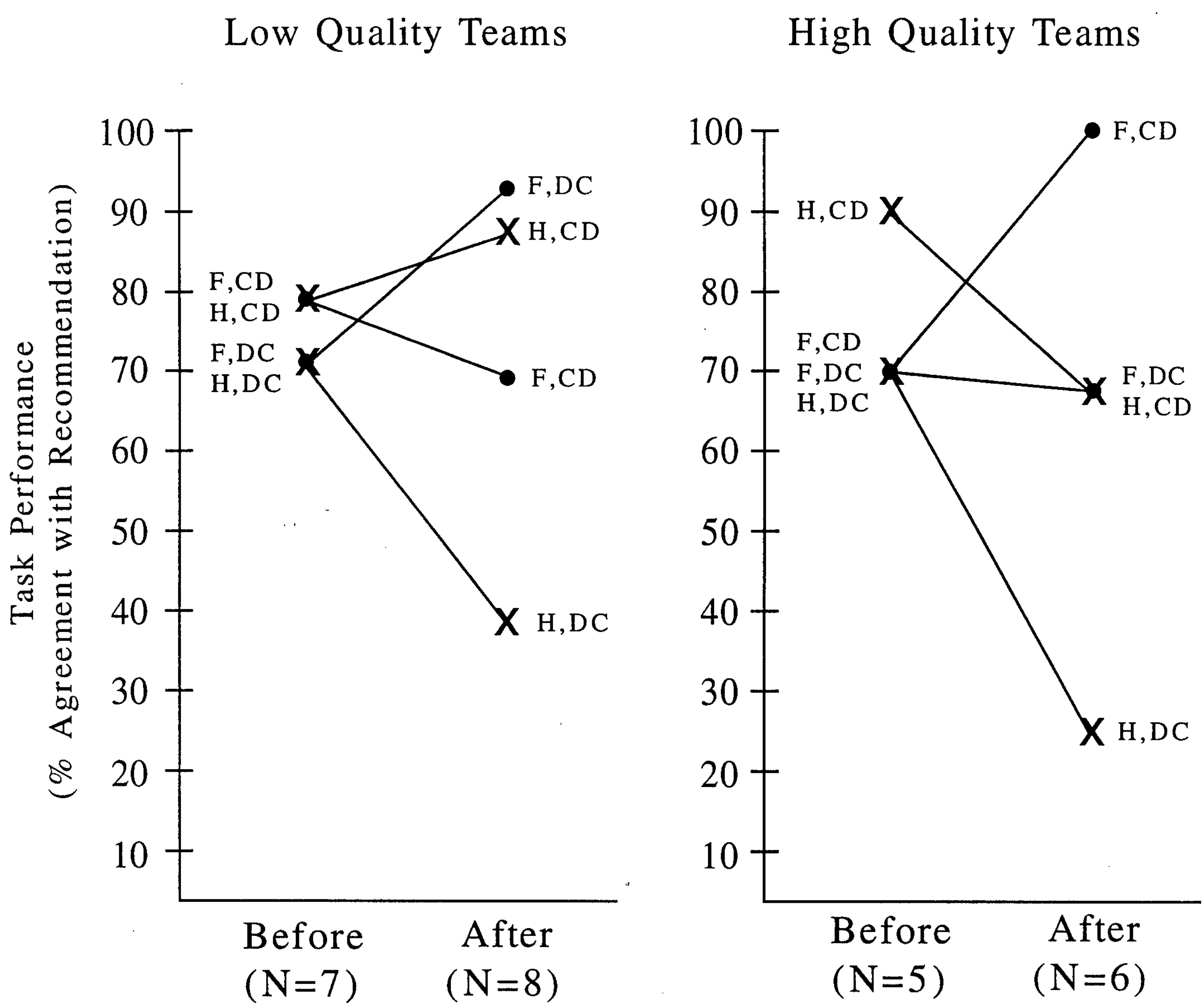

Figure 2. The team quality x briefing x late order x initial information interaction for layout 2.