Technical Research Note 219

AD

MAINTAINING TARGET DETECTION PROFICIENCY THROUGH TEAM CONSENSUS FEEDBACK

John T. Cockrell
System Development Corporation

SUPPORT SYSTEMS RESEARCH DIVISION

U. S. Army
Behavioral Science Research Laboratory

December 1969

Technical Research Note 219

AD

MAINTAINING TARGET DETECTION PROFICIENCY THROUGH TEAM CONSENSUS FEEDBACK

John T. Cockrell
System Development Corporation

Robert Sadacca, Program Director

SUPPORT SYSTEMS RESEARCH DIVISION
Joseph Zeidner, Chief

U. S. ARMY BEHAVIORAL SCIENCE RESEARCH LABORATORY

Office, Chief of Research and Development
Department of the Army

Room 239, The Commonwealth Building
1320 Wilson Boulevard, Arlington, Virginia 22209

December 1969

Army Project Number
2Q662704A721

Contract Number DA-HC 19-67-C-0040
Component Integration b-21

BESRL Technical Research Reports and Technical Research Notes are intended for sponsors of R&D tasks and other research and military agencies. Any findings ready for implementation at the time of publication are presented in the latter part of the Brief. Upon completion of a major phase of the task, formal recommendations for official action normally are conveyed to appropriate military agencies by briefing or Disposition Form.

FOREWORD

The SURVEILLANCE SYSTEMS research program of the U. S. Army Behavioral Science Research Laboratory has as its objective the production of scientific data bearing on the extraction of information from surveillance displays, and the efficient storage, retrieval, and transmission of this information within an advanced computerized image interpretation facility. Research results are used in future systems design and in the development of enhanced techniques for all phases of the interpretation process. Research is conducted under Army RDT&E Project No. 2Q662704A721, "Surveillance Systems," FY 1970 Work Program.

BESRL research in this area is conducted as an in-house research effort augmented by research contracts with organizations selected as having unique capabilities and facilities for research in intelligence systems. The present study was conducted jointly by personnel of the System Development Corporation and of the Behavioral Science Research Laboratory, under program direction of Robert Sadacca.

The IMAGE SYSTEMS Work Unit is one of four current research work units which focus on operationally meaningful segments of the Army's surveillance systems. Among the specific objectives of the work unit is the development of procedures to maintain and improve the proficiency of interpreters within an image interpretation facility. An exploratory study in this area was reported in BESRL Technical Research Note 195, "Maintaining image interpreter proficiency through team consensus feedback." The present publication reports on further study of team consensus feedback as a means of improving performance of individual interpreters, with emphasis on target detection skill.

J. E. UHLANER, Director
U. S. Army Behavioral Science
Research Laboratory

MAINTAINING TARGET DETECTION PROFICIENCY THROUGH TEAM CONSENSUS FEEDBACK

BRIEF

Requirement:

To continue the investigation of the effectiveness of team consensus feedback proficiency maintenance methods for maintaining and improving the proficiency of image interpreters--specifically, to determine if the target detection skill of individual interpreters can be improved by feedback which team members generate for themselves as they compare and discuss their work.

Procedure:

This experiment differed from a previous experiment in the series in that target detection only was required, rather than detection plus identification. Treatment was a three-day practice session. A pre-treatment and a post-treatment test were administered to each interpreter to assess detection proficiency. The interpreters assigned to feedback conditions practiced in teams; groups were arranged in a factorial design which allowed comparison of three-man teams versus two-man teams; discussion versus no discussion; heterogeneous teams in terms of initial proficiency versus homogeneous teams; and comparisons between interpreters of high, medium, and low initial detection proficiency. The no-feedback interpreters, who practiced alone, did not discuss or compare their work with anyone. None of the interpreters received ground truth feedback at any time.

Findings:

Interpreters working in teams with consensus feedback showed greater improvement than interpreters working alone in reducing inventive errors, but there was no difference in errors of omission. These results are in agreement with previous experimentation.

Interpreters working in heterogeneous teams made significantly greater improvement on all measures than interpreters in homogeneous teams. There was no difference between discussion versus no discussion and three-man teams versus two-man teams.

Interpreters initially low in proficiency made greater improvement in reducing inventive errors than did medium or high interpreters. Interpreters of medium initial skill improved more than high interpreters. Proficiency groups did not differ in number of omissions or total errors.

Utilization of Findings:

As a method of maintaining the proficiency of interpreters in an image interpretation facility, team consensus feedback can yield improvement in individual performance, particularly in target identification and reduction of inventive errors. The technique is especially useful where ground truth is not available. Operational imagery usually available within an operational image interpretation facility can be used in such practice.

MAINTAINING TARGET DETECTION PROFICIENCY THROUGH TEAM CONSENSUS FEEDBACK

CONTENTS

TABLES

FIGURES

MAINTAINING TARGET DETECTION PROFICIENCY THROUGH TEAM CONSENSUS FEEDBACK

CONTEXT OF THE STUDY

In a previous study,[1] the team consensus feedback method was developed and tested as a possible aid to proficiency maintenance for image interpreters. The method uses team operations as a means of improving the skills of individual interpreters. The essential difference between the method and more usual instructional methods is that the team members receive no knowledge as to the accuracy or completeness of their own interpretations except through comparison and discussion with their teammates.

The method was based on prior studies[2,3,4] which demonstrated that image interpreters working in teams can produce more complete and accurate intelligence information from aerial reconnaissance imagery than interpreters working alone. The consensual judgment of team members is especially effective in reducing the number of identification errors made by single interpreters. Since teams produce better reports than individuals, interpreters working in teams can receive more accurate knowledge of results than interpreters working alone. Image interpreters working alone on a mission are often unaware when they are doing a poor job of detecting and identifying targets. Seldom do they receive any feedback, and if they do, it is generally too late to be effective. In teams, however, interpreters can take stock of themselves whenever their teammates find targets and make interpretations at variance with their own. In conflict situations regarding targets and identifications, it has been found that teammates who discuss their conflicts frequently arrive at correct identifications.

In the first study testing the team consensus feedback method, results indicated that interpreters practicing in teams make greater performance gains than interpreters practicing alone. Although the evidence is not complete that the performance gains are due to the better feedback

---

[1] Cockrell, J. T. Maintaining image interpreter proficiency through team consensus feedback. BESRL Technical Research Note 1[illegible]5. April 1[illegible]8.

[2] Doten, G. W., J. T. Cockrell, and R. Sadacca. The use of teams in image interpretation: Information exchange, confidence, and resolving disagreements. BESRL Technical Research Report 11[illegible]1. October 1[illegible]6.

[3] Bolin, S. F., R. Sadacca, and H. Martinek. Team procedures in image interpretation. BESRL Technical Research Note 1[illegible]4. December 1[illegible]5.

[4] Sadacca, R., H. Martinek, and A. I. Schwartz. Image interpretation task--status report. BESRL Technical Research Report 112[illegible]. June 1[illegible]2.

which team members receive, the hypothesis is reasonable. Supporting the hypothesis is the result that the least amount of performance gain occurred under the work procedure which involved the greatest delay between initial interpretation and team discussion. This result is in keeping with general psychological evidence with regard to delay of feedback or reinforcement.

Other results of the first study indicated that there was much improvement in terms of errors of identification, some improvement in errors of invention (calling a non-target a target), but no improvement in errors of omission. Analysis of the procedures used in the experiment revealed that most of the practice was concentrated on errors of identification and errors of invention with very little practice on errors of omission. Accordingly, it was felt that a better assessment of the effect of team consensus feedback on errors of omission could be obtained through employing a procedure which concentrated on omissions and which greatly increased the number of detection practice units (frames) presented per unit of time.

OBJECTIVES

Field interpretation units typically have a relatively large number of inexperienced personnel and a relatively small number of experienced personnel. Some type of proficiency maintenance practice is necessary for these interpreters, especially for those who are recent graduates of interpretation schools or transferees from other kinds of work. The team consensus feedback method, if proved feasible, would offer a relatively simple and inexpensive method of providing this practice. The advantages of the method are that no elaborate and expensive materials need be acquired, and practice sessions can be initiated during any slack period by simply using rolls of off-the-shelf imagery.

A series of experiments is being conducted in an effort to develop team consensus feedback procedures which will lead to performance gains by individual interpreters. The first experiment was designed to obtain a general assessment of the usefulness of the consensus feedback process. The second experiment, described here, was designed to take a much closer look at the detection process to see if errors of omission could be reduced by consensus feedback practice. The primary objective of the present experiment was to concentrate practice on detection skill rather than on requiring the interpreters to identify any targets they detected.

Although the theoretical basis for consensus feedback is the effect of the improved feedback which teamwork provides, a number of other factors in the team setting also may influence individual performance. Team procedures and composition, for example, may play an important role not only in influencing the accuracy and completeness of the team report, but also in determining whether the feedback is accepted by the individual team members and how much they are motivated to improve their performance. Team discussion may also be an important factor in passing skills and concepts from high to low proficiency interpreters.

In addition to determining whether detection skills can be improved through team consensus, the present experiment investigated the impact on individual interpreter performance of 1) size of team, 2) discussion vs no discussion, 3) initial proficiency level of team members, and 4) homogeneous vs heterogeneous team composition with respect to initial proficiency level.

METHOD

Subjects

Sixty enlisted men who had just completed the image interpretation course at the U. S. Army Intelligence School comprised the experimental sample. These relatively inexperienced interpreters were judged to have proficiency levels consonant with the proficiency levels of interpreters who might benefit from participating in consensual feedback training programs in the field. All had met the school's entrance requirement of a score of 100 or above on the General Technical Aptitude Area (composite of the Verbal and Arithmetic Reasoning tests).

Imagery

One hundred stereo pairs of photographs with 40 to 60% stereo overlap were selected from rolls of aerial photography taken of military equipment being deployed in Army maneuvers. Each of the stereo pairs contained from 2 to 13 targets with scales ranging from 1:2000 to 1:5000. The stereo pairs were mounted on positive transparency roll film using 9" x 9" format. Six stereo pairs were used for orientation purposes, 12 pairs were used in the pre-training detection test, a maximum of 65 pairs were used in the practice phase, and 17 pairs were used in the post-training test.

Independent Variables

The variable of chief concern was feedback from team consensus versus individual practice with no feedback. Within the feedback method the following variables were introduced:

Team feedback procedure--discussion vs no discussion

Team size--3-man vs 2-man teams

Team composition--homogeneous with respect to initial proficiency level (high, medium, low) vs heterogeneous

Team Feedback Procedure. When the team discussion procedure was used, each man on the team began with the same stereo pair of aerial images. After each man had finished his initial interpretation, he recorded the position of all of his targets on a vellum overlay answer

sheet (shown in Figure 1). The answer sheets were transparent and could be placed on a light table. Positioning marks were provided so that the exact location of each target could be recorded with lead pencil. In addition to location, the interpreters also numbered each target and placed a confidence estimate (described below) beside each target on both the answer sheet in normal pencil and on the imagery in grease pencil. Each team member then passed his answer sheet to the team captain (captaincy was rotated from frame to frame), and the captain added any targets which had not already been marked on his answer sheet and imagery. These new targets were given a special designation to indicate their origin. The team members then gathered around the captain's light table and discussed each target in turn. After discussion, each team member called out his final confidence estimate to the team captain who recorded each man's estimate in a designated column. The final estimate did not necessarily have any relationship to the initial estimate, and the men were encouraged to consider the contents of the discussion before deciding on their final confidence estimate.

In the consensus feedback procedure without team discussion, team members were allowed to see and react to each other's answer sheets, but did not discuss the targets or talk to each other at any time. For the initial interpretation, each man on the team had a copy of the same stereo pair of aerial images. Each man worked by himself during initial interpretation, which was accomplished in the same way as in the discussion procedure. After all men on the team had finished the initial interpretation, each man passed his answer sheet to one of his teammates to be checked. Each checker could thus compare the answer sheet he received with the grease marks he had on his own imagery. Any targets which were on his imagery and not on his teammate's answer sheet were added to the answer sheet with a special designation. Next, the checker looked at all the targets on the answer sheet and placed a second confidence estimate beside the first for each target. The checker was instructed to consider his partner's confidence estimate, his own original estimate, and the appearance of the target in arriving at his revised confidence estimate. Checkers were told that they were not bound by their original estimates but could change their minds. After all checking was finished (for three-man teams, the answer sheets were rotated again for a second check), the answer sheets were passed back to the first interpreter. Each man thus received back his own answer sheet, which now contained all responses made by the team members. Each target on the answer sheet also had accumulated as many as three confidence statements. Each man now weighed all the evidence for each target and put down his final revised confidence estimate. Two types of feedback were considered to be present in this procedure. First, the checkers were receiving feedback by comparing the targets indicated on the answer sheets of their teammates with those they recorded on their own imagery. Second, the original interpreters were receiving feedback when their answer sheets were returned with the accumulated confidence estimates.

NAME ____________ MAN NUMBER __________ STEREO PAIR NUMBER ________ SCALE ________ START TIME ________ STOP TIME __________

1	2	3	4	5	6	7	8
Annot Number	Right Side	Initial Conf.	1st TM Conf	2nd TM Conf	Revised Conf.	Final Conf.	Team Score

Figure 1. Reproduction of data sheet used by subjects to indicate position of targets detected
(The data sheet has been reduced in size from 10" by 14" CLEARPRINT.)

Team Composition. Initial proficiency level as measured by a pre-treatment detection test individually administered to each interpreter served as a basis for categorizing the men as high, medium, or low in initial interpretation skill.

No-feedback Procedure. The interpreters examined the same imagery as under the feedback conditions, except that each interpreter worked by himself and did not discuss or compare target responses with any other interpreter. They received no feedback of any kind.

Experimental Design

The 60 interpreters participating in the experiment were assigned to feedback and no-feedback procedures and different feedback conditions as shown in Table 1. From each level of initial proficiency, men were drawn randomly for assignment to the feedback and no-feedback procedures, to two- and three-man teams, and to homogeneous and heterogeneous teams (Table 2). Twelve subjects, divided equally among the three proficiency levels, served in the no-feedback group.

Table 1

NUMBER OF SUBJECTS ASSIGNED TO EXPERIMENTAL PROCEDURES

Proficiency Level	Feedback Consensus Conditions								Total	No Feedback Condition
	3-Man Teams				2-Man Teams					
	Discussion		No Discussion		Discussion		No Discussion			
	Hom	Het	Hom	Het	Hom	Het	Hom	Het		
High	2	2	2	2	2	2	2	2	16	4
Medium	2	2	2	2	2	2	2	2	16	4
Low	2	2	2	2	2	2	2	2	16	4
Total	6	6	6	6	6	6	6	6	48	12

Table 2

MEN OF DIFFERENT PROFICIENCY LEVELS ASSIGNED TO TEAMS BY SIZE AND COMPOSITION OF TEAM

	3-Man Teams	2-Man Teams
Homogeneous Teams	1. High, High, Medium 2. Medium, Low, Low	1. High, High 2. Medium, Medium 3. Low, Low
Heterogeneous Teams	1. High, Medium, Low 2. High, Medium, Low	1. High, Low 2. High, Medium 3. Medium, Low

Conduct of the Experiment

The experiment was conducted over a five-day period. The first half of the first day was spent in explaining the purpose of the experiment, giving general instructions, and practicing response procedures with three large-scale stereo pairs containing easily detectable targets. After each stereo pair was finished, the response sheets and annotations of each interpreter were checked on an individual basis, and further explanation of the instructions was given where needed. During this period and subsequently throughout the experiment, each interpreter had available a set of photographic keys which contained photographs, scale drawings, and measurements for each target on the target list. The photographic keys also contained vertical photographs of each target in stereo at a scale within the range of those used in the experiment. During the instruction period, no feedback of any kind was given the interpreters.

After the initial instructional period, the interpreters were given an orientation test consisting of three stereo pairs. The imagery was similar to that used in the remainder of the experiment. The interpreters were required to accomplish the detection task by locating the targets on the imagery, circling the target with grease pencil, numbering the targets, and placing a confidence estimate beside each target. The list of required targets is shown in Table 3. As each interpreter finished a stereo pair, he was required to record his finish time and to sit quietly at his light table until all interpreters had finished.

Table 3

TARGET LIST

T	TRACKED VEHICLES	
	TT	Tanks
	TS	SP (Guns, Howitzers, Mortars, Antiaircraft)
	TA	APC's
	TB	Armored Bridge Launchers
	TR	Recovery Vehicles
	TP	Prime Mover/Tractor
A	ARTILLERY	
	AT	Towed Howitzers
	AM	Mortar
	AA	Antiaircraft
	AK	Antitank
M	MISSILES	
	MS	Surface-to-Surface Missile
	ML	Missile Launcher/Transporter
	MT	Missile Transporter
	MA	SAM
	MM	SAM Launcher/Transporter
W	WHEELED AND CONSTRUCTION VEHICLES	
	WL	Light Cargo Trucks 1/4-Ton, 3/4-Ton, Ambulance
	WH	Heavy Cargo Trucks, 2 1/2-Ton, 5-Ton, 10-Ton
	WK	Tank Trucks (Water, Fuel)
	WW	Wrecker Trucks
	WT	Truck Tractor (List Separate from Trailer)
	WV	Van Trucks (Generator, Shop, Communication, Radar)
	WD	Dump Truck
	WC	Construction Vehicles (Bulldozers, Cranes, Shovels, Scoops, etc.)
L	TRAILERS (ANNOTATE SEPARATELY FROM TRUCKS EVEN IF ATTACHED)	
	LL	Light Cargo, 1/4-Ton, 3/4-Ton
	LH	Heavy Cargo 1 1/2-Ton
	LS	Small Special Purpose (Ammo, Generator, Water, Fuel)
	LR	Large Special Purpose (Lc Boy, Tank Transporter, Van, Tanker)
	LE	House Trailers (Military)
C	CANVAS SHELTER	
	CS	Small Personnel Tents (Pup, Wall)
	CM	Medium Special Purpose Tents (CP, Hex, Kitchen)
	CL	Large Tents (GP, Maintenance, Hospital)
	CC	Miscellaneous (Latrine, Canvas Shelter, Canvas Water Tank, Canvas Covered Supplies, Canvas Covered Garbage Pits, Flys)

A pre-training test to determine initial proficiency was administered at the beginning of the second half of the first day. The procedure was identical to that of the orientation test with the exception that a ten-minute maximum time period was imposed for each stereo pair. Scoring was accomplished immediately so that individuals could be assigned to proficiency groups on the second day.

Team interpretation was started on the second day. During this phase, no time limit was imposed, the teams proceeding at their own pace. The subjects were in the laboratory for eight hours each day minus two 10-minute breaks and one 20-minute break each morning and afternoon and a one-hour lunch break. This phase lasted three days. The control group adhered to the same schedule but interpreted the imagery on an individual basis.

The post-training test administration was conducted during the fifth day of the experiment and consumed most of the day. The procedure for this test was identical to that of the pre-training test. During all the individual testing, the interpreters sat at their own light tables. No discussion was permitted and no feedback was given the interpreters.

Confidence estimates were required for each detection. Confidence estimates could range from 0 to 100% and were intended to reflect how confident the interpreters were that a target being recorded was in fact a target on the list. The interpreters were informed that the confidence level would affect their individual scores according to the following formula:

A real target assigned a confidence of 50% or more would count as 1 correct response.

A real target assigned a confidence of 40% or less would count as 1/2 correct response.

A non-target assigned a confidence of 50% or more would count as 1 incorrect response.

A non-target assigned a confidence of 40% or less would <u>not</u> count as an incorrect response.

The interpreters were also informed that scoring of team answer sheets would be on the same basis, with the exception that an average confidence estimate would be used.

A major reason for using this scoring method was to encourage a high rate of response, since interpreters could record doubtful targets without fear of penalty. Also, by assigning a low confidence estimate to a detection, an interpreter could indicate disagreement with his teammate(s).

Dependent Variables

Three dependent variables were based upon detection errors:

Omission Error Score. Number of military targets actually present in the imagery which are not recorded by the subject. One-half an error was counted for any actual target for which the confidence estimate was 49% or less.

Inventive Error Score. Number of targets recorded by the subject which he was specifically instructed to omit--imaginary targets, non-military targets, and military targets not on the target list. Inventive errors were not scored for targets for which the confidence estimate was 49% or less.

Total Error Score. The sum of omission score plus inventive error score. Error scores were computed separately for the pre- and post-training tests. Difference scores obtained by subtracting the error score made on the pre-training test from the error score made on the post-training test were used in the analysis.

RESULTS

A considerable number of errors were made by the subjects, more errors being made in the longer post-training test. Table 4 shows the mean total error scores made on the pre-training test and Table 5 shows the mean total error difference scores. Analysis of variance results for all variables appears as Table 6.

Since neither procedure (discussion-no discussion) nor team size (3-man vs 2-man teams) gave significantly different results, subjects were recombined into team composition (homegeneous and heterogeneous) and initial proficiency groups in order to test the main variable of the experiment, namely, team consensus feedback vs individual practice with no feedback. The means for this analysis are shown in Table 7 and the analysis of variance for groups with unequal numbers is shown in Table 8. The feedback vs no feedback method variable was significant at the .01 level. A comparison of team composition and no-feedback interpreters by means of t-tests showed that the interpreters from the heterogeneous teams differed significantly from the no-feedback team interpreters. No difference was found between the homogeneous teams and no-feedback interpreters.

Table 4

MEAN TOTAL ERROR SCORES ON INITIAL (PRE-TRAINING) PERFORMANCE TEST

Proficiency Level	Team Consensus Feedback								All Feedback Subjects	No Feedback
	3-Man Teams				2-Man Teams					
	Discussion		No Discussion		Discussion		No Discussion			
	Hom	Het	Hom	Het	Hom	Het	Hom	Het		
High	80.0	72.0	90.0	85.5	82.5	77.5	76.5	89.0	81.6	77.3
Medium	102.0	101.0	100.5	101.5	88.5	93.0	100.0	104.5	98.9	98.3
Low	105.5	124.5	130.0	119.0	114.5	135.5	110.5	128.0	122.3	124.5
All Groups	96.5	99.2	103.5	102.0	95.2	102.0	95.7	107.2	100.9	100.0

Table 5

MEAN DIFFERENCES IN TOTAL ERROR SCORE BETWEEN INITIAL AND FINAL PERFORMANCE TESTS

Proficiency Level	Team Consensus Feedback								All Feedback Subjects	No Feedback
	3-Man Teams				2-Man Teams					
	Discussion		No Discussion		Discussion		No Discussion			
	Hom	Het	Hom	Het	Hom	Het	Hom	Het		
High	42.0	31.0	33.5	52.0	77.0	44.0	47.0	43.0	47.1	81.8
Medium	57.0	14.5	72.0	26.0	60.5	34.5	78.0	36.5	47.4	45.8
Low	60.5	44.5	17.5	13.5	34.5	26.5	41.5	39.5	35.4	52.3
All Groups	55.3	30.0	41.0	30.5	58.7	35.0	55.5	39.7	43.3	59.9

Table 6

ANALYSIS OF VARIANCE FOR TOTAL ERROR DIFFERENCE SCORE FOR TEAM CONSENSUS FEEDBACK SUBJECTS

Source of Variation	Sum of Squares	df	Mean Square	F-Ratio
Proficiency Level (P)	1497	2	748	1.88
Team Composition (C)	4313	1	4313	10.83*
Feedback Method (M)	123	1	123	.28
Team Size (S)	744	1	744	1.87
PC	2410	2	1205	3.03
PM	1438	2	719	1.81
CM	402	1	402	1.01
PS	193	2	97	.24
CS	7	1	7	.02
MS	188	1	188	.47
PCM	892	2	446	1.12
PCS	466	2	233	.57
PMS	2228	2	1114	2.80
CMS	42	1	42	.11
PCMS	10	2	5	.02
Within (Error)	9556	24	398	
Total	24,513	47		

*P < .01

Table 7

MEAN DIFFERENCES IN TOTAL ERROR SCORE BETWEEN INITIAL AND FINAL PERFORMANCE TESTS

Initial Proficiency Level	Team Consensus Feedback			No Feedback Subjects
	Homogeneous Teams	Heterogeneous Teams	All Feedback Subjects	
High	51.6	42.5	47.1	81.8
Medium	66.9	27.9	47.4	45.8
Low	39.8	31.0	35.4	52.3
All	52.8	33.8	43.3	59.9

Table 8

ANALYSIS OF VARIANCE OF COMBINED FEEDBACK GROUPS AND NO-FEEDBACK GROUP ON TOTAL ERROR DIFFERENCE SCORE

Source of Variation	Sum of Squares	df	Mean Square	F-Ratio
Method (Feedback - No Feedback) (M)	6559	2	3279	6.87*
Team Composition (by Proficiency Level) (P)	2901	2	1451	3.04
M x P	4446	4	1112	2.33
Within Cells	24329	51	477	
Total	38235	59		

*P < .01

Insofar as total error score is concerned, the results of this experiment agree with the results of the previous experiment[5]; in both studies, team consensus feedback resulted in significantly larger performance gains than did the no-feedback method. However, this result held only for certain procedures in the earlier experiment and only for heterogeneous teams in the present experiment. In the first experiment, in which team type was not varied, all teams were heterogeneous in composition.

Errors of Omission

One of the major purposes of the present experiment was to determine if omission errors could be reduced by applying the team consensus feedback method over a larger number of detection practice units. In the previous team feedback experiment, only 15 frames were covered during team practice, whereas in the present experiment an average of 50 frames was covered. Table 9 shows the mean difference scores for omission error and Table 10 gives the associated F-ratios for team consensus feedback subjects. The only difference among the major factors was for team composition, the heterogeneous teams making fewer omission errors.

As with the total error score, the omission error scores were combined for homogeneous and heterogeneous teams and for all proficiency level groups for comparison with no-feedback subjects. None of the F-ratios for omission error were significant (Table 11). The team consensus feedback method had no beneficial effect insofar as omission errors were concerned. In fact, the no-feedback group had a considerably better score than the homogeneous team groups. The results for the present experiment agree with the results of the previous experiment, namely, that omission errors are not reduced by the team consensus method.

Table 9

MEAN DIFFERENCES IN OMISSION ERROR SCORE BETWEEN INITIAL AND FINAL PERFORMANCE TESTS

Initial Proficiency Level	Team Consensus Feedback			No Feedback Subjects
	Homogeneous Teams	Heterogeneous Teams	All Feedback Subjects	
High	45.0	42.2	43.6	38.3
Medium	54.6	43.5	49.1	40.0
Low	61.4	44.9	53.1	44.5
All	53.7	43.5	48.6	40.9

[5] op. cit.

Table 10

DIFFERENCE SCORE F-RATIOS FOR OMISSION AND INVENTIVE ERROR SCORE FOR TEAM CONSENSUS FEEDBACK SUBJECTS

Source	df	F-Ratios	
		Omission Error	Inventive Error
Proficiency Level (P)	2	1.28	12.37**
Team Composition (C)	1	4.34*	6.80*
Feedback Method (M)	1	.02	.55
Team Size (S)	1	.31	2.36
PC	2	.68	.17
PM	2	.42	.45
CM	1	.41	.63
PS	2	.08	3.26
CS	1	.61	.80
MS	1	.56	.01
PCM	2	.05	.35
PCS	2	.01	1.08
PMS	2	3.39	4.23*
CMS	1	.01	.55
PCMS	2	.40	.12
Error (Mean Square)	24	283.44	135.83
Total	47		

*P < .05
**P < .01

Table 11

F-RATIOS FOR COMPARISON BETWEEN COMBINED FEEDBACK GROUPS AND NO-FEEDBACK GROUP IN OMISSION AND INVENTIVE ERROR SCORES

Source	df	F-Ratios	
		Omission Error Score	Inventive Error Score
Method (M)	2	3.01	12.12*
Initial Proficiency Level (P)	2	1.18	6.73*
M x P	4	.30	.07
Within Cells (Mean Square)	51	270.65	322.20
Total	59		

*P < .01

Inventive Errors

If the total error variable shows significant performance gains for the team feedback method and the omission error variable shows no gains, then the gains must be concentrated in the inventive error variable. Table 12 compares the major independent variables on mean inventive error score. Table 10 gives the analysis of variance results for the feedback subjects. The major difference among the factors was agai team composition--heterogeneous teams showed the most improvement. The highly significant difference obtained for proficiency level indicates that interpreters who are initially low in proficiency gain the most from team feedback practice in reducing inventive errors.

Table 11 shows the comparison of the team feedback and the no-feedback groups for inventive error score, with a significant F-ratio at the .01 level for both instructional method and proficiency. These results again show that team feedback practice leads to substantial improvement insofar as inventive error score is concerned, and that the improvement is relatively greater for interpreters with initia' low proficiency scores.

Table 12

MEAN DIFFERENCES IN INVENTIVE ERROR SCORE BETWEEN INITIAL AND FINAL PERFORMANCE TESTS

Initial Proficiency Level	Team Consensus Feedback			No Feedback Subjects
	Homogeneous Teams	Heterogeneous Teams	All Feedback Subjects	
High	8.0	-0.2	3.9	29.0
Medium	2.2	-9.2	-3.5	23.0
Low	-13.0	-19.8	-16.4	5.0
All	-0.9	-9.8	-5.3	19.0

IMPLICATIONS AND CONCLUSIONS

The overall conclusions of the present experiment are essentially the same as those in the previous experiment testing the team feedback method. The interpreters showed a reduction in inventive errors, but no improvement in terms of omission errors. Despite the greater amount of practice imagery provided in this experiment, the teams were evidently not detecting enough targets during practice to provide adequate feedback for omission error avoidance. Methods which lead to more detections by the team might result in improved individual proficiency. One such method would be to permit the entire team to search the same frame at the same time, with each interpreter always aware of all targets which have been found. Since the interpreters would not have to waste time searching for targets which had already been detected, a greater concentration of effort could be applied to every part of each frame. This method would provide instant feedback to team members on a target by target basis. Such minimum delay of reinforcement may lead to greater individual learning. Research testing other team methods is currently under way.

Other conclusions from the present experiment are concerned with differences in feedback procedure. These conclusions hold only for target omissions and inventive errors since the present experiment did not include identification.

The one factor found to be very effective in the present experiment was team composition with respect to initial proficiency level. Teams composed of members whose initial proficiency is heterogeneous show greater gain than do homogeneous teams. Team members who are initially low in proficiency improve relatively more than those who are initially high in proficiency. The poorer interpreters are most probably learning from their interactions with the better interpreters. Evidence indicates that discussion has no effect on the learning of the individual team members. Written communication seems to be as effective as verbal. Whether the team is composed of two or three men also seems to have no effect. However, the possibility exists that teams of more than three men might be more effective.

From both experiments conducted to date, the general conclusion is that on-job training based on team consensus feedback shows promise for reducing identification and inventive errors but limited effectiveness in increasing the number of targets detected. The method should be considered for maintaining and enhancing the performance of interpreters in field units, especially where skilled interpreters can be mixed with relatively inexperienced men. Although there are still many unanswered questions, it appears that such factors as team discussion and team size are probably not as important as having interpreters who are heterogeneous in terms of proficiency assigned to the teams.

Unclassified
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY *(Corporate author)*
System Development Corporation, Falls Church, Virginia
U. S. Army Behavioral Science Research Laboratory, Arlington, Virginia

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
MAINTAINING TARGET DETECTION PROFICIENCY THROUGH TEAM CONSENSUS FEEDBACK

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

5. AUTHOR(S) *(First name, middle initial, last name)*
John T. Cockrell, SDC

6. REPORT DATE
December 1969

7a. TOTAL NO. OF PAGES
36

7b. NO. OF REFS

8a. CONTRACT OR GRANT NO.
DA HC 19-67-C-0040

b. PROJECT NO.
DA R&D Proj. No. 20662704A721

c. Component Integration

d. b-21

9a. ORIGINATOR'S REPORT NUMBER(S)
Technical Research Note 219

9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*

10. DISTRIBUTION STATEMENT
This document has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Office, Chief of Research and Development, Wash., D. C.

13. ABSTRACT
A series of studies, monitored by the Behavioral Science Research Laboratory, is being undertaken in an effort to develop team consensus feedback procedures which will lead to the enhancement of performance of individual interpreters. An exploratory study in the series, reported on in Technical Research Note 195, "Maintaining Image Interpreter Proficiency Through Team Consensus Feedback", was designed to assess the usefulness of the team consensus feedback process as a possible aid to proficiency maintenance for interpreters in an image interpretation facility. The present publication reports on further study in this area, with emphasis on target detection skill. Specifically, the objective of the present experiment was to determine if the target detection skill of individual interpreters can be improved by feedback which team members generate for themselves as they compare and discuss their work. This experiment differed from the preceding exploratory study in that target detection only was required, rather than detection plus identification. In addition, the experimenter investigated the impact on individual interpreter performance of 1) size of team (3-man vs 2-man); 2) discussion vs no discussion; 3) initial proficiency level of team members, and 4) team composition (heterogeneous vs homogeneous) with respect to initial proficiency level. Sixty USAIS graduates participated in the experiment. Treatment was a 3-day practice session. A pre- and post-treatment test was administered to each interpreter to assess detection proficiency. Interpreters assigned to feedback conditions practiced in teams and were permitted to either discuss or compare their work; the no-feedback interpreters practiced alone and were not permitted to discuss or compare their work with anyone. Neither group received ground truth feedback under the experimental procedure.

DD FORM 1 NOV 68 1473 REPLACES DD FORM 1473, 1 JAN 64, WHICH IS OBSOLETE FOR ARMY USE.

Unclassified
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Surveillance systems						
Target detection--skill, proficiency						
*Interpreter proficiency						
Image interpretation facility						
*Team consensual judgment						
*Feedback techniques						
Image interpretation systems						
Laboratory facilities						
*Image interpreter team performance						
Military psychology						
Detection errors						
Team composition						
Initial proficiency level						

DD Form 1473

13. ABSTRACT continued

It was found, as in previous experimentation, that interpreters working in teams with consensus feedback showed greater improvement than interpreters working alone in reducing inventive errors; there was no difference, however, in errors of omission. No difference obtained between discussion vs no-discussion and three-man teams vs two-man teams, but interpreters working in heterogeneous teams showed significantly greater gain in performance on all measures than interpreters on homogeneous teams. Findings also indicated a relatively greater improvement in performance of team members who are initially low in proficiency than those who are initially high in proficiency. From both experiments conducted to date, evidence points to the effectiveness of team consensus feedback in maintaining and enhancing performance of interpreters in field units, particularly in target identification and reduction of inventive errors. The technique appears to be especially useful where ground truth is not available.