1968-06

An experimental study of interpreter proficiency as a criterion for image interpretation personnel assignments.

Schwabe, William Lawrence

Monterey, California. Naval Postgraduate School

https://hdl.handle.net/10945/40097

UNITED STATES
NAVAL POSTGRADUATE SCHOOL

THESIS

AN EXPERIMENTAL STUDY OF INTERPRETER PROFICIENCY

AS A CRITERION FOR IMAGE INTERPRETATION

PERSONNEL ASSIGNMENTS

by

William Lawrence Schwabe

June 1968

[redacted]

AN EXPERIMENTAL STUDY OF INTERPRETER PROFICIENCY

AS A CRITERION FOR IMAGE INTERPRETATION

PERSONNEL ASSIGNMENTS

by

William Lawrence Schwabe
Lieutenant, United States Naval Reserve
B.A., Vanderbilt University, 1964

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 1968

Signature of Author William L. Schwabe

Approved by Gary Poock
Thesis Advisor

Roy H Hurdle for J R Borsting
Chairman, Department of Operations Analysis

W. F. Koehler for R. F. Rinehart
Academic Dean

ABSTRACT

Methods of improving image interpretation system output through use of interpreter proficiency as a criterion for making interpreter personnel assignments were investigated. An experiment was conducted to determine if either of two personnel assignment methods using interpreter proficiency as the assignment criterion would yield significantly improved team performance. No significant difference in performance due to either of the methods tested were found. A second experiment was conducted to determine if assigning the more difficult imagery to the more proficient interpreter would result in higher team performance than random assignment of imagery to team members. Analysis indicated no significant differences in interpreter performance due to either of the methods tested.

The image interpreter personnel assignment problem was formulated as a linear integer program.

TABLE OF CONTENTS

CHAPTER		PAGE

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Image interpretation is one of the best sources of tactical and strategic intelligence. Rapid availability of such intelligence is becoming increasingly important in order to counter the mobility of enemy forces and to utilize fully the rapid strike capabilities of our forces. New sensors, platforms, transmission systems, and "real time" systems are being developed which generate large volumes of imagery from which human interpreters must extract accurate, timely, complete, and relevant information. This has prompted research efforts directed toward improving speed and quality of image interpretation.

I. STATEMENT OF THE PROBLEM

In order to meet future image interpretation requirements it is necessary to find ways to improve and increase image interpretation output. This can be accomplished by training more and better interpreters, improving the performance of interpreters, or making image interpretation tasks less demanding on human interpreters. The purpose of this study was to evaluate two possible methods of improving interpretation output.

Studies sponsored by the U. S. Army Behavioral Science Research Laboratory[(6) (12)], formerly known as the Army Personnel Research Office, indicated that having interpreted imagery checked by another interpreter resulted in improvement in certain measures of interpreter output. It seemed reasonable to assume that interpreters at an interpretation facility would differ in proficiency and that often the relative proficiency of available interpreters would be known. If a check procedure was decided upon and available interpreters could be ranked according to proficiency, task assignments could be made either with or without regard to interpreter proficiency. Given a two-man team, it appeared reasonable that assignment of the initial interpretation task to the lower proficiency interpreter and having the higher proficiency man do the checking would yield higher output than other possible procedures. Experiment I was designed to determine if this was true.

Research has been directed toward development of pre-processing techniques that would provide the interpreter with "advance" information on interpretability of imagery. Such information would be worthwhile if its use resulted in improved interpreter performance. If imagery were pre-processed in such a way that its difficulty of interpretation was known in advance of human interpretation, assigning higher proficiency interpreters to the more difficult imagery might improve output. Experiment II was designed to determine if this was true.

II. IMPORTANCE OF THE STUDY

Several studies of human factors in image interpretation have been initiated since about 1961. Many of the experimental results are not in consensus with the image interpretation community; few of these experimental results can be considered definitive. Some studies have recommended procedures whose feasibility is questionable on account of military considerations or time costs. This study investigated two no-cost, easily implementable interpretation procedures which, if found justified by controlled experimentation, should pose no major feasibility problems.

CHAPTER II

REVIEW OF THE LITERATURE

Interpreter performance is dependent upon characteristics of the imagery, pre-processing of the imagery, training of the interpreters, previous experience and attained level of competence of the interpreters, tactical and strategic information available, equipment available, interpretation procedures, and personnel organization. Research has been undertaken in most of these areas.

I. CHARACTERISTICS OF IMAGERY

The overall quality of photographic and other imagery is improving, due to technological advances; yet imagery quality varies because of variations in conditions under which the imagery is made. A study by Applied Psychology Corporation [2] found that the increase in completeness over time became greater as the quality of imagery was improved. Poor quality imagery yielded negligible increases in completeness over time. It was suggested that imagery below certain quality levels need not be interpreted.

Aerial photographs can be made such that targets present either vertical or oblique aspects. Studies have been made to determine the effects of vertical only, oblique only, and both vertical and oblique. In a pilot study, J. E. Ranes[16] found that simultaneous use of both views--vertical and oblique--of the target area yielded no significant improvement over the vertical view alone. The interpretation task was limited to identification of vehicles in convoy. Results of another study[3] indicate that for mensuration and plotting, vertical views should be used. For objects with major dimensions in the vertical plane, oblique views should be used. Test imagery was limited to one vertical and one oblique view each of a bridge and an airfield, for a total of four photographs. Defending the use of both aspects, R. N. Colwell[10], an eminent member of the photo interpretation community, cited examples where both vertical and oblique views were necessary for correct interpretation of objects.

In the same article Professor Colwell also defended the use of stereo imagery, that is, two photographs of the same area taken by two cameras a small distance apart. The resulting dual photographs are presented to the interpreter in such a way that he is able to use the stereoscopic parallax to obtain three-dimensional information. Schwartz and Zeidner[16], however, found no significant difference between stereo and non-stereo viewing. Their measures of effectiveness were number right and number wrong. No consistent pattern or trend was found to indicate superiority of either stereo or non-stereo viewing. They suggested selective use of stereo viewing.

II. PRE-PROCESSING OF IMAGERY

Interpreter performance might be improved if it were possible to pre-process the imagery in such a way as to reduce the human interpretation task. Ultimately, this would mean complete non-human image interpretation. Research efforts have been made in that direction.

Readers interested in efforts to automate photo interpretation are referred to W. S. Holmes' paper, "Automatic Photo Interpretation and Target Location," in IEEE Proceedings, Vol. 54, No. 12, Dec. 1966, pp. 1679-86, which cites twenty-four references. The attainment of complete automation is not envisioned in the immediate future; however, limited automatic assistance appears to be within the capability of current technology.

One type of automatic assistance which has been investigated is automatic quantification of image quality. If, as several researchers have assumed, interpretability is dependent upon image quality, then knowledge of image quality might be used to predict image interpretability. Cornell Aeronautical Laboratory[11] has developed reliable microdensitometric techniques for measuring and specifying contrast, resolution, edge sharpness, and granularity directly from photographic imagery. Presumably these could be used to quantify image quality and eliminate poor quality imagery as uninterpretable. Measurement of band-widths associated with transition from one tone to another in photographic imagery has been demonstrated by Minneapolis-Honeywell[17] to be a convenient, reliable, and objective method for estimating the ground resolution of photography.

Manual determination of image quality might be helpful, especially if it did not require highly trained personnel. A catalog technique[8] has been developed by which interpreters compare their imagery with catalog imagery and assign predicted interpretability values. Discrimination of target areas from non-target areas by the catalog technique was correlated with results of actual interpretations, yielding correlations of .77 for trained interpreters and .70 for untrained personnel. Correlations of predicted accuracy of target identification with interpreted accuracy of target indentification were .54 and .51, respectively. Average time per test image was 45 seconds. The close correlation values between trained and untrained interpreter performance suggests this might be an effective way to reduce the work load of photo interpreters by using less skilled personnel.

Another approach to predicting interpretability[14] is to record various data from an initial interpretation and from these compute probable accuracy of the interpretation and probable utility of future search. This information could form the basis of a decision rule to indicate whether or not the imagery should be check interpreted.

Pre-processing might also take the form of automatic enhancement of image interpretability. One technique[7] involves obtaining a video signal from a transparency and adding to this signal its negative second derivative. This so-called "differentation enhancement technique" appeared to improve performance principally by increasing the number of correct responses, and, to a lesser extent, by decreasing the number of incorrect responses. It has been found to be better suited for more difficult imagery.

A Boeing study[5] recommended that interpreters view alternately flashing superimposed photographs of the same area taken at two different times. This technique causes an apparent motion of elements in the photography which changed during the time interval between exposures. As noted in the study, the effectiveness of the technique is dependent upon the amount and complexity of background image disparities.

III. FEEDBACK INFORMATION AVAILABLE TO INTERPRETERS

Feedback information can come either from external intelligence sources or from the image interpretation operation itself. A. E. Castelnovo[16] investigated the effects of different levels of externally provided intelligence information on photo interpretation and found that, for a sequence of imagery, increased information aided the photo interpreters initially, but after a short period of time it had no effect. He pointed out that the negative effects of erroneous intelligence must also be considered--something he did not measure experimentally. This suggested the possibility that an increased amount of accurate intelligence information might not necessarily be of significant value in the long run.

Photo interpreters commonly assign subjective confidence estimates to their interpretations. The reliability of these estimates varies. Measured reliability for completeness judgments ranges from .45 to .88[14] and for accuracy judgments from .27 to .83.[14] [12] If evaluations of their previous performance are fed back to photo interpreters, their subsequent confidence estimates of accuracy are significantly improved.[21] One problem with this, however, is that such feedback is not available in actual photo interpretation situations.

A similar but more sophisticated technique is to record several performance measures, such as time to first target detection, as well as confidence ratings, and compute probabilistic ratings of accuracy and completeness which are fed back to the interpreters. Use of this procedure has been found to reduce the number of subsequent incorrect identifications, but did not significantly improve accuracy, completeness, or speed of identification.[14] If reliable confidence ratings can be obtained, a decision rule must be formulated to determine which imagery is to be check interpreted. No experimental work has yet been attempted to select such an optimal decision rule.

IV. TRAINING OF INTERPRETERS

R. N. Colwell's article[9] summarized some of the methods currently used in training photo interpreters. H. W. Leibowitz[13] has advocated using the programmed instruction technique. He argued convincingly from results of experimental psychology in perceptual learning that programmed instruction would be significantly more efficient than presently used lecture presentation. RCA[18] conducted a special four-day training program in photo interpretation using tachistoscopic techniques similar to those used in reading improvement courses to increase interpreters' speed of detection and classification of targets. Comparisons between experimental and control group proficiency measures showed statistically significant performance improvement due to the special training. The experimental group extracted information from the test photography in one half the viewing time required by the control group, with slight gains in completeness and accuracy. Moreover, the effect of the training was such as to counteract deteriorating effects of diminished scale and increased number of targets per photograph.

V. INTERPRETATION PROCEDURES AND WORKING CONDITIONS

Work load can be expected to vary, especially at "front line" interpretation facilities. For conventional (non-stereo) large scale imagery, 60 ft./90 min. is considered an acceptable low input quantity for average interpreters to view. A rate of 120 ft./90 min. would constitute a high input.[1]

Performance has been found to fluctuate during the working day, but not in a consistent manner. A one-day experiment suggested that there was no decrement in performance even during a work day of extended length, twelve hours, containing no rest periods and providing only short periods for lunch and dinner.[1] The same study found low correlation between expressions of fatigue and performance.

Interpreters can vary their performance as a function of the relative weights given to accuracy and completeness, but unless they are given guidance, they will base their work methods on their own subjective and highly variable conception of the intelligence objectives.[20]

Another study[2] found that completeness increased generally with increased viewing time. The study suggested that, given large quantities of photography on which features are to be identified accurately, one minute viewing time per photograph yielded high performance. A 48 minutes on and 5 minutes off work-rest cycle was recommended.

Aero Service Corporation[1] found that when short (25 ft.) samples of imagery were interpreted, an acceptable methodology was to proceed directly to interpretation without first rapidly screening the imagery. Willmorth and Birnbaum[22] [23] recommended that neither screening nor overlapping of imagery be used for rapid interpretation.

Investigating interpreter team organization, Bolin, Sadacca, and Martinek[6] found no single factor or principle of team organization that led to improved performance in all types of missions. Continuing this line of research, Doten, Cockrell, and Sadacca[12] found that teams in which the check interpreter had complete knowledge of the initial interpreter's work produced more complete results with higher efficiency than did procedures utilizing only partial knowledge of initial interpretation. Arbitrary checking (where the checker made final judgments without consulting the initial interpreter), consensus checking (where only those interpretations agreed upon without discussion were recorded), and discussion-consensus checking (where only those interpretations agreed upon after discussion were recorded) procedures were tested. Introduction of a third man provided more completeness but reduced efficiency. No differences in team output from different procedures with the three-man team were noted. The checking procedure with arbitrary scoring resulted in the highest completeness but lowest accuracy. Checking procedure with consensus yielded higher accuracy but less complete interpretation. Discussion with consensus scoring gave both high accuracy and completeness but reduced efficiency.

VI. RELATION TO PREVIOUS RESEARCH

No analyses based on operational data were found in the literature on human factors research in image interpretation. Most studies conducted to date have used advanced photo interpreter trainees as subjects in controlled experiments. The all-but-insurmountable difficulty encountered when attempting to use operational environments as sources of data is the measurement of interpreter performance. More precisely, the difficulty is in defining the imagery's ground truth. All standard measures of interpreter performance--accuracy, completeness, conciseness (the ratio of accuracy to time), and efficiency--are dependent upon ground truth. Definition of ground truth is a tedious process, generally accomplished by consensus decision following careful interpretation by a team of expert photo interpreters. Nevertheless, if ground truth is known, it should be possible to insert that imagery between sequences of operational imagery. Such a procedure might provide more reliable indication of operational interpreter performance than that obtained from presently employed procedures.

This study was constrained by lack of trained image interpreters; however, it was felt that the important factors in team studies would be found in experiments using untrained subjects.

Simulated photo imagery was constructed because (1) its composition could be controlled exactly, (2) ground truth could be determined easily, and (3) symbolic targets could be used. Untrained subjects would be expected to find identification of objects in aerial photographs inordinately difficult, on account of their lack of experience in identifying objects from vertical or high oblique aspects. It was felt that use of more familiar symbolic targets would result in a better balance of identification, classification, and evaluation difficulties for untrained subjects than would use of actual aerial photography. Test image interpretability was designed to be dependent upon target density, shape, markings, scale, contrast with background, resolution, detail, and spatial location, as well as background noise.

Experiments I and II were meant to complement the APRO team studies cited. Experiment II, in addition, was designed to complement the pre-processing studies cited.

CHAPTER III

EXPERIMENTAL DESIGN AND RESULTS OF THREE EXPERIMENTS ON IMAGE INTERPRETATION PERSONNEL ASSIGNMENT CRITERIA

I. ELEMENTS COMMON TO ALL EXPERIMENTS

A Preliminary Experiment was conducted in order to provide data necessary for Experiments I and II, each of which was directed toward one of the primary objectives of the present study. Certain methodological elements were common to the three experiments.

Interpretation Tasks

The interpretation tasks used in the study consisted of two subsets of activities. These were:

1. Initial interpretation. Interpreters worked independently on separate parts of the imagery, completing annotations and target identifications.

2. Checking. Interpreters checked their teammates' initial interpretations and looked for additional targets.

Team Scoring Rules

A scoring rule was defined as a means of combining individual output into a team output. The two basic scoring rules were:

1. Arbitrary. Score all responses which checkers approve or make.

2. Combined. Score and sum all responses which both teammates make using the initial interpretation procedure.

Dependent Variables.

Three measures of individual interpreter performance were used:

1. Accuracy. Ratio of right interpretation to the sum of right plus wrong interpretations.

2. Completeness. Ratio of right interpretations to the total possible rights, i.e., the total number of scored targets in the imagery.

3. Efficiency. Number of right interpretations divided by the total amount of time required in minutes.

Experimental Subjects

Twenty-four Army and Marine Corps officers enrolled in the Operations Research program at the Naval Postgraduate School constituted the population of subjects for the three experiments. Their rank distribution was: 12 captains, 10 majors, and 2 lieutenant colonels. One subject was a pilot. One subject had previous experience in photo interpretation.

Subject Proficiency

It was assumed that the subjects had acquired some degree of proficiency in detection and classification tasks other than image interpretation which would give them individually varying proficiency in image interpretation tasks. It was further assumed that the subjects' proficiencies could be measured and the subjects ordered according to those proficiency measurements, and that this ordering would not change during the course of experimentation, due to learning or any other cause.

Experimental Imagery

Sixty 7" x 8" image frames were hand drawn on 8 1/2" x 11" white paper. These image frames were assembled into six different ten frame imagery sets. Twenty-four Xerox copies of each set were made; this quantity was sufficient to insure that no subject would view any of the imagery more than once during the course of the experimentation. Original copies of the imagery were drawn in red, black, blue, and green ink; this produced controlled differences in contrast ratios in the Xerox copies.

The image frames were intended to simulate photographs of targets in the vicinity of a border between two countires. Fifteen classes of targets were represented. Appendix A contains samples of the test imagery.

II. PRELIMINARY EXPERIMENT: MEASUREMENT OF INTERPRETER PROFICIENCY AND IMAGERY DIFFICULTY

Experimental Objectives

The objectives of the Preliminary Experiment were (1) to measure interpreter proficiency, (2) to measure imagery difficulty, and (3) to determine if there was any significant difference in difficulty among the six sets of imagery used.

Dependent Variables

Individual interpreter proficiency was calculated from individual accuracy, completeness, and efficiency scores according to the following formula:

$$\text{Proficiency}_i = 1/3\,(\text{Accuracy}_i + \text{Completeness}_i + \text{Normalized Efficiency}_i)$$

Where

$$\text{Normalized Efficiency}_i = \frac{\text{Efficiency}_i - \min_i(\text{Efficiency}_i)}{\max_i(\text{Efficiency}_i) - \min_i(\text{Efficiency}_i)}$$

$$i = 1,2,\ldots,24$$

Image frame difficulty was determined according to:

$$\text{Difficulty}_j = 1 - 1/2\,(\text{Mean Accuracy Frame}_j + \text{Mean Completeness Frame}_j)$$

$$j = 1,2,\ldots,60$$

Experimental Design

The experimental design to test effects of different imagery sets on interpreter performance is shown in Figure 1. Assignment of imagery to subjects was random, subject to the balance requirements that (1) each subject interpret two different imagery sets, and (2) each imagery set be interpreted by eight subjects.

Experimental Procedures

Each subject was given two imagery sets of ten image frames each. A separate interpretation key, showing examples of each type of target, was provided. The interpreter was required to circle or draw an arrow to each target detected and label each with a number. The numbers were

PERFORMANCE MEASURE	Set A	Set B	Set C	Set D	Set E	Set F
Accuracy	S 1 2 7 9 13 16 19 23	S 2 3 8 10 14 17 20 24	S 3 4 9 11 15 18 19 21	S 4 5 10 12 13 16 20 22	S 5 6 7 11 14 17 21 23	S 1 6 8 12 15 18 22 24
Completeness	S 1 2 7 9 13 16 19 23	S 2 3 8 10 14 17 20 24	S 3 4 9 11 15 18 19 21	S 4 5 10 12 13 16 20 22	S 5 6 7 11 14 17 21 23	S 1 6 8 12 15 18 22 24
Efficiency	S 1 2 7 9 13 16 19 23	S 2 3 8 10 14 17 20 24	S 3 4 9 11 15 18 19 21	S 4 5 10 12 13 16 20 22	S 5 6 7 11 14 17 21 23	S 1 6 8 12 15 18 22 24

IMAGERY

Figure 1. Preliminary Experiment Design

then entered on appropriate lines on the target identification form printed below the image frame.

After an interpreter completed his first set of ten image frames, he recorded the time, measured in 15 second increments, and commenced work on the remaining ten image frame set immediately. On completing the entire twenty image frames, total time was recorded. Interpreters were instructed to work independently, without going back to completed frames, pacing themselves in order to maximize their accuracy, completeness, and efficiency scores.

Results

Accuracy, completeness, and efficiency scores for each subject were tabulated and are presented in Table I. From these, proficiency scores were calculated, and subjects were ranked in order of decreasing proficiency scores, as shown in Table II.

Image frames were ranked in order of decreasing computed difficulty within each imagery set, as shown in Table III.

Analysis of variance. A 6 x 3 factorial analysis of variance was performed using the data summarized in Table I. Results of the analysis of variance are shown in Table IV. All tests of hypotheses were made at a five per cent significance level. Differences due to performance measures were statistically significant. Differences due to imagery sets were not significant. Interaction between imagery sets and performance measures was not significant.

TABLE I

ACCURACY, COMPLETENESS, AND EFFICIENCY SCORES FOR SUBJECTS IN THE PRELIMINARY EXPERIMENT

Imagery	Accuracy	Completeness	Efficiency
Set A	.981	.684	4.952
	1.000	.829	5.040
	.972	.897	4.667
	.986	.947	5.053
	.928	.842	5.224
	.970	.842	7.758
	1.000	.921	7.778
	1.000	.868	8.000
Set B	.984	.795	6.359
	.848	.719	3.672
	.932	.885	5.520
	.986	.923	5.878
	.969	.795	5.905
	.969	.808	6.811
	.983	.731	8.769
	.901	.936	6.791
Set C	.896	.759	4.898
	.986	.886	4.118
	.959	.899	5.796
	.986	.886	5.000
	.914	.810	4.830
	.947	.899	4.982
	.929	.823	7.647
	.987	.949	8.571
Set D	.957	.868	4.800
	.938	.904	5.000
	.985	.882	5.154
	.926	.829	4.500
	.941	.842	6.919
	.892	.763	5.800
	1.000	.961	6.952
	.957	.882	7.053

TABLE I (continued)

Imagery	Accuracy	Completeness	Efficiency
Set E	.956	.929	5.098
	.938	.779	4.286
	.952	.779	4.898
	.973	.922	7.100
	.970	.844	6.341
	1.000	.844	4.906
	1.000	.909	6.829
	1.000	.870	5.360
Set F	.969	.539	4.824
	.947	.934	4.897
	.973	.934	5.680
	.986	.908	7.459
	.933	.737	4.148
	.962	.671	3.778
	.956	.855	4.561
	.984	.829	8.129

TABLE II

RANKING OF SUBJECTS IN ORDER OF DECREASING PROFICIENCY, BASED ON PRELIMINARY EXPERIMENT DATA

Subject Number	Proficiency	Rank
1	.606	22
2	.853	5
3	.706	17
4	.955	1
5	.873	4
6	.612	21
7	.790	10
8	.740	11
9	.697	18
10	.737	14
11	.652	20
12	.602	23
13	.719	16
14	.547	24
15	.739	12
16	.810	8-9
17	.820	7
18	.927	2
19	.738	13
20	.926	3
21	.674	19
22	.810	8-9
23	.831	6
24	.720	15

TABLE III

RANKING OF IMAGE FRAMES WITHIN SETS IN ORDER OF DECREASING DIFFICULTY, BASED ON PRELIMINARY EXPERIMENT DATA

	Frame	Difficulty	Rank
Set A	1	.030	10
	2	.082	2
	3	.039	8
	4	.055	7
	5	.223	1
	6	.078	3
	7	.069	4
	8	.031	9
	9	.063	6
	10	.065	5
Set B	1	.076	8
	2	.103	6
	3	.162	1
	4	.112	4
	5	.110	5
	6	.141	2
	7	.133	3
	8	.087	7
	9	.000	10
	10	.031	9
Set C	1	.050	8
	2	.063	6
	3	.088	4
	4	.068	5
	5	.062	7
	6	.096	3
	7	.113	2
	8	.290	1
	9	.000	10
	10	.042	9

TABLE III (continued)

	Frame	Difficulty	Rank
Set D	1	.089	6
	2	.096	5
	3	.079	7
	4	.100	4
	5	.057	8
	6	.104	3
	7	.254	2
	8	.035	9
	9	.312	1
	10	.031	10
Set E	1	.038	8
	2	.167	2
	3	.096	4
	4	.093	5
	5	.013	9
	6	.297	1
	7	.156	3
	8	.053	7
	9	.000	10
	10	.063	6
Set F	1	.129	5
	2	.088	10
	3	.120	6
	4	.090	9
	5	.108	7
	6	.143	1
	7	.141	2
	8	.107	8
	9	.131	4
	10	.133	3

TABLE IV

PRELIMINARY EXPERIMENT
ANALYSIS OF VARIANCE

Source	df	SS	MS	F
Performance Measure	2	764.245	382.123	
Imagery	5	1.216	0.243	0.356
PM x I	10	2.445	0.245	0.359
Residual	126	85.934	0.682	
Total	143	853.840		

Discussion and conclusions. The formula used to calculate interpreter proficiency was selected arbitrarily to yield proficiency scores in the unit interval [0,1]. This formula weighted efficiency more highly than accuracy or completeness, which did not seem unreasonable. The relative importance of accuracy, completeness, and efficiency in the real world can be expected to vary with changing tactical and strategic image interpretation requirements; hence, no one formula for proficiency can be said to be best for all situations. Likewise, the formula for calculating image difficulty was chosen arbitrarily to yield scores in the unit interval [0,1].

Data on performance of interpreters using different sets of imagery indicated that use of any particular set of imagery did not bias an interpreter's performance scores relative to those of other interpreters. There was no significant interaction between imagery and performance measurements. The importance of these results was that it permitted comparisons of scores among interpreters using dissimilar test imagery. The significant main effect due to performance measures indicated that the performance measure factor should be included in the design of Experiments I and II.

III. EXPERIMENT I: PROFICIENCY AS THE CRITERION FOR ARBITRARY CHECK PROCEDURE PERSONNEL ASSIGNMENTS

Experimental Objective

The objective of Experiment I was to determine if either of two personnel assignment methods using interpreter proficiency as the assignment criterion would yield significantly improved team performance.

Personnel Assignment Methods

The following three personnel assignment methods were employed:

1. Low-initial/High-check. Initial interpretation was performed by the lower proficiency team member. Checking was done by the higher proficiency team member.

2. High-initial/Low-check. Initial interpretation was performed by the higher proficiency team member. Checking was done by the lower proficiency team member.

3. Random initial/check. Initial interpretation and check interpretation personnel assignments were made without regard to interpreter proficiency.

All personnel assignment methods were scored according to the arbitrary scoring rule.

Experimental Design

Experiment I design, to test effects of proficiency as a criterion for check task assignments, is shown in Figure 2. Subjects were assigned to two-man teams in restricted randomized fashion, subject to the requirements that (1) subjects who interpreted any of the same imagery in the Preliminary Experiment were ineligible for membership on the same team, and (2) each team was composed of one subject from among the twelve most proficient interpreters and one subject from among the twelve least proficient interpreters. Restriction (1) was necessary in order to use imagery annotated in the Preliminary Experiment as material to be checked in Experiment I without any subject's checking

ASSIGNMENT CRITERION	Accuracy	Completeness	Efficiency
High-Initial/Low Check	<u>Ss</u> 4,24 5,3 16,11 18,13 2,6 17,1 23,12 20,19	<u>Ss</u> 4,24 5,3 16,11 18,13 2,6 17,1 23,12 20,19	<u>Ss</u> 4,24 5,3 16,11 18,13 2,6 17,1 23,12 20,19
Low-Initial/High-Check	<u>Ss</u> 1,17 3,5 21,22 19,20 14,15 9,8 24,4 11,16	<u>Ss</u> 1,17 3,5 21,22 19,20 14,15 9,8 24,4 11,16	<u>Ss</u> 1,17 3,5 21,22 19,20 14,15 9,8 24,4 11,16
Random Initial/Check	<u>Ss</u> 10,7 13,18 15,14 22,21 12,23 7,10 6,2 8,9	<u>Ss</u> 10,7 13,18 15,14 22,21 12,23 7,10 6,2 8,9	<u>Ss</u> 10,7 13,18 15,14 22,21 12,23 7,10 6,2 8,9

PERFORMANCE MEASURE

Figure 2. Experiment I Design.

imagery he had interpreted previously. Data generated during the Preliminary Experiment test session and during the Experiment I test session were combined, so that, in effect, each team interpreted the same imagery using both the Low-initial/High-check and High-initial/Low-check procedures. Restriction (2) was designed to accentuate any differences between the two check procedures by encouraging wider range in proficiency between team members. Teams were grouped so that Group 1 consisted of eight subjects using High-initial/Low-check procedure, Group 2 consisted of eight subjects using Low-initial/High-check procedure, and Group 3 consisted of eight subjects using the random initial/check procedure. This grouping is graphically illustrated in Figure 3.

Experimental Procedures

A group testing session was held one day after the Preliminary Experiment, during which session both Experiment I and Experiment II were conducted.

Each subject was given the two imagery sets that had been interpreted in the Preliminary Experiment by his teammate. Subjects were instructed to check their teammates' interpretations, making corrections when appropriate, and to interpret additional targets missed by the initial interpreters. Time was recorded after the first ten frames were checked and after completion of the entire twenty frames. Checkers were instructed to work independently without going back to completed frames, pacing themselves in order to maximize team accuracy, completeness, and efficiency scores. Separate interpretation keys were provided.

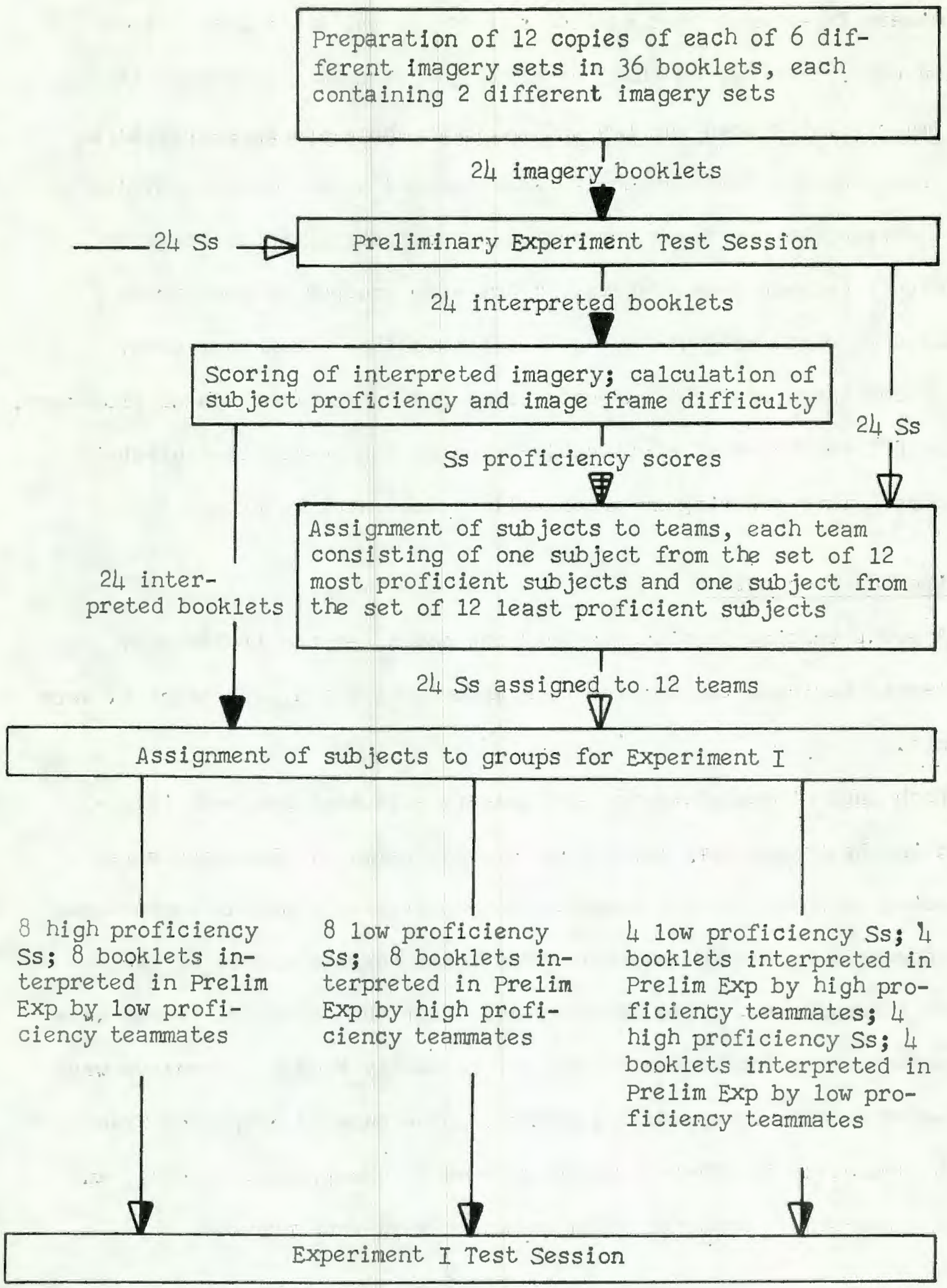


Figure 3. Experiment I Flow Chart.

Results

Team accuracy, completeness, and efficiency scores are shown in Table V. Table VI presents initial interpretation scores and incremental scores resulting from check interpretation.

Analysis of variance. A 3 x 3 factorial analysis of variance was performed. Data used are presented in Table V. Results are shown in Table VII. All tests of hypotheses were made at a five per cent significance level. Differences due to performance measures were statistically significant. Differences due to personnel assignment criteria were not significant. Interaction between personnel assignment criteria and performance measures was not significant.

Discussion and conclusions. Data on performance of interpretation teams using different personnel assignment criteria indicated that none of the three criteria was to be preferred to any other. This was an unexpected result, for it had been assumed that the Low-initial/High-check procedure would prove superior to the other procedures. An attempt was made to account for this result. It was noted from the data in Table VI that mean low proficiency initial interpretation scores were below mean high proficiency initial interpretation scores; the Low/High ratios for accuracy, completeness, and efficiency were .985, .948, and .732, respectively. In addition, the means of low proficiency incremental scores due to checking were below those of high proficiency incremental checking scores; the Low/High ratios in this case were .188, .505, and .667, respectively. Team performance was determined by combining initial and incremental checking scores.

TABLE V

ACCURACY, COMPLETENESS, AND EFFICIENCY SCORES FOR TEAMS IN EXPERIMENT I

Procedure	Accuracy	Completeness	Efficiency
	.973	.942	3.058
	.979	.922	4.241
	.943	.868	4.000
High-Initial Low-Check	.971	.877	3.750
	.943	.955	4.975
	.965	.710	3.359
	1.000	.941	4.028
	.992	.857	4.981
	.960	.783	3.748
	.986	.917	3.623
	.979	.960	3.425
Low-Initial High-Check	.986	.910	3.639
	.925	.877	2.941
	.961	.930	3.539
	.973	.942	3.314
	.957	.859	3.646
	.960	.942	4.920
	.987	.925	3.322
	.980	.942	3.240
Random Initial Check	.980	.947	3.972
	.960	.928	3.337
	.966	.928	4.028
	.986	.915	3.256
	.973	.922	4.000

TABLE VI

INITIAL INTERPRETATION SCORES
AND INCREMENTAL SCORES DUE TO CHECKING

	Accuracy	Completeness	Efficiency	Accuracy Increment	Completeness Increment	Efficiency Increment
	.892	.748	3.966	.021	.026	- 2.068
	.938	.777	5.545	- .021	.046	- 2.631
	.930	.783	4.372	.000	.072	- 2.382
	.978	.865	5.414	.000	.013	- 2.666
	.969	.612	4.895	- .005	- .116	- 3.205
Lower	.961	.925	5.026	- .014	.029	- 2.902
Half	.939	.909	5.185	.000	.026	- 4.722
	.932	.795	4.863	.008	.065	- 2.685
	.972	.896	5,308	.015	.046	- 2.039
	.971	.906	4,538	- .006	.052	- 2.140
	.979	.892	4.828	.047	.026	- 1.520
	.983	.758	4.336	- .014	.020	- 3.019
Mean Low	.954	.822	4.855	.003	.045	- 2.665
	.959	.916	5.308	.033	.129	- 1.025
	1.000	.876	6.872	.048	.140	- 1.922
	1.000	.869	6.410	.030	.145	- .990
	.992	.844	7.647	.008	.045	- 1.775
	.970	.826	6.564	- .009	.171	- 1.147
Higher	.955	.826	7.877	.026	.000	- 1.704
Half	.973	.916	7.780	.034	.033	- 1.871
	.931	.803	6.685	.025	.064	- 1.217
	.951	.882	6.067	- .012	.046	- .488
	.986	.895	6.112	.008	.054	- 1.083
	.926	.896	5.520	- .018	.038	- 1.289
	.985	.857	6.769	.003	.157	- 1.080
Mean High	.969	.867	6.634	.015	.085	- 1.299

TABLE VII

EXPERIMENT I
ANALYSIS OF VARIANCE

Source	df	SS	MS	F
Performance Measure	2	125.823	62.912	
Assignment Criterion	2	0.598	0.299	1.246
PM x AC	4	0.856	0.214	0.892
Residual	63	15.103	0.240	
Total	71	142.380		

The entries in Table VIII were obtained by adding mean low proficiency initial scores to mean high proficiency incremental check scores and by adding mean high proficiency initial scores to mean low proficiency check incremental scores for accuracy, completeness, and efficiency. The ratios of the Low/High sums to the High/Low sums were .997, .995, and .896. These were closer to unity than were the initial ratios or the incremental ratios. Thus, checking served to balance out differences between high and low initial interpretations. Lest there by any temptation to conclude from these figures that a High-initial/High-check procedure would yield significantly higher team performance, it should be noted here that Doten, Cockrell, and Sadacca(12) found the performance of High/Low proficiency teams (each man checking the other's initial interpretations) to be better than High/High proficiency teams. These two findings are not necessarily inconsistent because Experiment I yielded data on high proficiency check incremental scores to low proficiency initial scores from which nothing can be deduced about high proficiency increments to high proficiency initial scores. If the results of both the Doten study and Experiment I were valid, then it would follow that mean High/High incremental scores could be expected to be less than mean Low/High incremental scores.

TABLE VIII

INITIAL AND INCREMENTAL PERFORMANCE RATIOS

Mean Scores	Accuracy	Completeness	Efficiency
Mean Low Initial	.954	.822	4.855
Mean High Initial	.969	.867	6.634
Mean Low / Mean High Initial	.985	.948	.732
Mean Low Increment	.003	.045	- 2.665
Mean High Increment	.015	.085	- 1.299
Mean Low / Mean High Increment	.188	.505	.667
Mean Low Initial + Mean High Increment	.969	.907	3.969
Mean Low/Mean High / Mean High/Mean Low	.997	.995	.896

IV. EXPERIMENT II: INTERPRETER PROFICIENCY AND IMAGERY DIFFICULTY AS CRITERIA FOR ASSIGNMENT OF IMAGERY TO INTERPRETERS

Experimental Objective

The objective of Experiment II was to determine if assigning the more difficult imagery to the more proficient interpreter and the easier imagery to the less proficient interpreter would result in significantly higher team performance than random assignment of imagery to team members.

Imagery Assignment Methods

The two imagery assignment methods were:

1. Presorted. Each interpreter received equal quantities of imagery for interpretation. All imagery given to the lower proficiency team member was less difficult than any of the imagery given to the higher proficiency team member.

2. Unsorted. Imagery was assigned to team members without regard to its difficulty or interpreter proficiency.

Experimental Design

Experiment II design, to test effects of interpreter proficiency and imagery difficulty as criteria for interpretation task assignments, is shown in Figure 4. Composition of the twelve teams remained the same as in Experiment I. Each team interpreted unannotated imagery of known difficulty not previously viewed by either team member. Teams were randomly grouped such that Group 1 consisted of six teams using the

ASSIGNMENT CRITERION	Accuracy	Completeness	Efficiency
Unsorted	<u>Ss</u> 14,15 3,5 12,23 19,20 1,17 13,18	<u>Ss</u> 14,15 3,5 12,23 19,20 1,17 13,18	<u>Ss</u> 14,15 3,5 12,23 19,20 1,17 13,18
Sorted	<u>Ss</u> 24,4 11,16 10,7 21,22 9,8 6,2	<u>Ss</u> 24,4 11,16 10,7 21,22 9,8 6,2	<u>Ss</u> 24,4 11,16 10,7 21,22 9,8 6,2

PERFORMANCE MEASURE

Figure 4. Experiment II Design.

Unsorted imagery interpretation procedure while Group 2 consisted of six teams using the Presorted imagery interpretation procedure. Team grouping and imagery flow is shown in Figure 5.

Experimental Procedures

Interpretation procedures were similar to those of the Preliminary Experiment, except as noted. Each team was given twenty image frames to interpret. Those using the Presorted assignment method were given imagery in ten frame presorted booklets. Those using the Unsorted assignment method were given a stack of twenty unsorted imagery frames with team members being instructed to take an image frame off the top of the stack after each frame was interpreted, until the stack was exhausted. Each interpreter recorded the time when he finished all the imagery assigned to him. Interpreters were instructed to work independently without going back to completed frames, pacing themselves in order to maximize team accuracy, completeness, and efficiency scores. Users of presorted imagery were not told the imagery was presorted.

Results

Accuracy, completeness, and efficiency scores for each team were tabulated and are presented in Table IX.

Analysis of variance. A 2 x 3 factorial analysis of variance was performed. Results of the analysis are shown in Table X. All tests of hypotheses were made at a five per cent significance level. As before, differences due to performance measures were statistically significant.

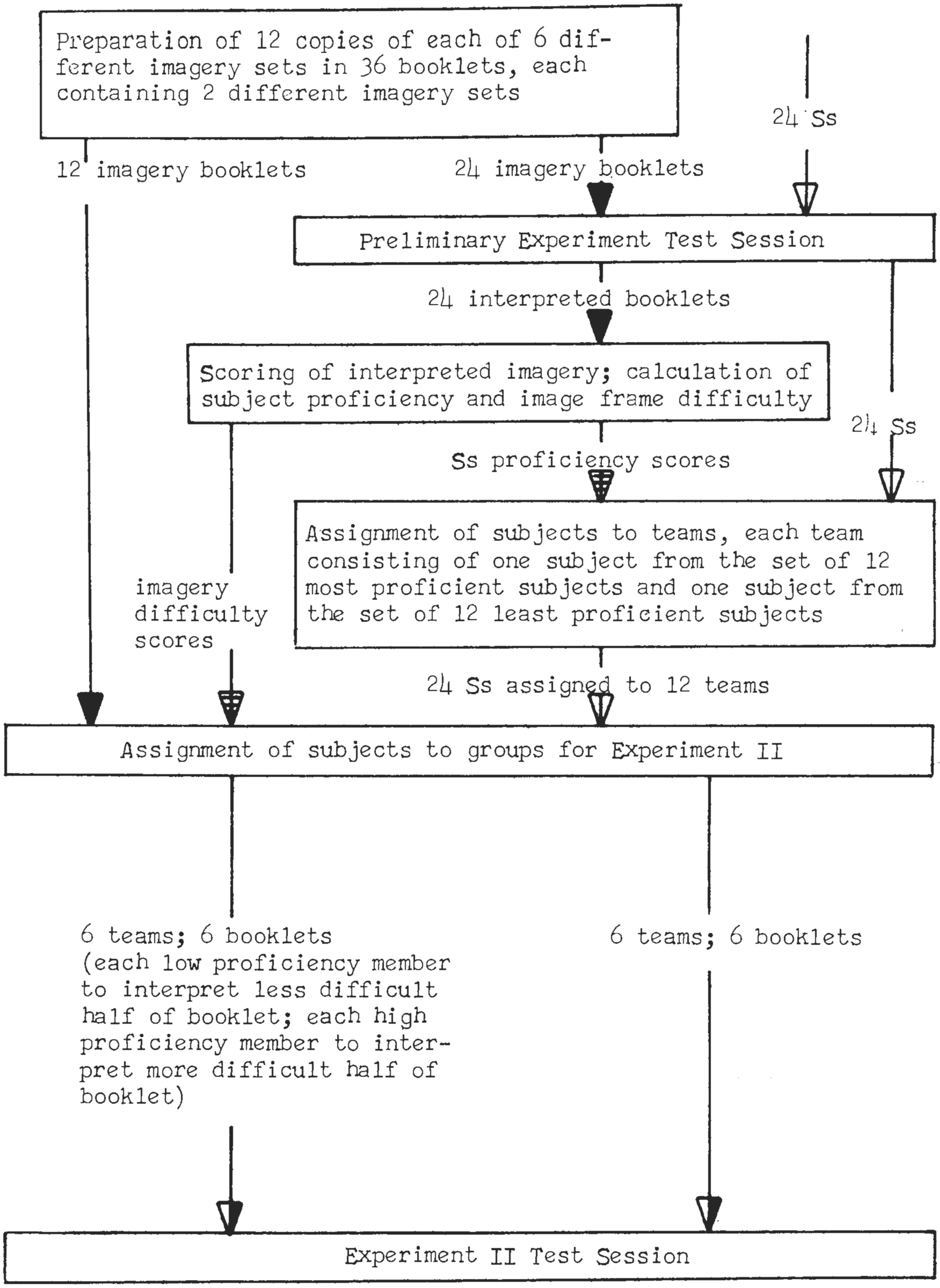


Figure 5. Experiment II Flow Chart.

TABLE IX

ACCURACY, COMPLETENESS, AND EFFICIENCY SCORES FOR TEAMS IN EXPERIMENT II

Procedure	Accuracy	Completeness	Efficiency
Unsorted	.960	.774	7.619
	.980	.955	7.688
	.960	.941	7.526
	.965	.902	8.118
	.993	.962	7.023
	.944	.882	6.022
Pre-sorted	.993	.954	7.392
	.973	.922	7.889
	.965	.890	7.667
	.972	.890	7.211
	.986	.903	6.829
	.980	.961	7.840

TABLE X

EXPERIMENT II
ANALYSIS OF VARIANCE

Source	df	SS	MS	F
Performance Measure	2	333.880		
Assignment Criterion	1	0.028	0.028	0.237
PM x AC	2	0.031	0.016	0.136
Residual	30	3.542	0.118	
Total	35	337.481		

Differences due to task assignment criteria for accuracy, completeness, and efficiency were not significant. Interaction between task assignment criteria and measures of team performance was not significant.

Discussion and conclusions. Data on team performance using different task assignment criteria indicated that neither the Presorted method nor the Unsorted method was to be preferred. If Presorting involved additional cost, the Unsorted method would be preferred. Results of this experiment suggested that development and subsequent procurement of equipment to pre-process imagery by predicting image difficulty would not be cost effective.

CHAPTER IV

OPTIMAL ASSIGNMENT OF PERSONNEL TO IMAGE INTERPRETATION TASKS

A linear integer programming formulation of the problem of optimal utilization of personnel was developed and is presented in this Chapter. The solution is dependent upon knowledge of several assumed constant terms; these are:

1. Number of each class (high proficiency and low proficiency) of interpreters available;

2. Flow of imagery into the system, expressed in expected number of targets per unit time;

3. Expected efficiency of each class of interpreter for initial interpretation, for check interpretation of initial work done by an interpreter of his class, and for check interpretation for initial work done by an interpreter of the other class; and

4. Expected performance measures (accuracy or completeness) of each class of interpreter corresponding to the various efficiency measures.

The system can be depicted in the flow chart format of Figure 6, where

1. X_i is the number of class i interpreters available, i = 1,2;

2. x_i is the number of class i interpreters utilized as initial interpreters, i = 1,2;

3. x_{ij} is the number of class j interpreters utilized as checkers of class i initial interpretations, i = 1,2, j = 1,2;

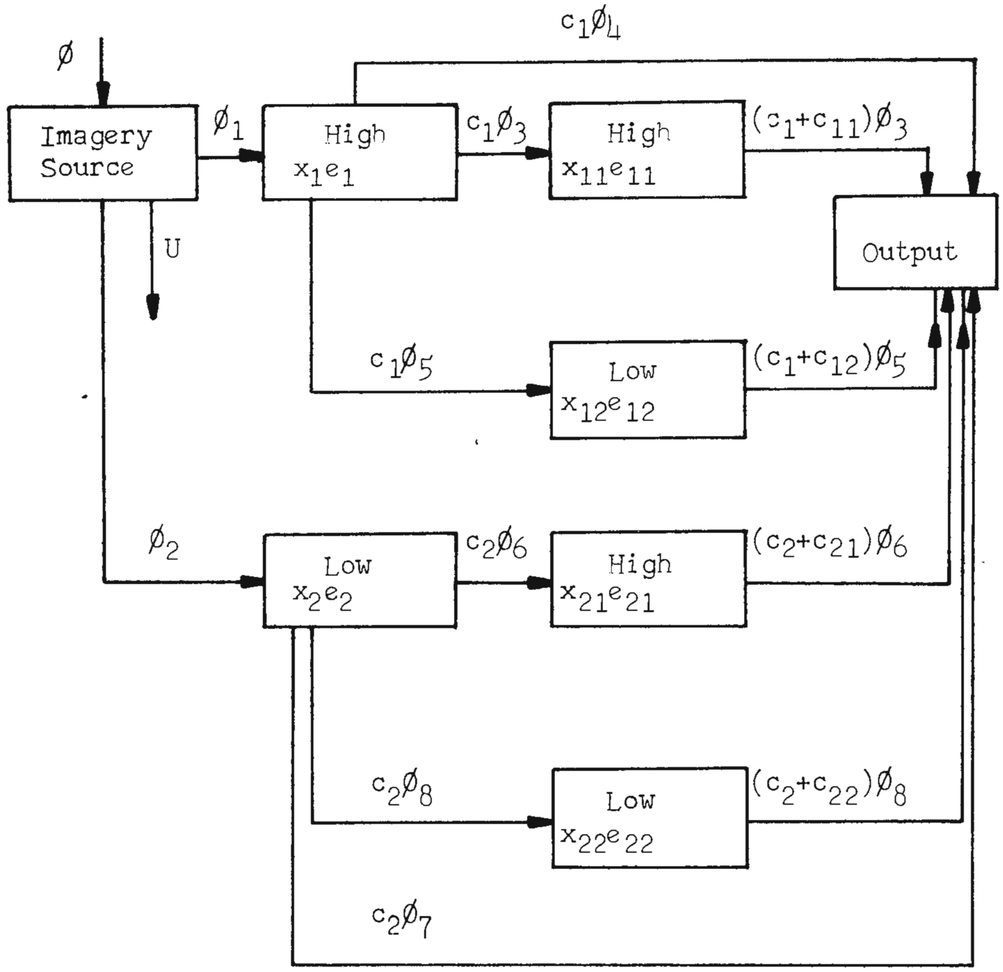


Figure 6. Image Interpretation Flow Chart.

4. ϕ is the flow of interpretable targets into the system;

5. ϕ_i is the flow of interpretable targets in arc i, $i = 1,\ldots,8$;

6. e_i is the efficiency of class i interpreters utilized as initial interpreters, $i = 1,2$;

7. e_{ij} is the efficiency of class j interpreters utilized as checkers of class i initial interpretations, $i = 1,2$, $j = 1,2$;

8. c_i is the expected completeness (accuracy) score of a class i interpreter utilized as an initial interpreter, $i = 1,2$;

9. c_{ij} is the expected completeness (accuracy) score of a class j interpreter utilized as a checker of calss i initial interpretation, $i = 1, 2$, $j = 1,2$; and

10. U is the flow of uninterpreted interpretable targets in excess of system capacity.

This was written in linear program format as:

$$\text{MAX } \phi_3 (c_1 + c_{11}) + \phi_4 c_1 + \phi_5 (c_1 + c_{12}) + \phi_6 (c_2 + c_{21}) + \phi_7 c_2 + \phi_8 (c_2 + c_{22})$$

Subject to

$$\phi = U + \phi_1 + \phi_2$$

$$\phi_1 \leq e_1 x_1$$

$$\phi_2 \leq e_2 x_2$$

$$\phi_1 = \phi_3 + \phi_4 + \phi_5$$

$$\phi_2 = \phi_6 + \phi_7 + \phi_8$$

$\phi_3 \leq e_{11}x_{11}$

$\phi_5 \leq e_{12}x_{12}$

$\phi_6 \leq e_{21}x_{21}$

$\phi_8 \leq e_{22}x_{22}$

$x_1 + x_{11} + x_{21} = X_1$

$x_2 + x_{22} + x_{12} = X_2$

$x_1, x_2, (x_{12} + x_{22}), (x_{21} + x_{11})$ non-negative integer

$\phi_1, \phi_2, \phi_{12}, \phi_{21}, \phi_{11}, \phi_{22} \geq 0$

The $(x_{12} + x_{22}), (x_{21} + x_{11})$ non-negative integer constraint, rather than $x_{11}, x_{12}, x_{21}, x_{22}$ non-negative, was necessary in order to permit the possibility of one checker serving both high and low proficiency initial interpreters.

The problem can be written in terms of the x's only. Adding slack variables to the inequality constraints,

$\phi_1 + S_1 = e_1x_1$

$\phi_2 + S_2 = e_2x_2$

$\phi_3 + S_3 = e_{11}x_{11}$

$\phi_5 + S_5 = e_{12}x_{12}$

$\phi_6 + S_6 = e_{21}x_{21}$

$\phi_8 + S_8 = e_{22}x_{22}$

solving for the ϕ's

$$\phi_1 = e_1x_1 - S_1$$

$$\phi_2 = e_2x_2 - S_2$$

$$\phi_3 = e_{11}x_{11} - S_3$$

$$\phi_5 = e_{12}x_{12} - S_5$$

$$\phi_6 = e_{21}x_{21} - S_6$$

$$\phi_8 = e_{22}x_{22} - S_8$$

noting that

$$\phi_4 = \phi_1 - \phi_3 - \phi_5$$

$$\phi_7 = \phi_2 - \phi_6 - \phi_8$$

The objective function can be written

$$\text{MAX} \quad (e_{11}x_{11} - S_3)(c_1 + c_{11}) + e_1x_1 - S_1 - (e_{11}x_{11} - S_3) - (e_{12}x_{12} - S_5))c_1$$
$$+ (e_{12}x_{12} - S_5)(c_1 + c_{12}) + (e_{21}x_{21} - S_6)(c_2 + c_{21})$$
$$+ (e_2x_2 - S_2 - (e_{21}x_{21} - S_6) - (e_{22}x_{22} - S_8))\, c_2 + (e_{22}x_{22} - S_8)(c_2 + c_{22})$$

which can be reduced to

$$\text{MAX} \quad c_{11}e_{11}x_{11} + c_1e_1x_1 + c_{12}e_{12}x_{12} + c_{21}e_{21}x_{21} + c_2e_2x_2 + c_{22}e_{22}x_2 - c_{11}S_3$$
$$- c_1S_1 - c_{12}S_5 - c_{21}S_6 - c_2S_2 - c_{22}S_8$$

Eliminating the variable ϕ's, the constraints become:

$$e_1x_1 + e_2x_2 - S_1 - S_2 + U = \phi$$

$$x_1 + x_{11} + x_{21} = X_1$$

$$x_2 + x_{22} + x_{12} = X_2$$

$x_1, x_2, (x_{12} + x_{22}), (x_{21} + x_{11})$ non-negative integer

$U \geq 0.$

CHAPTER V

SUMMARY AND CONCLUSIONS

I. SUMMARY

Methods of improving image interpretation system output through use of interpreter proficiency as a criterion for making interpreter personnel assignments were investigated. A Preliminary Experiment was conducted to determine subject proficiency and imagery difficulty. Analysis of variance indicated that the imagery used was sufficiently homogenous that measures of interpreter performance based on interpretation of dissimilar imagery sets could be compared. Experiment I was designed to determine if either of two personnel assignment methods using interpreter proficiency as the assignment criterion would yield significantly improved team performance. Analysis of variance revealed no significant differences in performance due to either of the methods tested. Experiment II was designed to determine if assigning the more difficult imagery to the more proficient interpreter would result in a significantly higher team performance than random assignment of imagery to team members. Analysis of variance indicated no significant differences in interpreter performance due to either of the methods tested.

The image interpreter personnel assignment problem was formulated as a linear integer program.

II. CONCLUSIONS

Insofar as image interpretation operations resemble the experimental conditions of this study, the relative proficiency of image interpreters need not be considered in making personnel task assignments. Subject to the same qualification, pre-sorting of imagery by predicted difficulty of interpretation with subsequent assignment of the more difficult imagery to the more proficient interpreters cannot be expected to result in improved system output.

If measures of expected interpreter proficiency and expected input rate of interpretable targets are known, optimal assignment of interpreter personnel can be made using an integer linear programming formulation of the problem.

SELECTED BIBLIOGRAPHY

SELECTED BIBLIOGRAPHY

1. Aero Service Corporation, "Specification and Maintenance of Interpreter Performance," Rome Air Development Center, RADC-TR-63-539, October 1966.

2. Applied Psychology Corporation, "Performance of Photographic Interpreters as a Function of Time and Image Characteristics," Rome Air Development Center, RADC-TDR-63-313, September 1964.

3. Birnbaum, Abraham H., "Exploratory Study in Interpretation of Vertical and High Oblique Photographs," Army Personnel Research Office, APRO TRN 174, June 1966.

4. Birnbaum, Abraham H., "Human Factors Research in Image Systems--Status Report, 30 June 1962," Army Personnel Research Office, APRO TRN 122, June 1962.

5. Boeing Company, "A Study of Photo Interpreter Performance in Change Discrimination," Rome Air Development Center, RADC-TDR-63-482, 1963.

6. Bolin, Stanley F., Sadacca, Robert, and Martinek, Harold, "Team Procedures in Image Interpretation," Army Personnel Research Office, APRO TRN 164, December 1965.

7. Brainard, Robert W. and Caum, Kenneth B., "Evaluation of an Image Quality Enhancement Technique," Aerospace Medical Research Laboratories, AMRL-TR-65-143, September 1965.

8. Brainard, Robert W., and others, "Development and Evaluation of a Catalog Technique for Measuring Image Quality," Army Personnel Research Office, APRO TRR 1150, August 1966.

9. Colwell, Robert N., "Aids for the Selection and Training of Photo Interpreters," Photogrammetric Engineering, Vol. 31, pp. 326-339, 1965.

10. Colwell, Robert N., "To Measure Is To Know--Or Is It?," Photogrammetric Engineering, Vol. 29, pp. 71-83, 1963.

11. Cornell Aeronautical Laboratory, "Quality Categorization of Aerial Reconnaissance Photography," Rome Air Development Center, RADC-TDR-63-279, 1963.

12. Doten, George W., Cockrell, John T., and Sadacca, Robert, "The Use of Teams in Image Interpretation: Information Exchange, Confidence, and Resolving Disagreements," Army Personnel Research Office, APRO TRR 1151, October 1966.

13. Leibowitz, Herschel W., "The Human Visual System and Image Interpretation," Institute for Defense Analyses, Research Paper P-319, June 1967.

14. Levy, Girard W., and others, "Probability Indexes of Image Interpreter Performance: Development and Evaluation," Behavioral Science Research Laboratory, TRN 183, June 1967.

15. MacLeod, Shelton, "Photointerpreter Performance Studies," Rome Air Development Center, RADC-TDR-64-326, September 1964.

16. Martinek, Harold, and others, "Human Factors Studies in Image Interpretation," Photogrammetric Engineering, Vol. 27, pp. 714-728, December 1961.

17. Minneapolis-Honeywell Regulator Company, "Ground Resolution Study Final Report," Rome Air Development Center, RADC-TDR-63-421, 1963.

18. Radio Corporation of America, "Rapid Identification and Interpretation Techniques," Rome Air Development Center, RADC-TDR-63-421, 1963.

19. Rome Air Development Center, "Proceedings of Symposium on Human Factors Aspects of Photo Interpretation," Rome Air Development Center, RADC-TDR-63-324, September 1963.

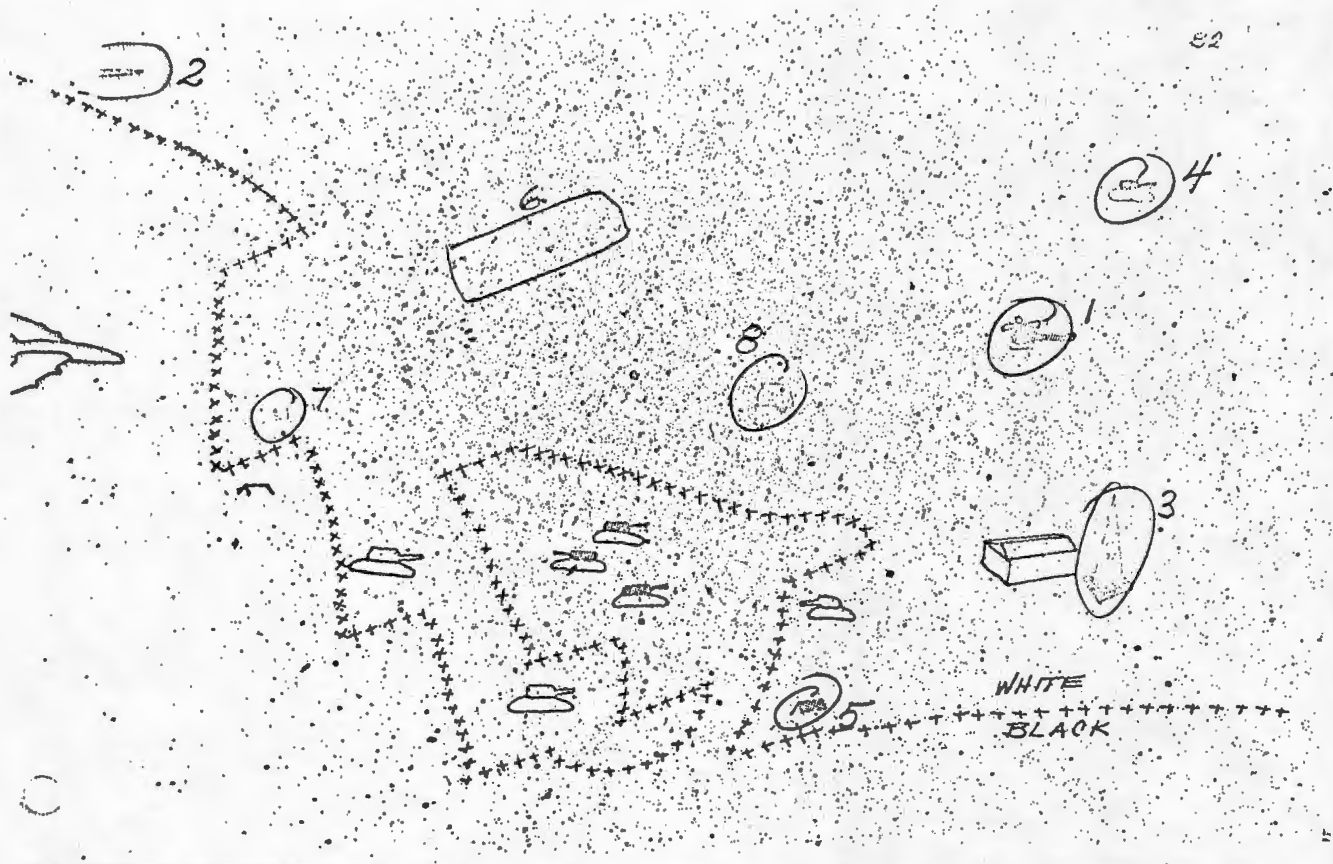
20. Thomas, James A., and Sadacca, Robert, "Ability of Image Interpreters to Adapt Output to Varying Requirements for Completeness and Accuracy," Army Personnel Research Office, APRO TRN 165, December 1965.

21. Thomas, James A., and Sadacca, Robert, "Impact of Feedback on Accuracy of Confidence Levels Assigned by Interpreters," Behavioral Science Research Laboratory, BESRL TRN 187, June 1967.

22. Willmorth, N. E., "Influence of Overlap on Speed and Accuracy in Screening Imagery," Army Personnel Research Office, APRO TRN 180, February 1967.

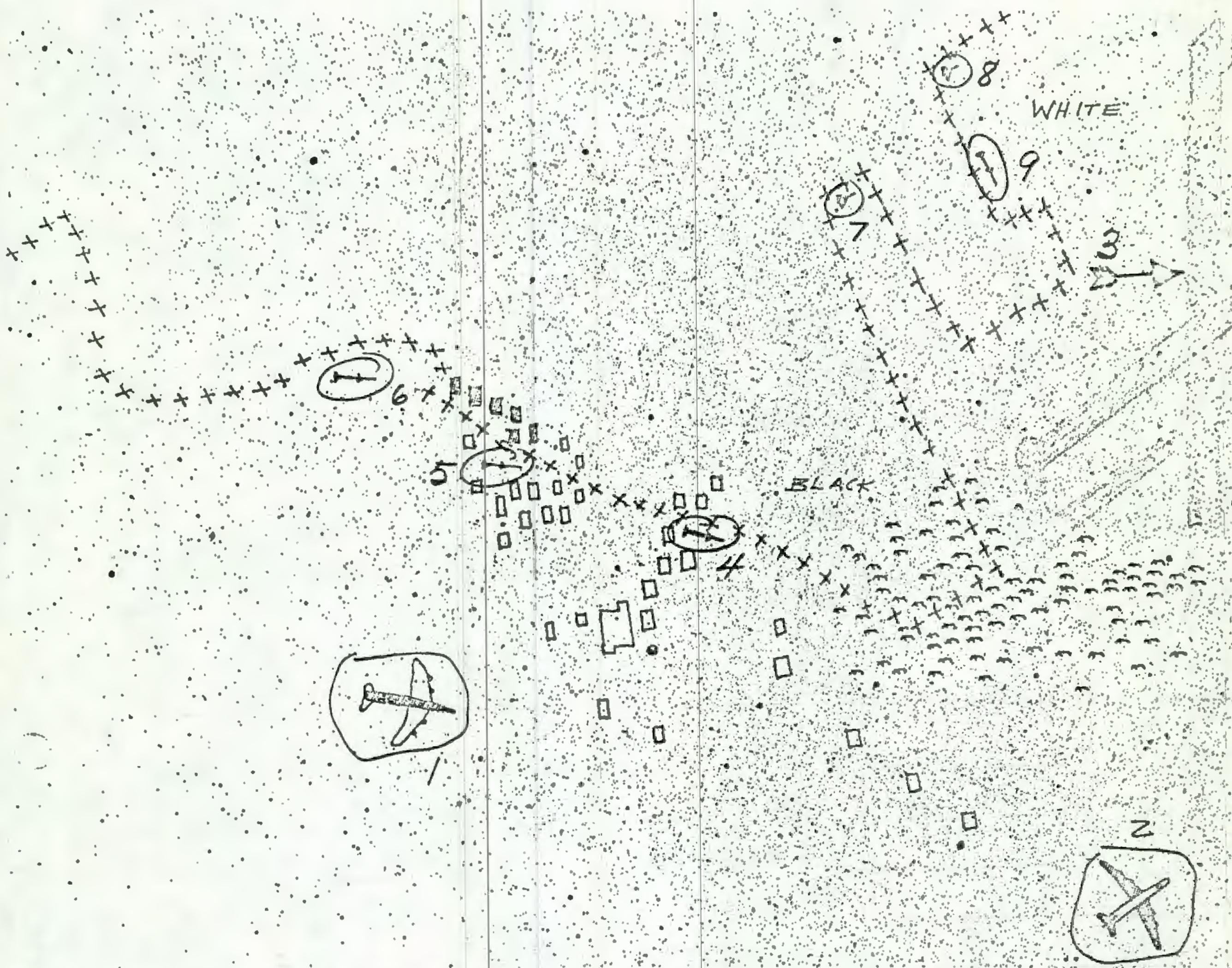
23. Willmorth, Norman E., and Birnbaum, Abraham H., "Influence of Screening and Overlapping Imagery on Speed and Accuracy of Photo Interpretation," Behavioral Science Research Laboratory, BESRL TRN 182, April, 1967.

APPENDIX A



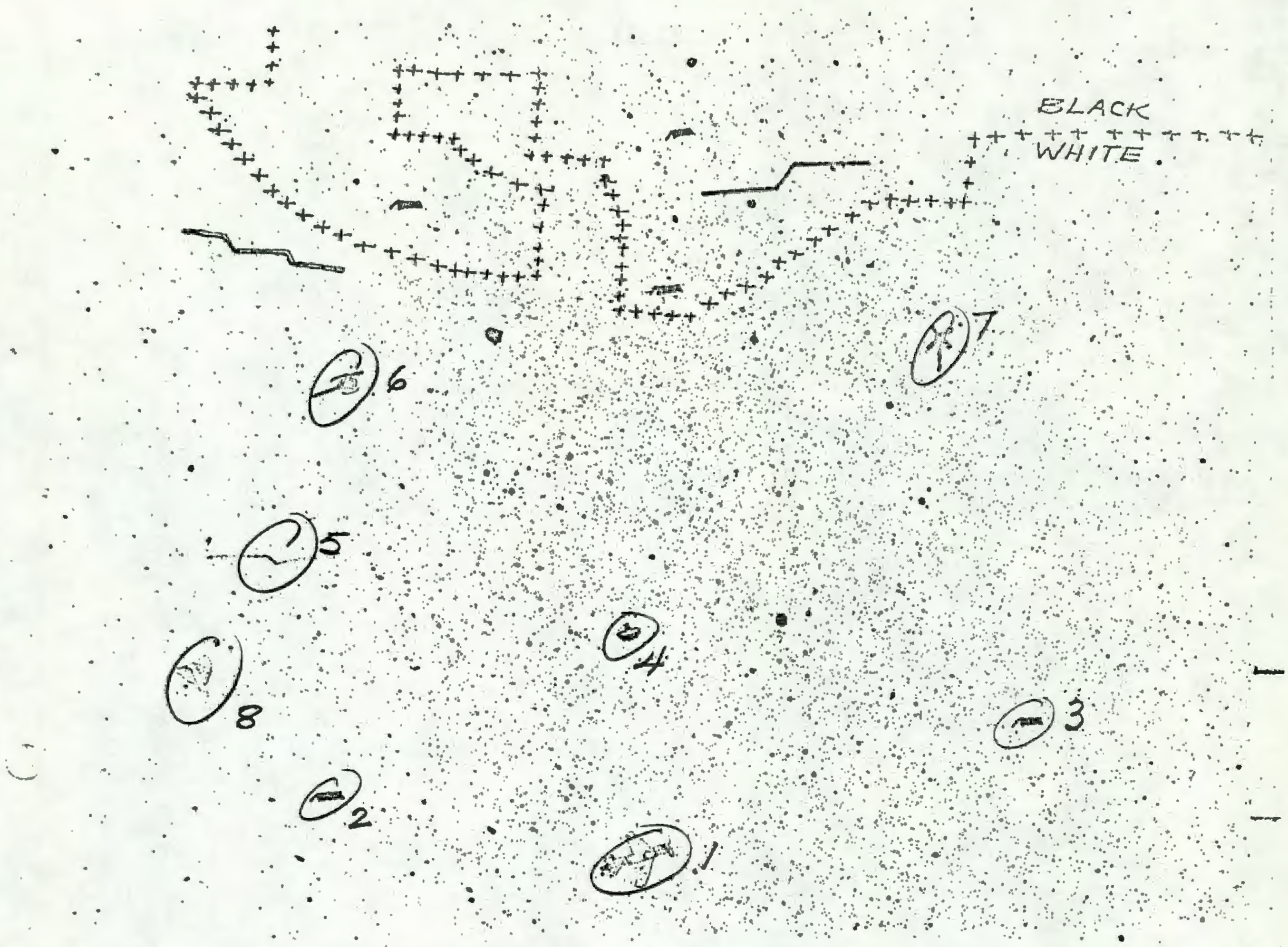
A/C, Jet fighter ______	Helicopter 1	Tank 4
A/C, Prop fighter ______	Missile 2	Tent 5,7
A/C, Multi-engine ______	Radar Antenna ______	Trench ______
Airfield ______	Radio Antenna 3	Truck, long 6
Building ______	Road ______	Truck, short 8

Figure 7. Image Frame B2.



A/C, Jet fighter	7,8	Helicopter		Tank	
A/C, Prop fighter		Missile	4,5,6, 9	Tent	
A/C, Multi-engine	1,2	Radar Antenna		Trench	
Airfield	3	Radio Antenna		Truck, long	
Building		Road		Truck, short	

Figure 8. Image Frame F7.



A/C, Jet fighter 8	Helicopter 6,7	Tank
A/C, Prop fighter 1	Missile	Tent 2,3
A/C, Multiengine	Radar Antenna 4	Trench 5
Airfield	Radio Antenna	Truck, long
Building	Road	Truck, short

Figure 9. Image Frame B3.

APPENDIX B

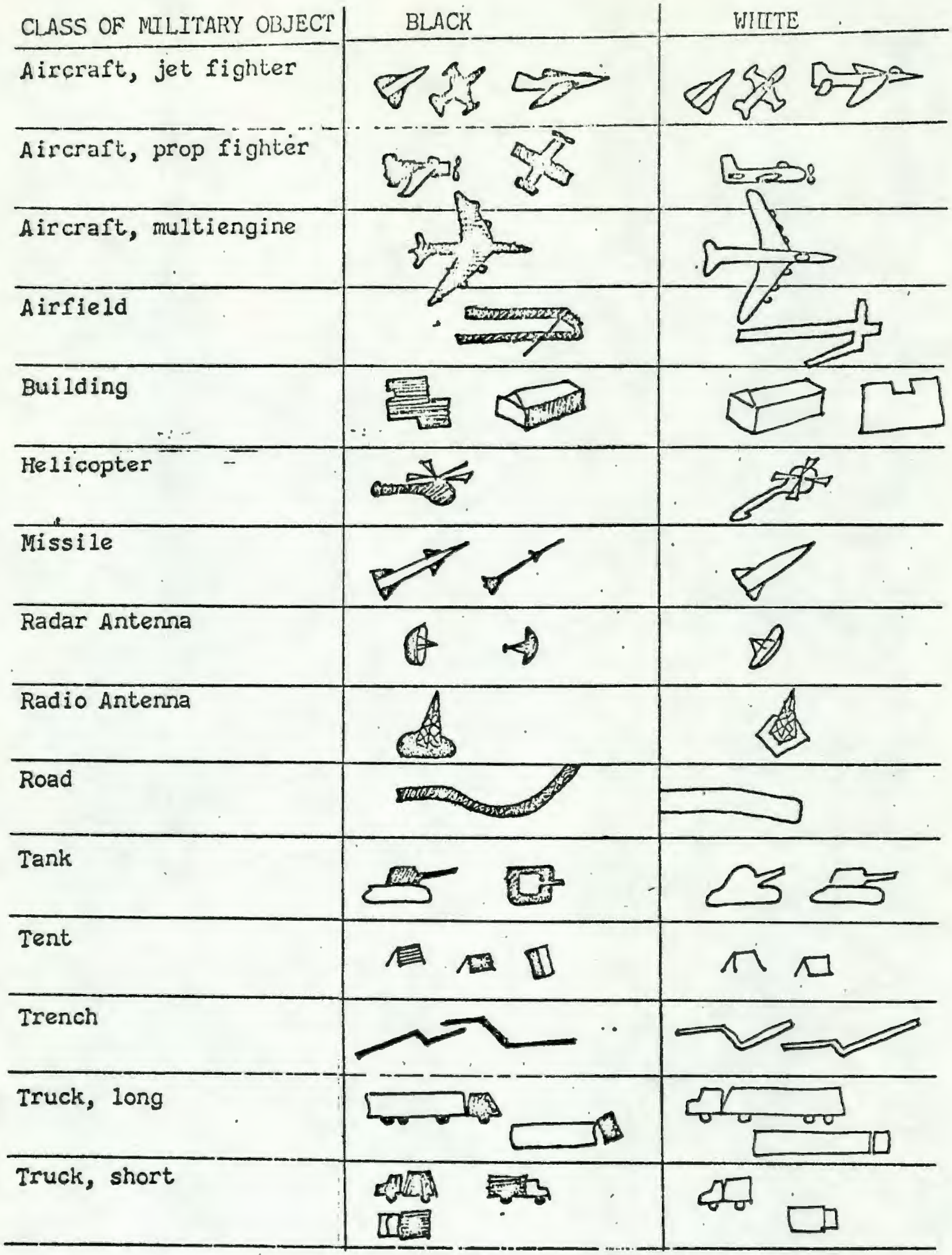


Figure 10. Interpretation Key.

APPENDIX C

APPENDIX C

INSTRUCTIONS TO SUBJECTS

The following instructions were given to subjects prior to the Preliminary Experiment:

"The imagery you will be given is designed to represent aerial photography. The territory depicted shows a border between two countries called WHITE and BLACK. You are WHITE photo interpreters who have been given the task of detecting all BLACK military objects on WHITE's side of the border.

"In the imagery the border is indicated by a line of x's. The WHITE side and the BLACK side are clearly labelled. The border will often be unrealistically irregular in configuration. On the other hand, the military objects will often be unrealistically simple. WHITE and BLACK military objects are similar in appearance; they differ in that BLACK forces are drawn with portions shaded, whereas WHITE forces are drawn in outline with no shaded portions. You are not to report any WHITE forces--no matter which side of the border they are on. You are not to report any BLACK forces on BLACK's side of the border. Report only those BLACK forces that have violated WHITE's territory. Is that clear? (Wait for response.)

"You have an interpretation key before you. It shows examples of the symbols you will see on the imagery. You may refer to the key as you interpret the imagery. The key lists the fifteen different types of military objects you are looking for. For your purposes no other types of military objects exist. The key is arranged in alphabetical order. Note that the types of military objects are:

Aircraft, jet fighter

Aircraft, prop fighter

Aircraft, multiengine

Airfield

Building

Helicopter

Missile

Radar Antenna

Radio Antenna

Road

Tank

Tent

Trench

Truck, long

Truck, short

"Please open your imagery booklet to the first page, labelled EXAMPLE 1. This image frame is similar to those in the rest of the booklet. In this frame BLACK's territory is roughly the upper right quadrant. There are seven military objects:

one BLACK radio antenna

two WHITE tents

two BLACK tanks

one BLACK trench

one BLACK tent

Recall, however, that you are a WHITE interpreter concerned only with reporting BLACK border violations. Therefore only the BLACK tent, tank, and trench are of interest to you, as they are located in WHITE territory.

"Your task will be to look at each image frame, circle or draw an arrow to each BLACK border violator, assign a number to each violator, and write the numbers on the appropriate lines of the evaluation form below the image frame. Please turn to the next page for an example.

"On the imagery the tent has been circled and numbered 1; the number 1 has been entered on the appropriate line. The tank is circled and labelled 2; note the 2 on the line opposite 'tank.' An arrow is drawn to the trench, which is numbered 5; the number 5 appears opposite 'trench' below. Circles around any of the other objects would be scored as errors. Your choice of labelling numbers is immaterial, just as long as each number used is not repeated on the same image frame.

"Please turn the page to EXAMPLE 2. Mark all BLACK forces on WHITE's side, and fill in the evaluation form below. Look up when you have finished. (Pause.)

"Now turn to the next page. The long truck and two helicopters have been labelled and recorded below. Note the double entry on the helicopter line. The BLACK tent is not recorded because it is on BLACK's side of the border. The black shape near the border in the upper part of the frame does not represent any military object. Note that the helicopter to the right contrasts less with the background than does the helicopter to the left. You can expect some objects to be difficult to see on account of low contrast ratio. You may have to guess the identity of an indistinct or lightly drawn object or guess if it is shaded or not. It may be costly to WHITE for you to miss an intruder; on the other hand, false alarms may also be costly. Do as well as you can.

"When I say START, turn to the next page and begin your image interpretation. Work as accurately, completely, and quickly as you can. As soon as you finish one image frame, go immediately to the next. After ten frames you will come to an instruction page. When you reach it, record the time as indicated on the flip cards I have on the desk here in front, and go on to the next set of frames immediately. After completing ten more frames, stop, and record the time from the flip cards. Do not look back at any image frame you have completed. You are then free to leave. You will be scored on accuracy, completeness, and speed--so pace yourselves to maximize your score. Are there any questions? (Pause.) You may start in ten seconds. (Pause.) START."

The following instructions were given to subjects prior to Experiment I:

"You have each been given an imagery booklet that was used by an initial interpreter in yesterday's experiment. It contains marked image frames and evaluation forms. Your task in this experiment is to check-interpret the imagery. You should correct any omissive or commissive errors you find. If you find a commissive error, X through the initial interpreter's marks and mark the frame according to your interpretation. Please do not erase any of the initial interpreter's marks--X through them.

If you find omissive errors--that is, BLACK forces on WHITE's side of the border that were overlooked by the initial interpreter--label them and make appropriate entries on the evaluation form beneath the imagery. (Demonstrate on blackboard.) Is this clear? (Pause.)

"In this experiment you and the initial interpreter are considered a two man team. Your team will be scored on the basis of accuracy, completeness, and speed. Your teammate has, in effect, already done the initial interpretation. You should not change his correct interpretations. Any corrections you make will be final judgments--that is, your teammate will not be checking your corrections.

"When you are told to START, check-interpret the first ten image frames without stopping. Do not go back to a frame you have checked. When you reach the instruction page, record the time you see on the flip card here in front, and go immediately to the next set of image frames. When you finish, stop and record the time. Please remain seated until everyone finishes.

"We'll take a short break when everyone is finished. The final experiment--which is a short one--will follow the break. Are there any questions? (Pause.) You may START in ten seconds. (Pause.) START."

The following instructions were given to subjects prior to Experiment II:

"You have been assigned to teams and should be seated next to your teammate. Some of you have been given sets of ten image frames to interpret. Others of you have a stack of twenty image frames which should be placed within reach of both members of the team.

"When you are told to START, you should interpret your image frames just as you did in yesterday's experiment. If you have your own individual booklet, work through the frames without stopping. When finished, record the time from the flip cards. Those of you with a shared stack of imagery: When told to START, each team member should take one image frame from the top of the stack and interpret it. As soon as you've finished a frame, take another from the top of the stack. Continue working until your team has exhausted the imagery. Each team member should call out his number and record the time when he finishes his last frame. Please remain seated until everyone is finished. Are there any questions? (Pause.) You may START in ten seconds. (Pause.) START."

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Documentation Center Cameron Station Alexandria, Virginia 22314	20
2.	Library Naval Postgraduate School, Monterey, California	2
3.	Director, Systems Analysis Division (OP-96) Office of the Chief of Naval Operations Washington, D. C. 20350	1
4.	Prof. Gary K. Poock (Thesis Advisor) Department of Operations Analysis Naval Postgraduate School, Monterey, California	1
5.	Operations Analysis Department Naval Postgraduate School, Monterey, California	1
6.	LT William L. Schwabe, USNR 7823 Loomis Street Lake Worth, Florida 33460	1

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) Naval Postgraduate School Monterey, California	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE
An Experimental Study of Interpreter Proficiency as a Criterion for Image Interpretation Personnel Assignments.

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
None

5. AUTHOR(S) (First name, middle initial, last name)
William Lawrence Schwabe

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
June 1968	76	23

8a. CONTRACT OR GRANT NO. N/A b. PROJECT NO. c. d.	9a. ORIGINATOR'S REPORT NUMBER(S) N/A 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) N/A

10. DISTRIBUTION STATEMENT
[redacted]

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Naval Postgraduate School

13. ABSTRACT

Methods of improving image interpretation system output through use of interpreter proficiency as a criterion for making interpreter personnel assignments were investigated. An experiment was conducted to determine if either of two personnel assignment methods using interpreter proficiency as the assignment criterion would yield significantly improved team performance. No significant difference in performance due to either of the methods tested were found. A second experiment was conducted to determine if assigning the more difficult imagery to the more proficient interpreter would result in higher team performance than random assignment of imagery to team members. Analysis indicated no significant differences in interpreter performance due to either of the methods tested.

The image interpreter personnel assignment problem was formulated as a linear integer program.

DD FORM 1 NOV 65 1473 (PAGE 1)
S/N 0101-807-6811

UNCLASSIFIED
Security Classification

A-31408

UNCLASSIFIED

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
IMAGE INTERPRETATION						
*IMAGE INTERPRETER TEAM OPERATION						
TEAM METHODS						
HUMAN FACTORS IN IMAGE INTERPRETATION						
PERSONNEL ASSIGNMENT						

DD FORM 1 NOV 65 1473 (BACK)

S/N 0101-807-6821

UNCLASSIFIED

Security Classification

A-31409

[redacted]