Technical Report 72-5

HumRRO-TR 72-5

HumRRO

Studies of Aircraft Recognition Training

Paul G. Whitmore, William C. Rankin, Robert D. Baldwin, and Sandra Garcia

HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street • Alexandria, Virginia 22314

February 1972

Prepared for

Office of the Chief of Research and Development
Department of the Army
Washington, D.C. 20310

54

BIBLIOGRAPHIC DATA SHEET	1. Report No. HumRRO-TR-72-5	2.	3. Recipient's Accession No.
4. Title and Subtitle STUDIES OF AIRCRAFT RECOGNITION TRAINING			5. Report Date Feb 72 6.
7. Author(s) Paul G. Whitmore, William C. Rankin, Robert D. Baldwin, and Sandra Garcia			8. Performing Organization Rept. No. TR 72-5
9. Performing Organization Name and Address Human Resources Research Organization (HumRRO) 300 North Washington Street Alexandria, Virginia 22314			10. Project/Task/Work Unit No. 2Q062107A712 11. Contract/Grant No. DAHC-19-70-C-0012
12. Sponsoring Organization Name and Address Office, Chief of Research and Development Department of the Army Washington, D.C. 22314			13. Type of Report & Period Covered Technical Report 14.

15. Supplementary Notes Work Unit STAR, Aircraft Recognition Training, HumRRO Division No. 5, Fort Bliss, Texas

16. Abstracts

The research dealt with three problem areas: selection of the minimum number of views of each aircraft required for effective recognition training, determination of an appropriate exposure duration for test images, and determination of the relative emphasis needed on friendly and hostile aircraft to produce adequate identification performance. The uniformity of performance on a posttraining test was a function of the number and distribution of the views used in training and the similarity level of the aircraft. Differences in duration from one to five seconds were critical only for the most highly similar aircraft. Both friendly and hostile aircraft need to be given equal training emphasis.

17. Key Words and Document Analysis. 17a. Descriptors

*Military training
*Training devices

17b. Identifiers/Open-Ended Terms

Aircraft recognition
Forward area air defense
Imagery-exposure time

17c. COSATI Field/Group 05 09 Behavioral and social sciences; Personnel selection, training, and evaluation

18. Availability Statement	19. Security Class (This Report)	21. No. of Pages
Approved for public release; distribution unlimited.	UNCLASSIFIED	48
	20. Security Class (This Page) UNCLASSIFIED	22. Price

HumRRO
Technical
Report
72-5

Studies of Aircraft Recognition Training

Paul G. Whitmore, William C. Rankin,
Robert D. Baldwin, and Sandra Garcia

February 1972

HumRRO Division No. 5
Fort Bliss, Texas

HUMAN RESOURCES RESEARCH ORGANIZATION

The Human Resources Research Organization (HumRRO) is a nonprofit corporation established in 1969 to conduct research in the field of training and education. It is a continuation of The George Washington University Human Resources Research Office. HumRRO's general purpose is to improve human performance, particularly in organizational settings, through behavioral and social science research, development, and consultation. HumRRO's mission in work performed under contract with the Department of the Army is to conduct research in the fields of training, motivation and leadership.

Published
February 1972
by
HUMAN RESOURCES RESEARCH ORGANIZATION
300 North Washington Street
Alexandria, Virginia 22314

FOREWORD

The development of a variety of forward area air defense weapons in recent years has revived interest in visual aircraft recognition. This report describes research pertinent to the selection of aircraft views for aircraft recognition training, to the exposure duration of test images used to evaluate aircraft recognition achievement, and to the relative training emphasis upon friendly and hostile aircraft.

These research efforts were conducted by the Human Resources Research Organization under Sub-Unit I of Work Unit STAR. Subsequent Sub-Units are concerned with simulation of field conditions and with the development of an individualized aircraft recognition training program using printed rather than projected images.

The view selection studies described in this report were conceptualized by Dr. Paul G. Whitmore, and were designed and conducted by Mr. William C. Rankin. The exposure duration study was conceptualized and designed by Dr. Whitmore and conducted by Mrs. Sandra Garcia. The studies concerned with the relative training emphasis given to friendly and hostile aircraft were conceptualized by Dr. Robert D. Baldwin, and were designed and conducted by Dr. Whitmore.

STAR research, begun in 1965, is being conducted at HumRRO Division No. 5, Fort Bliss, Texas. Dr. Robert D. Baldwin was Director of Research during the period in which the research described in this report was performed. Dr. Albert L. Kubala is the present Director.

Military support has been provided by the U.S. Army Air Defense Human Research Unit and by the U.S. Army Air Defense Center. The Military Chief of the Human Research Unit at the time these studies were initiated was MAJ A.D. Bell. They were completed during the tenure of LTC J.W. Feiger.

HumRRO research for the Department of the Army is conducted under Contract DAHC 19-70-C-0012. Training, Motivation, and Leadership Research is conducted under Army Project 2Q062107A712.

Meredith P. Crawford
President
Human Resources Research Organization

MILITARY PROBLEMS

This report is concerned with formulating answers to three basic questions regarding the conduct of aircraft recognition training:

(1) What is the smallest number of views of each aircraft that needs to be included in an aircraft recognition training slide kit?

(2) In conducting aircraft recognition testing, what is the most valid duration of each exposure with respect to operational conditions?

(3) Can the amount of training time required to bring observers to an adequate level of friendly or hostile identification performance be reduced by training them to recognize either friendly aircraft or hostile aircraft, but not both?

RESEARCH PROBLEMS

The research effort had three objectives:

(1) To select some minimum number of views so that training on these views would generalize (or transfer) to other views to produce a uniformly high level of recognition performance across all views of operational significance.

(2) To determine the precision required in establishing an operationally valid image exposure duration for aircraft recognition testing.

(3) To determine whether learning to recognize either friendly aircraft or hostile aircraft (but not both friendly and hostile aircraft) will produce a satisfactorily high level of friendly or hostile identification performance; that is, to determine the extent to which an observer can accurately distinguish between "familiar" and "unfamiliar" aircraft.

THE VIEW-GENERALIZATION STUDIES

To select some minimum number of views so that training would transfer to all other views of operational significance, a series of three studies on view-generalization (i.e., transfer studies) were conducted. The first study explored the general effect of systematically varied training views on the pattern of performance on a test; the second and third studies sought to select training views that would produce a uniform high level of performance on the test. In each of these studies, trainees were trained to recognize a given number of selected views of six aircraft used in training plus additional views of the same six aircraft.

It was clearly established that the uniformity of performance on the views in the test is a function of the number and distributions of the views used in training. Performance on those used in training was essentially the same in all three studies. However, it appeared that different sets of training views produced different degrees of generalization to nontraining views. Generalization tended to increase as the number of views used in training increased. However, generalization is not simply a function of the number of views used in training. Use of the three traditional planform views in training did not produce as much generalization to other views as did the use of three particular oblique views.

Generalization appeared to be most restricted around the direct head-on view of the aircraft, but improved as either heading angle or climb angle increased. Training views should be selected to satisfy either of two criteria:

(1) A broad generalization to other views of interest.

(2) Operational criticality despite little or no generalization from other views.

The results of the third study were analyzed for differential effects because of varying degrees of similarity among the aircraft. As would be expected, trainees performed best on the least similar aircraft and worst on the most similar aircraft. This difference did not exist during training, but showed up only in the posttraining test in which less time was available than in the training tests for responding to each image. The results of the subsequent exposure duration study suggested that the poorer performance on the highly similar aircraft was due, at least in part, to the restricted total amount of time available to respond to each image. Highly similar aircraft may not be so much more difficult to learn to recognize, but the act of recognizing them may require more time than the act of recognizing less similar aircraft.

THE EXPOSURE DURATION STUDY

To determine the precision required in establishing an operationally valid image exposure duration for aircraft recognition testing, two classes of trainees, who had been trained to recognize six aircraft, were separated into thirds. Each third was administered the same posttraining test, but the images were exposed for different durations. Each image was exposed to one group for one second, to another group for three seconds, and to the last group for five seconds. All groups were given a five-second blank between images to write the answer. The six aircraft again were selected to represent low, moderate, and high levels of similarity.

Differences in performance on the posttraining test were insignificant for different exposure durations of the low- and moderate-similarity aircraft. Overall performance on the highly similar aircraft was lower. Furthermore, performance on the highly similar aircraft exposed for only one second was poorer than performance on the same aircraft exposed for three or five seconds.

The degree of similarity represented by the two high-similarity aircraft is relatively uncommon among the aircraft of the world. This high degree of similarity would most likely occur among aircraft produced in the same country. Two aircraft with this high degree of similarity are likely to be both friendly or both hostile, rather than one being friendly and the other hostile.

STUDIES OF DIFFERENTIAL REPRESENTATION OF FRIENDLY AND HOSTILE AIRCRAFT IN TRAINING

Two studies were conducted to determine whether observers can accurately distinguish between familiar and unfamiliar aircraft. In the first study, one group of trainees was trained to recognize six U.S. aircraft and another group was trained to recognize six non-U.S. aircraft. Both groups were administered a posttraining test that included all 12 aircraft. The trainees were required to identify each test image as either "friendly" or

"hostile." Trainees in both groups performed significantly lower on the unfamiliar aircraft (71.5%) than on the familiar aircraft (86.9%).

Previous research (HumRRO Technical Report 68-1, January 1968) had already established that a two-category approach in which students are required to learn to differentiate equally among all aircraft, in both the friendly and hostile categories, is effective. The study described above established that single-category approach is not acceptably effective.

After the first study was completed, it was hypothesized that the effectiveness of the single category approach might be increased, or bolstered, by providing paired-comparisons between similar U.S. and non-U.S. aircraft during training. Three training conditions were evaluated:

(1) One class received 42 different paired comparisons between U.S. and non-U.S. aircraft repeatedly during training. They were told the designation of the U.S. aircraft in each pair. However, they were told only that the non-U.S. aircraft was hostile.

(2) A second class received the same treatment, except that they were told the designation of the non-U.S. aircraft in each pair.

(3) A third class received paired-comparison training involving only U.S. aircraft.

All three groups were tested during training on their recognition accuracy of the U.S. aircraft only. They were administered a posttraining test in which they were instructed to recognize each U.S. aircraft image by name or number designation and to identify each non-U.S. (or unfamiliar) aircraft as hostile. There were no effective differences in performance on the posttraining test between the three groups. Performance on the U.S. aircraft ranged from 86.2-89.7%. Performance on the non-U.S. aircraft ranged from 47.1-50.7%.

These two studies clearly indicate that a single category approach to aircraft recognition training, whether bolstered or unbolstered, does not provide an acceptable level of identification accuracy for aircraft in the nonincluded category. Both friendly and hostile aircraft should receive equal emphasis during training.

CONCLUSIONS

(1) The views used in training should be systematically selected to provide for uniformly high recognition performance across all views of operational significance.

(2) The exposure duration of test images is not critical except for those instances in which the trainee is required to discriminate between highly similar aircraft.

(3) All aircraft that the observer is expected to identify should receive equal emphasis in training and testing.

(4) Learning to recognize aircraft occurs in a relative rather than in an absolute sense. One learns to recognize a single aircraft in a set of similar aircraft, rather than by simply learning to name each single aircraft independently of the others in the set. Images of similar aircraft should be presented in an intermixed random order in practice and in testing.

CONTENTS

Studies of
Aircraft Recognition Training

Chapter 1

INTRODUCTION

BACKGROUND

A "state-of-the-art" method of administering aircraft recognition training in the classroom was developed by the Human Resources Research Organization in 1966. The training method, which was described in a HumRRO Technical Report,[1] employed teaching methods and training aids that were different from those currently prescribed by military training literature. Subsequent to that training experiment and the report of its results, the HumRRO research team conducted behavioral and operational analyses to identify the characteristics of effective visual aids to support aircraft recognition instruction.

Number and Distribution of Training Views

Analysis indicated that an effective and useful visual aid kit for aircraft recognition instruction would have the following:

(1) A large number of views that represent all views critical in operational situations.

- (a) Aircraft image size equal for all aircraft.
- (b) Image size small enough to represent aircraft at some distance from an observer.
- (c) Slide backgrounds of uniform tone, lighter than the image and providing moderate to low contrast.
- (d) Slide images uniformly illuminated and monochromatic with a minimum of highlighting. All main features visible, (i.e., intakes, exhausts, and canopies) but no nationality markings.

(2) Comparable views available for each aircraft in the kit to facilitate direct comparison of views of different aircraft.

(3) One set of aircraft views reserved exclusively for proficiency testing to obtain a valid measure of the utility of the training for transferring to new situations.

(4) A large number of duplicate slides for each image to permit slide trays to be assembled for the entire training program at one time.

(5) Single image slides presented as stimulus-feedback pairs for recognition practice and review. These slides would contain the same image as the stimulus slide in each pair, but would also give the aircraft's name or number designation.

(6) Selected pairs of slides showing the same view of two different aircraft for paired-comparisons training.

To select and specify the views for the kit, it was necessary to devise a system for describing aircraft views with respect to an observer. There are three possible characteristics of aircraft views for this purpose: (a) the heading angle, (b) the climb angle, and (c) the roll angle.

[1] Paul G. Whitmore, John A. Cox, and Don J. Friel. *A Classroom Method of Training Aircraft Recognition*, HumRRO Technical Report 68-1, January 1968.

To simplify the system, the roll angle requirement was dropped because it is largely redundant; most views of an aircraft specified in terms of a heading and climb angle and a non-zero roll angle can be adequately approximated by a view of the aircraft at a zero roll angle, but at some other heading and/or climb angle.

Heading angle was specified as shown in Figure 1. If the aircraft is heading directly toward the observer, it is designated as having a 0° heading. If it is heading in a direction perpendicular to the observer's line of sight and to the observer's left, it is designated as having a 270° heading; if it is heading in a direction perpendicular to the observer's line of sight but to his right, it is designated as having a 90° heading. If it is moving directly away from the observer, it is designated as having a 180° heading. It should be noted that views on one side of the 0—180° axis are mirror images of comparable views on the other side of the 0—180° axis.

Schema for Defining Aircraft Heading Angle

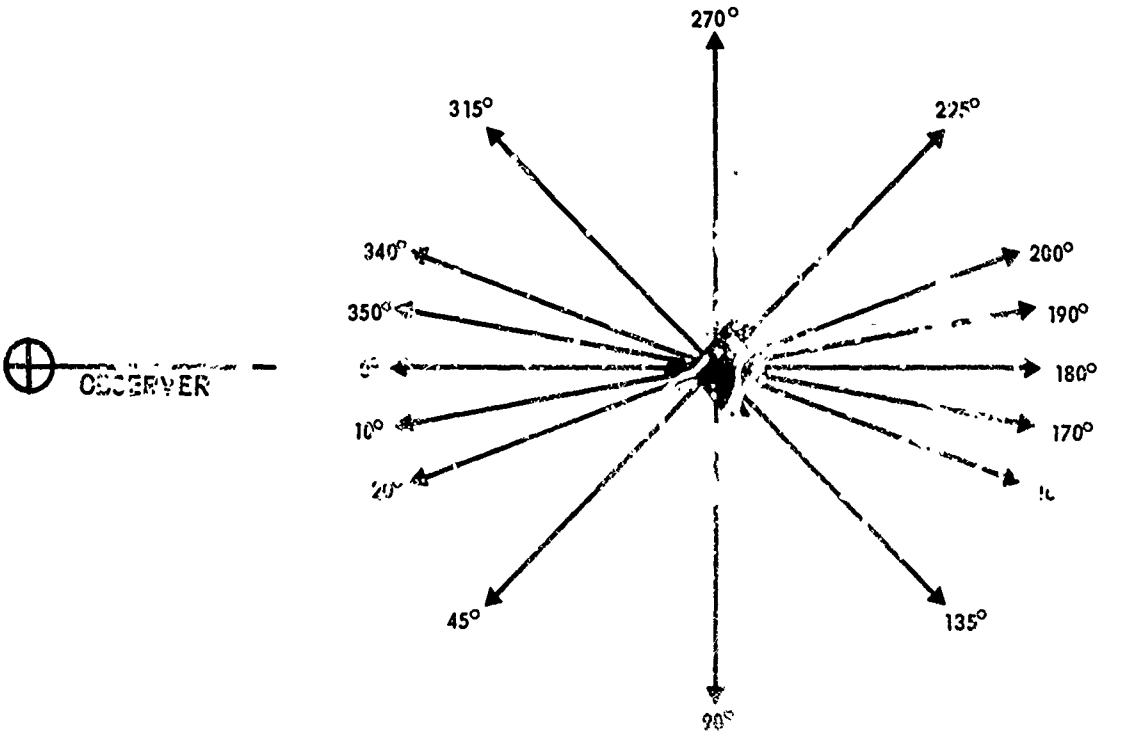


Figure 1

Climb angle was specified as shown in Figure 2. Climb angle is the angle between the aircraft's direction of movement and the horizontal plane containing the observer's line of sight. If the aircraft's direction of movement lies in the observer's sight plane, it has a 0° climb. If it is crossing perpendicularly to the observer's sight plane so that its nose is straight up, it has a 90° climb angle.

Views for the slide kit were selected to be representative of all views that might be critical in the operational situation. Since low-flying aircraft were the major problem area, climb angles were sampled most densely at the lower values. In addition, it was believed that generalization to adjacent views would be least for views at the lower climb angles. Heading angles were sampled most densely around the 0—180° axis for the same reason.

Schema for Defining Aircraft Climb Angle

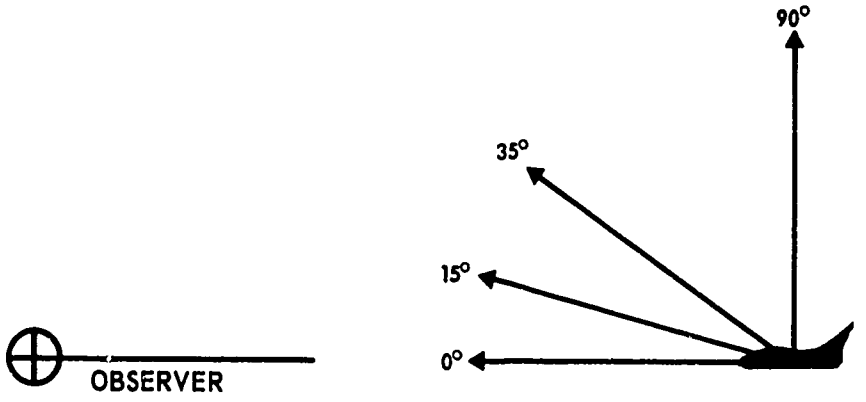


Figure 2

Twenty-four approaching views (Figure 3), and 21 receding views (Figure 4) were selected, for a total of 45 views of each aircraft. Views at 90° and 270° headings were indicated for only the 0° climb angle, since increasing the climb angle at these headings serves only to rotate the image without changing its configuration. The 45 views are shown in Figure 5.

Approaching Views

Climb Angle (degree)	270	315	340	350	0	10	20	45	90
90					X				
35		X	X	X	X	X	X	X	
15		X	X	X	X	X	X	X	
0	X	X	X	X	X	X	X	X	X

Heading Angle (degree)

Figure 3

Receding Views

Climb Angle (degree)	90	135	160	170	180	190	200	225	270
35		X	X	X	X	X	X	X	
15		X	X	X	X	X	X	X	
0		X	X	X	X	X	X	X	

Heading Angle (degree)

Figure 4

Sample Illustration of the GOAR Kit Images for One Aircraft, Soviet MIG-19, Farmer

0°–35°

10°–35°
(350°–35°)

20°–35°
(340°–35°)

45°–35°
(315°–35°)

180°–35°

0°–15°

10°–15°
(350°–15°)

20°–15°
(340°–15°)

45°–15°
(315°–15°)

180°–15°

0°–0°

10°–0°
(350°–0°)

20°–0°
(340°–0°)

45°–0°
(315°–0°)

180°–0°

Figure 5 *(Continued)*

Sample Illustration of the GOAR Kit Images for One Aircraft, Soviet MIG-19, Farmer *(Continued)*

190°–35°
(170°–35°)

200°–35°
(160°–35°)

225°–35°
(135°–35°)

0°–90°

190°–15°
(170°–15°)

200°–15°
(160°–15°)

225°–15°
(135°–15°)

90°–0°
(270°–0°)

190°–0°
(170°–0°)

200°–0°
(160°–0°)

225°–0°
(135°–0°)

NOTE: In each pair, the first value is the heading angle and the second is the climb angle. Values in parens denote a mirror image.

Figure 5

A prototype aircraft recognition slide kit that had the desired image and background characteristics, as well as the other requirements already discussed, was developed. It was designated as the Prototype GOAR (Ground Observer Aircraft Recognition) Slide Kit.

In addition, an instruction manual "A Manual for Conducting Aircraft Recognition Training in the Classroom," was prepared specifically for training to be used in conjunction with the Prototype GOAR Slide Kit. The development of a slide kit and the troop test of the kit and the manual are described in Appendix A.

Descriptions of the Prototype GOAR Slide Kit, and the kit itself, were submitted to the U.S. Army Air Defense School for review. As a result of this review, the Air Defense School prepared a draft small-device requirement that recommended development of a replacement for the current SLARK #1 slide kit. The recommended replacement contained a number of the characteristics of the Prototype GOAR Slide Kit. However, since the GOAR Slide Kit contained 45 views of each aircraft, the Air Defense School requested that HumRRo conduct studies to reduce the number of views required for instructional purposes. This report describes the results of a number of experiments that were designed to determine the amount of transfer of training that occurs when limited numbers of views are employed for aircraft recognition training.

The training method developed earlier by HumRRO had been informally questioned in two other respects. The first question concerned the duration of the exposure of training and test images during instruction. The second question concerned the need to provide aircraft recognition training for both friendly and hostile (or [illegible]tile) aircraft; some critics felt that training would be most efficient if presented only friendly or only hostile aircraft, not both.

Duration of Test Exposures

The training methods used in World War II for aircraft recognition customarily used very short exposures (less than one second), particularly during the later sessions of aircraft recognition instruction. Very brief exposures had been originally recommended by Renshaw to prevent trainees from analyzing the aircraft into their component parts.[1] Renshaw believed that analysis of the image interfered with recognition learning, although no experimental data existed to support his hypothesis.

After World War II, Gibson reported upon research that provided data of relevance to the issue of short- versus long-image exposure.[2] The results of this research have been discussed in an earlier HumRRO Technical Report.[3] Within the range of conditions studied by Gibson, test performance was found to be independent of the duration of image exposures used during training, except for a condition in which testing and training had exposure durations of 1/50 of a second. Test performance on 1/50-second exposures was significantly better for students who had exposures at 1/50 second than for students who had longer training exposures. Gibson's research indicated that test performance could be a function of the duration of the test exposure. Within the conditions of that study, longer test exposures produced higher test scores.

In the context of the HumRRO-designed training method, it was concluded that additional studies were needed to identify the optimum test exposure interval to be used for evaluating training achievement. An examination of possible air defense situations

[1] Samuel Renshaw. "The Visual Perception and Reproduction of Forms by Tachistoscopic Methods," *Journal of Psychology*, vol. 20 1945, pp. 217-232.

[2] James J. Gibson (Ed.). *Motion Picture Testing and Research*, Armed Forces Aviation Psychology Program Research Reports, Report No. 7, U.S. Government Printing Office, Washington, D.C., 1947.

[3] Elmo E. Miller and Arthur C. Vicory. *Comparison and Evaluation of Printed Programs for Aircraft Recognition*, HumRRO Technical Report 71-22, October 1971.

involving aircraft recognition suggested that recognition judgments might have to occur during intervals varying between one and five seconds. This range of exposure duration was, therefore, selected for experimental evaluation.

Representation of Friendly and Hostile Aircraft

The second question concerned the inclusion of both friendly and hostile aircraft in aircraft recognition training programs. It has been customary in the past to train aircraft observers to recognize all aircraft that they might reasonably be expected to encounter in tactical situations. Some military planners have suggested, however, that observers might be trained to recognize only hostile aircraft or only friendly aircraft, but not both. This would reduce the amount of training time required and the memory burden imposed upon the observer. Under this rationale for instruction, if an observer had been taught to recognize only hostile aircraft, he would identify an aircraft as friendly if he did not recognize it; conversely, if taught to recognize only friendly aircraft, he would conclude that any aircraft he did not recognize was hostile.

In selecting training procedures, it makes no difference whether observers are trained to recognize only friendly aircraft or only hostile aircraft. However, in the development of training materials and the specification of time required for training, the set of aircraft to be used does make a difference. Although information is not always available on aircraft of potential enemies, such information is available for our own aircraft and those of our allies. In this report, two experiments are described that were designed to evaluate this concept of recognition and training.

RESEARCH OBJECTIVES

Chapter 2 describes a series of studies designed to select the minimum number of views from the Prototype GOAR Slide Kit to produce the greatest amount of generalization or transfer to all views in the kit. The results of the study that evaluated the effect of various durations of image exposure of aircraft slides on recognition test proficiency are given in Chapter 3. Chapter 4 is concerned with results of studies designed to determine whether limiting instruction in aircraft recognition to either friendly or hostile aircraft would produce a satisfactorily high level of accuracy in identification.

TRAINEES

Most of the trainees used in each experiment were either draftees or volunteers in their first enlistment—primarily, young men in their early twenties. All groups of trainees were members of air defense or automatic weapons units; for example, Redeye AIT,[1] or quad-fifty and twin-forth batteries. Mean GT[2] scores varied moderately from group to group; however, the range in every group was quite large—from the low 80s to over 130.

Trainees were obtained by requesting a certain number for a given day through regular post channels. Sometimes all trainees in a group came from the same unit and sometimes they came from several different units. The only stipulation placed on the request was that each man be free of visual anomalies.

[1] Advanced Individual Training (AIT).
[2] General Technical Aptitude Area tests (GT).

Chapter 2

THE VIEW-GENERALIZATION STUDIES

METHOD

General

The general procedure in this series of three studies consisted of training a group of men to recognize a given number of selected views of six aircraft and then testing them, immediately after training, on their ability to recognize the views of the aircraft used in training, plus additional views of the same six aircraft.

The first study explored the general effect of systematically varied training views on the performance pattern in the test. The second and third studies sought to select training views that would produce a uniform high level of performance on the test. The specific conditions used in each study are summarized in Table 1.

Table 1

Summary of Conditions Used in Each of the View-Generalization Studies

Study	Number of Training Views of Each Aircraft	Aircraft	Number of Trainees	Test View
Study I				All approaching views, including mirror images.
Experiment 1	1	Set A	13	
Experiment 2	1	Set A	11	
Experiment 3	3 (planform)	Set A	13	
Experiment 4	3 (oblique)	Set A	13	
Study II				All approaching views, including mirror images.
Group 1	6	Set B	10	
Group 2	6	Set B	10	
Study III				All approaching and receding views, excluding mirror images.
Group 1	9	Set A	20	
Group 2	9	Set A	20	

Groups

The first study consisted of four concurrent experiments. Trainees were randomly assigned to each of the four experiments, so the results of the four experiments are

directly comparable. Different views of the aircraft were used in training these four experimental groups.

The second study was conducted at a later time with trainees drawn from a different source. The same training treatment was administered to both groups in this study. Subsequently, the third study was conducted, drawing trainees from yet a different source. Again, the same training treatment was administered to all trainees in the study.

Materials

The first and third studies used the same six aircraft. These aircraft were selected from those used in the similarity scaling study described in Appendix A so as to represent three levels of similarity—high, moderate, and low. These similarity levels were designated as Set A:

High Similarity (HS)
- Fishbed (Mig-21)
- Fishpot

Moderate Similarity (MS)
- F-4 (Phantom)
- A-4 (Skyhawk)

Low Similarity (LS)
- F-5 (Freedom Fighter)
- Flashlight (Yak-25)

These similarity levels represent net similarity among all six aircraft rather than simply the degree of similarity between the two aircraft at each level. The low-similarity aircraft are not only dissimilar from each other, but also dissimilar from the aircraft in the other two levels. The moderate-similarity aircraft are not only similar to each other, but also similar to the high-similarity aircraft.

The second study used Set B, which had three of the same aircraft as Set A, plus an additional three:

- Fishpot
- F-4 (Phantom)
- Flashlight (Yak-25)
- F-8 (Crusader)
- F-100 (Super Sabre)
- F-101 (Voodoo)

All these aircraft cluster toward the moderate-to-low end of the similarity scale with respect to each other. The first and second experiments in the first study used only one view in training. The third experiment used the three traditional planform views (head-on, full-belly, and full-crossing) in training. The fourth experiment also used three views in training, but three oblique views rather than three planform views.

The second study consisted of two replications of one experiment,[1] which used the same three oblique views used in the fourth experiment of the first study plus two of the planform views used in the third experiment. The two planform views were selected to bolster low points in the generalization gradients (i.e., performance patterns) resulting from the last experiment in the first study.

The third study, like the second, consisted of only one experiment, which built upon the results of the second study by adding four more training views to bolster low points in the generalization gradients resulting from the second study.

[1] The original intent of this study had been to train one group of 20 trainees for two days. However, this requirement could not be met. Instead, it was administratively necessary to train two groups of 10 trainees for one day each.

The end-of-training tests used in the first and second studies included only approaching views of the aircraft (Figure 3). The test used in the third study included both the approaching and receding views (Figures 3 and 4).

Procedure

Training in each study proceeded in successive 50-minute sessions with 10- to 15-minute breaks between sessions. An achievement test was administered toward the end of each 50-minute session. Each image was exposed for five seconds with a five-second blank between images to allow the trainees to write their recognition responses on their answer sheets.

Training in the first and second studies was conducted as follows:

(1) Orientation and introduction to the task (5-10 minutes).

(2) Learning of the names of the six aircraft (5-10 minutes).

(3) First classroom hour—informal recognition feature learning for each aircraft using slides paced by a military instructor.

(4) Succeeding hours of instruction—five-second projector pacing of stimulus-response-feedback practice, during which trainees practiced with paper and pencil as a group, or, in some cases, individually, by responding orally.

(5) When the group average reached 90% on the periodic achievement tests, the instructor paced the slide exposures at approximately 0.5 second, and the trainees responded orally, individually, or orally as a group.

(6) Training continued for each group until all possible trainees attained an individual achievement level of 80%. One or more trainees in each group progressed so slowly as to result in their being dropped from the study.

Training in the third study was conducted as follows:

(1) Orientation and introduction to the task (10 minutes).

(2) Learning of the names and recognition features of each aircraft, using a specially prepared booklet containing one page for each aircraft with images of the three planform views, the 340° heading - 15° climb view of the aircraft, its name, and a listing of its recognition features (20-45 minutes).

(3) Succeeding hours of instruction—instructor-paced slide exposures ranging from approximately 10 seconds to 0.5 second. Trainees responded orally, in turn, or as a group. The instructor provided oral feedback, which frequently included information regarding the recognition features of the displayed aircraft image.

(4) Trainees were released from training individually at the end of the session in which they attained 90% on a periodic achievement test.

All trainees in each group of each study were assembled immediately following the conclusion of training and administered an end-of-training test (ETT). In the first study, the ETT consisted of all the approaching views with the exception of the full-belly (0° heading - 90° climb) view (Figure 4). These 23 views were shown once for each of the six aircraft—138 images. The second study used these same views plus the full-belly view—144 images. The ETT for the third study was extended to include receding views also. However, to keep this ETT from being too long, mirror images were omitted (with one exception). The full-belly view was included, and two views were added at heading angles not previously represented—65° (15° climb) and 295° (35° climb). In addition, half the aircraft were represented at 0° heading at each of two new climb angles—7.5 and 25°. Thus, a total of 30 different views were shown for each of the six aircraft.

All images in each of the ETTs were presented in random order. Each image was exposed for five seconds. There were no blanks between images. Trainees wrote their responses on their answer sheets while the image was still exposed on the screen. The ETT for the first and second studies required 12 minutes for exposure of all the images. The ETT for the third study required 30 minutes for exposure of all the images.

RESULTS

Training

Experimental training times were as follows:

Study	Number of 50-Minute Sessions[1]
Study I	
Experiment 1	3.2
Experiment 2	2.2
Experiment 3	5.2
Experiment 4	2.2
Study II	
Group I	2.2
Group II	3.2
Study III	
Mean	3.5
Standard Deviation	0.8

All trainees in each group in the first and second studies received the same amount of training. In the third study, trainees were released from training individually at the end of the session in which they achieved 90%. Consequently, different trainees in the same group in the third study received different amount of training.

Not all trainees met the achievement criterion in each study. The rate of progress of some was so slow that training could not reasonably be continued in the expectation that they would attain the achievement criterion. Many of those who met the achievement criterion in the first and second studies tended to be overtrained since all trainees were trained for the same length of time in each of these studies. In two instances, the mean group achievement was higher at the end of the next-to-the-last session than at the end of the last session. The higher figure is considered the more valid indicator of the group's achievement. It seems reasonable to assume that the decrement from the next-to-the-last to the last session was due to fatigue and boredom rather than to assume that the higher test performance is an overestimate of actual achievement. The number of trainees who met the achievement criterion and the mean achievement of these trainees in each study are shown in Table 2.

Even though the third study used a different achievement criterion than was used in the first and second studies, the differences among the achievement means are negligible. Fewer trainees met the achievement criterion in the third study than in the first and second studies, because the third study used a 90% rather than an 80% individual criterion, in order to be consistent with the newly adopted achievement criterion for Redeye gunners. Although only 20 out of 40 met the 90% achievement criterion used in this study, 35 out of 40 met the lower 80% level used in the previous studies. This is comparable to the proportions of trainees who attained the 80% level in the previous studies, as shown in Table 2.

[1]The last session in Studies I and II lasted 10 minutes instead of 50 minutes, because it was only necessary to improve the performance of a few trainees by a few percentage points. Other trainees who had already attained the minimum achievement criterion had become quite restless, so the session was reduced to the shortest possible time required for most of the remaining trainees to reach the criterion.

Table 2

Number of Trainees Meeting Achievement Criterion

Study Group	Criterion	Number of Trainees		Highest Mean Achievement of Those Meeting Criterion (%)
		Achieving Criterion	Total	
Study I	80% (Group)			
Experiment 1		11	13	96.2[a]
Experiment 2		10	11	100.0
Experiment 3		11	13	95.5
Experiment 4		11	13	99.5[b]
Study II	80% (Group)	16	20	96.3
Study III	90% (Individual)	20	40	98.0

[a]The mean achievement for the next-to-the-last training session was used. The drop from that session to the last was 0.6%.

[b]The mean achievement for the next-to-the-last training session was used. The drop from that session to the last was 3.8%.

End-of-Training Test

The number of training views and the number of nontraining views in the ETT used in each study and the mean percent correct for both categories are shown in Table 3. The number of nontraining approaching views is low in the third study, not only because more views were used in training but because mirror images were not used in this ETT.

Table 3

Performance on End-of-Training Test (ETT) on Views Used and Not Used in Training

Study Group	Number of Training Views	Mean Percent Correct in ETT on Views Used in Training	Number of Nontraining Views Used in ETT	Mean Percent Correct in ETT on Views Not Used in Training	Decrement in ETT Performance on Views Not Used in Training
Study I					
Experiment 1	1	78.8	22	55.2	−23.6
Experiment 2	1	85.0	22	53.7	−31.3
Experiment 3	3	81.8	20	54.7	−27.1
Experiment 4	3	90.9	20	74.9	−16.0
Study II	5	89.2	19	80.8	− 8.4
Study III	9	85.0	7	79.9	− 5.1

The third study also included receding views in the ETT. Since the previous studies did not include receding views, only the nontraining approaching views were used in computing the mean percent correct for the third study as shown in Table 3.

Because a different achievement criterion was used in the third study than in the first and second studies, it cannot be statistically compared to the earlier studies.

The differences among the groups with regard to percent correct on nontraining views in the ETTs of the first and second studies were evaluated by means of a single-factor analysis of variance. The treatment effect is statistically significant ($F = 1.38$, df 4,54, $p < .05$). The differences among the first five means on nontraining views in Table 3 were evaluated by means of the Newman-Keuls procedure, which showed that the treatment means for the first three experiments do not differ significantly from one another and that the means from the fourth experiment and Study II do not differ significantly from each other, but do differ significantly ($p < .05$) from the first three means.

The differences among the groups with regard to percent correct on the training views in the ETTs of the first and second studies were also evaluated by means of an analysis of variance. The treatment effect is not statistically significant ($F = 1.38$, $df = 4,54$, NS). These means are also shown in Table 3.

Figures 6 through 9 display the performance gradients by view on the ETTs administered at the end of each study; in each case, the views that had been used during training are listed, to permit comparison of test performance on training and nontraining views. The points along each line show percent correct for each view in the ETT.

The graphs for all but the second experiment in the first study display a marked symmetry between the left and right halves of the gradients. Although symmetry in the second experiment is less marked, it is noticeable. Performance on mirror images tends to be similar.

For the most part, mirror images were not used in the ETT administered at the end of the third study. However, in order to graph the results on the same kind of coordinate system used for the results of the first two studies, performance on most views was plotted twice, once for the designation of the view as it appeared in the ETT and once for the designation of its mirror image. Hence, the perfect symmetry displayed in Figures 8 and 9 is an artifact of the graphing technique.

In all three studies, performance on training views tends to be higher than performance on nontraining views. The difference, however, tends to become smaller as the number of training views increases (Table 3).

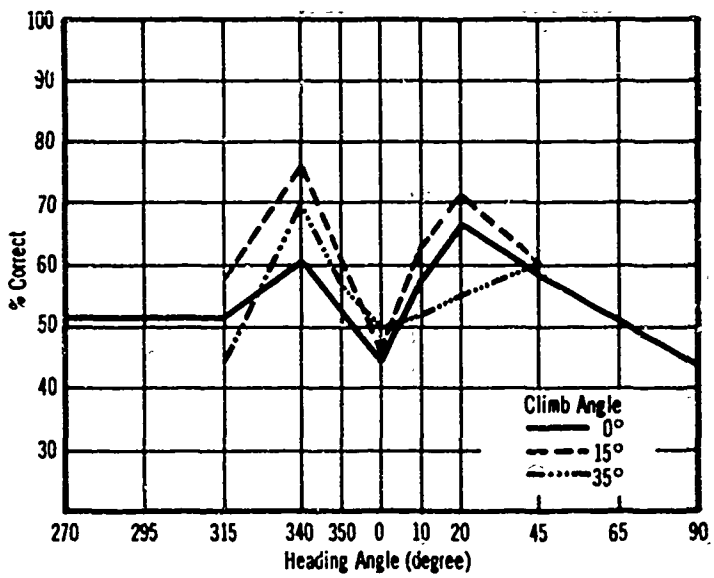
The ETT of the third study included receding views with heading angles from 135 to 225°. The two views at each extreme end of the 0° climb graph are the same as the two views at each extreme end of the 0° climb graph in Figure 8; that is, they are the full crossing views. All the receding views are nontraining views.

Since the third study was the final one in the series and included more trainees than the others, it was extensively analyzed. The basic design consisted of three levels of aircraft similarity (high, moderate, and low) and three levels of view (training, nontraining approaching, and nontraining receding). In addition, two levels of a training achievement factor (criterion and noncriterion) were added as a consequence of the fact that 20 of the 40 trainees attained the 90% achievement criterion and the remaining 20 did not attain it. The basic analysis consisted of a 2 x 3 x 3 analysis of variance with repeated measures on the last two factors. A summary of this analysis is presented in Table 4. Percent scores were used in this analysis to account for the smaller number of receding nontraining views in the ETT. All main effects and all three two-factor interactions are statistically significant at, or beyond, the .05 level.

Study I: End-of-Training Test (ETT) Performance Gradients for Training Views Used in Four Experiments

A – Experiment 1, One Training View

Training View: Heading 340°, Climb 15°



B – Experiment 2, One Training View

Training View: Heading 10°, Climb 15°

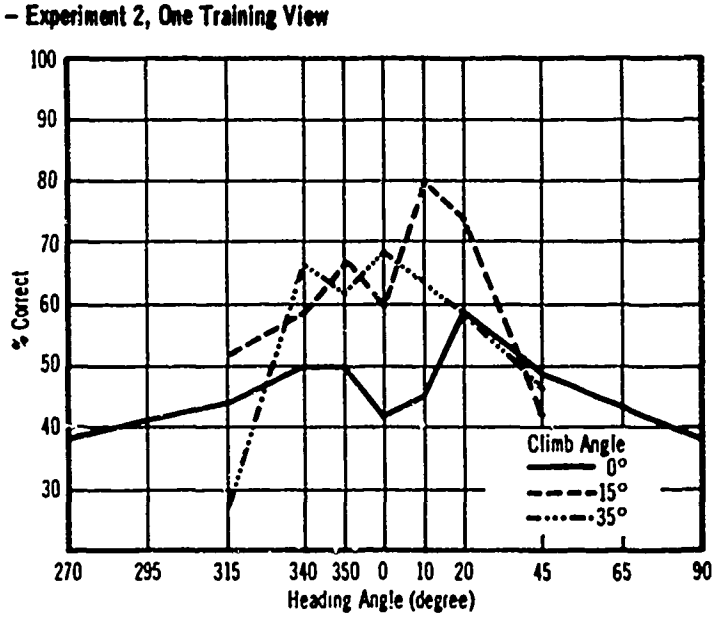
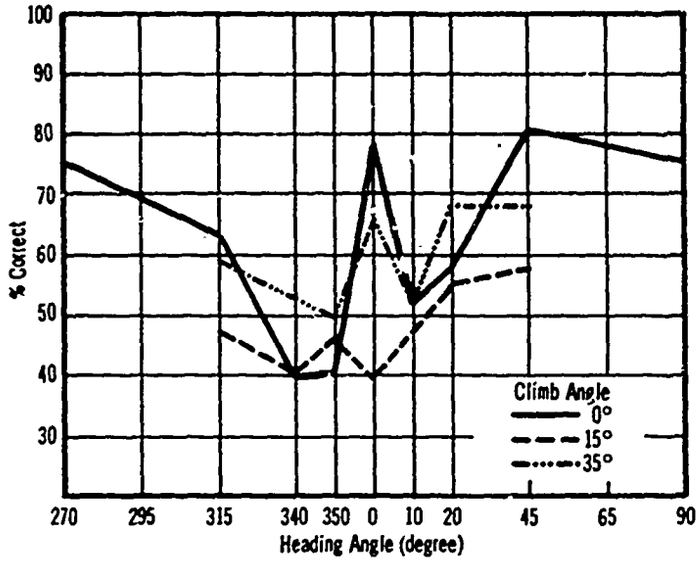


Figure 6 *(Continued)*

Study I: End-of-Training Test (ETT) Performance Gradients for Training Views Used in Four Experiments *(Continued)*

C – Experiment 3, Three Planform Training Views

Training Views: Heading 0°, Climb 0°
Heading 90°, Climb 0°
Heading 0°, Climb 90°



D – Experiment 4, Three Oblique Training Views

Training Views: Heading 10°, Climb 15°
Heading 340°, Climb 15°
Heading 45°, Climb 15°

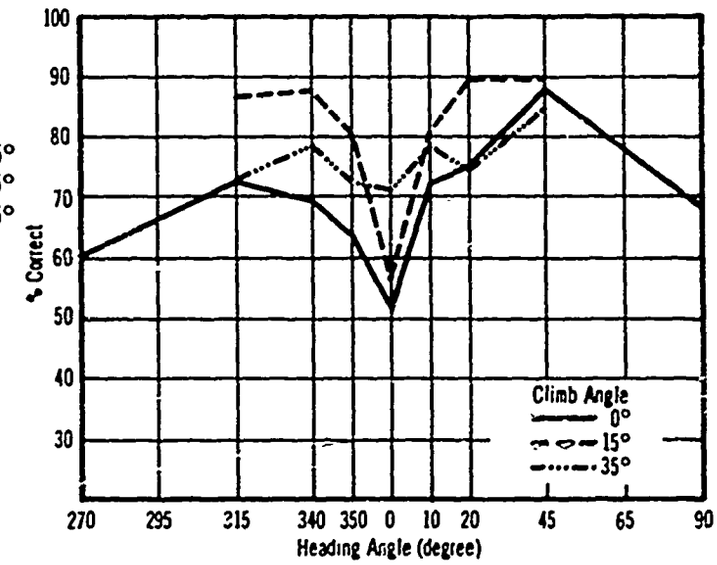


Figure 6

Study II: ETT Performance Gradients After Training on Five Views

Training Views: Heading 0°, Climb 0°
Heading 0°, Climb 90°
Heading 10°, Climb 15°
Heading 315°, Climb 15°
Heading 340°, Climb 15°

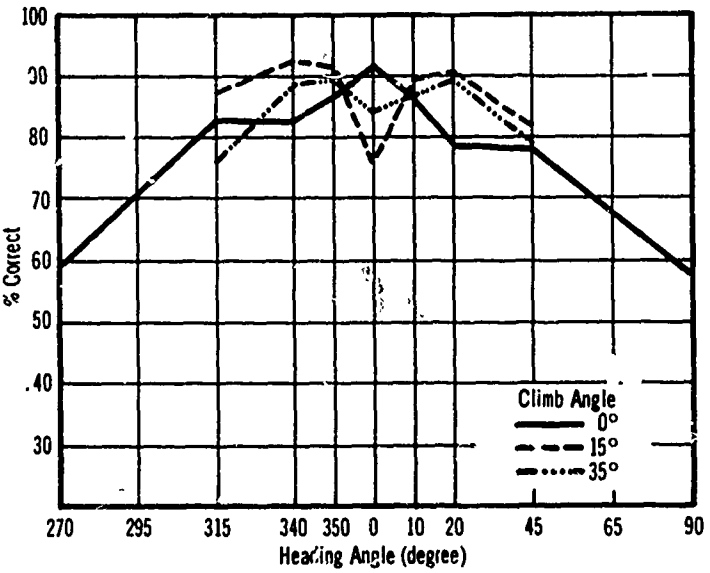


Figure 7

Study III: ETT Performance Gradients (Approaching) After Training on Nine Views

Training Views: Heading 0°, Climb 0°
Heading 20°, Climb 0°
Heading 270°, Climb 0°
Heading 0°, Climb 90°
Heading 0°, Climb 15°
Heading 10°, Climb 15°
Heading 45°, Climb 15°
Heading 340°, Climb 15°
Heading 315°, Climb 35°

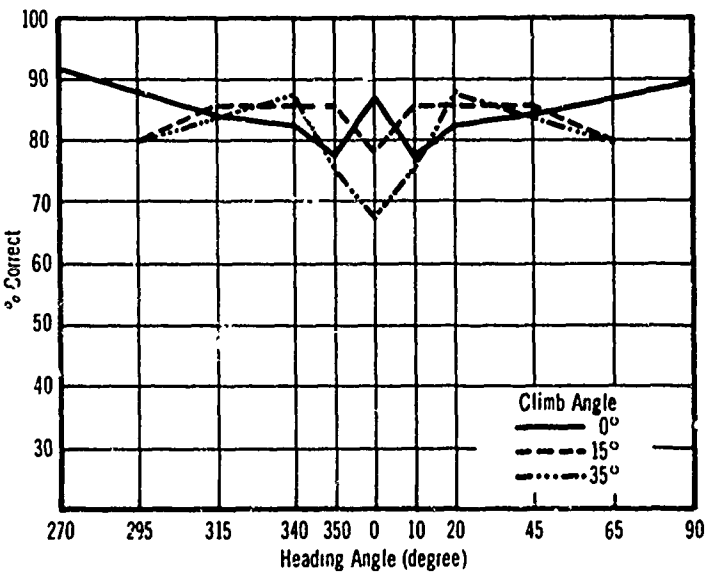


Figure 8

Study III: ETT Performance Gradients (Receding) After Training on Nine Views

Training Views: Heading 0°, Climb 0°
Heading 20°, Climb 0°
Heading 270°, Climb 0°
Heading 0°, Climb 90°
Heading 0°, Climb 15°
Heading 10°, Climb 15°
Heading 45°, Climb 15°
Heading 340°, Climb 15°
Heading 315°, Climb 15°

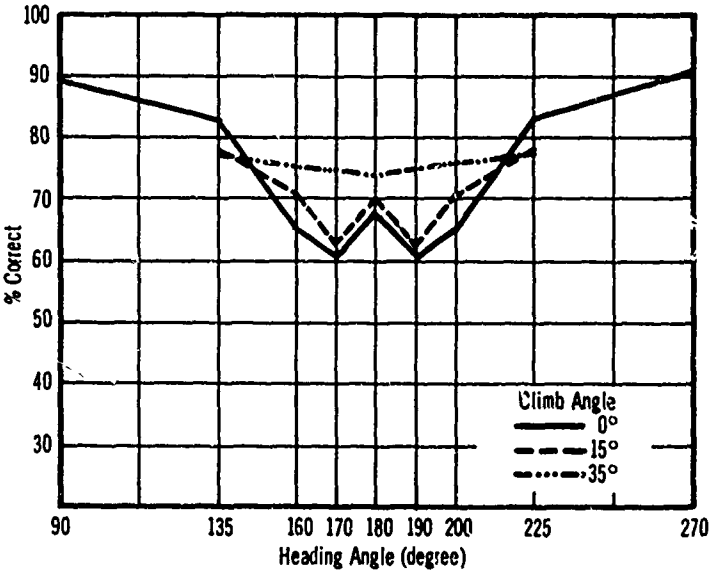


Figure 9

Table 4

Analysis of Variance for Achievement Level, Similarity, and Views for Study III

Source	*df*	Mean Square	*F*	*p*
Between Subjects				
Achievement (A)	1	17515.6	18.70	<.01
Error A	38	936.5		
Within Subjects				
Similarity (B)	2	12560.0	63.11	<.01
AB	2	776.6	3.90	<.05
Error B	76	199.0		
Views (C)	2	29098.9	359.62	<.01
AC	2	394.9	4.88	<.05
Error C	76	80.9		
BC	4	320.2	5.78	<.01
ABC	4	58.6	1.06	NS
Error BC	152	55.4		

As shown in Figure 10, ETT performance for both training and nontraining views decreases with increasing aircraft similarity. To determine whether this relationship exists during training is important. To make this determination, each trainee's last achievement test was analyzed into similarity levels. Two different achievement tests were alternated so, since different trainees in the criterion group took different amounts of time to reach the achievement criterion, the last achievement test was not the same for all trainees in this group. However, both achievement tests contained four images of one aircraft and five of the other at each similarity level. In addition, all nine training views were represented at each similarity level of each test. The means for each similarity level on the last achievement test, the corresponding means from the ETT training views, and the difference between corresponding means for each group are presented in Table 5.

It is apparent that for the criterion group the differences among similarity levels in the last achievement test are trivial. These same differences in the noncriterion group were evaluated by means of a single-factor analysis of variance of the raw scores. These differences are not statistically significant ($F = 3.07$, $df = 2,38$, NS).

The difference between the percent correct on the last achievement test and the percent correct on the training views of the ETT at each level of similarity for each

ETT Performance at Each Similarity Level for Training and Nontraining Views: Third Experiment, Study I

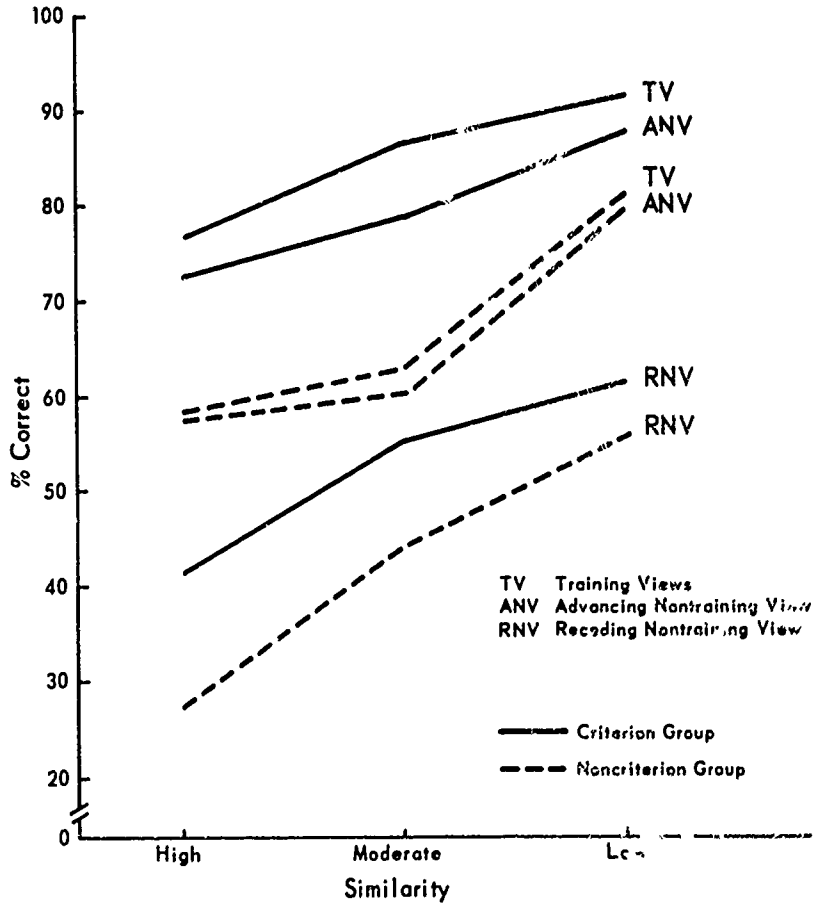


Figure 10

Table 5

Performance on Last Achievement Test and Training Views of the End-of-Training Test (ETT)
(Percent)

Group	Last Achievement Test	ETT Training Views	Difference
Criterion			
High Similarity	94.4	76.7	16.7
Moderate Similarity	95.6	86.7	8.9
Low Similarity	98.3	91.7	6.6
Noncriterion			
High Similarity	80.6	58.3	22.3
Moderate Similarity	68.9	63.1	5.8
Low Similarity	86.7	81.1	5.6

group was evaluated by means of a 2 x 3 analysis of variance with repeated measures on the second factor. A summary of this analysis is presented in Table 6. Only the F-ratio for the main effects of similarity is statistically significant ($p < .05$).

Table 6

Analysis of Variance of Differences Between Last Achievement Test and Training Views of the ETT at Each Level of Similarity for Each Group

Source	*df*	Mean Square	*F*	*p*
Between Subjects				
Groups (A)	1	0.2	--	NS
Error A	38	393.8		
Within Subjects				
Similarity (B)	2	2356.8	17.80	<.01
AB	2	147.9	1.12	NS
Error B	76	132.4		

Differences among similarity levels for the two groups combined were evaluated by means of the Newman-Keuls procedure. The drop in performance from the last test during training to the ETT is essentially the same for the moderate- and low-similarity aircraft. The drop for the high-similarity aircraft, however, is of significantly greater magnitude, being two to three times as great as the drop for the moderate- and low-similarity aircraft.

DISCUSSION

The major findings of this series of studies are exhibited in the ETT performance gradients shown in Figures 6 through 9. These gradients indicate that stimulus generalization from training to nontraining views does occur in a regular manner across both heading and climb dimensions. This is most clearly evident in the results of the first experiment of the first study (as shown in the left half of Figure 6a), the highest performance being attained on the single training view used in the experiment. Decrements in performance from this high tended to be a function of the distance of a given nontraining view from the single training view along either a heading or climb dimension. The right half of Figure 6a indicates the existence of a mirror image effect, the highest performance on these views being attained on the mirror image of the single training view; decrements in performance from this high tended to be a function of the distance of a given nontraining view from the mirror image of the single training view.

Scanning the performance gradients (Figures 6 through 8) clearly shows that the uniformity of performance is a function of the number and the distribution of the views used in training. Differences in performance on the ETT training views of the first and second studies are not statistically significant. As shown in Table 3, performance on the ETT training views of the third study falls well within the range of the first and second studies. Thus, there is no reason to believe that different sets of training views produce different levels of learning on those views. However, since differences in performance on nontraining views are statistically significant for the first and second studies (Table 3), it would appear that different sets of training views do produce different levels of generalization to other views.

It is interesting to note that generalization is not simply a function of the number of training views. The third and fourth experiments in the first study used three views in training, but produced significantly different amounts of generalization. The three training views used in the third experiment did not produce greater generalization than did the single training views used in the first and second experiments. On this basis, it would appear that the three planform views used for training in the third experiment provide a poor basis for generalization. Yet these are the three views that, historically, have most often been used when an effort was being made to restrict the number of views presented in a training program.

A comparison of the gradients obtained from the third experiment (Figure 6c) with those obtained from the fourth experiment (Figure 6d) dramatizes the extent to which the distribution of the training views can affect the shape of the gradients. Both studies used the same number of training views, but the shapes of the resulting gradients are virtual inversions of one another.

Examination of the gradients from the several studies (Figures 6 through 8) suggests that generalization is most restricted about the 0° heading–0° climb view, but improves as either heading angle or climb angle increases. Training views should be selected to satisfy either of the following two criteria:

(1) A view provides broad generalization to other views of interest.

(2) A view is operationally critical but receives little or no generalization from other views. Views of aircraft heading directly toward the observer (0° heading angle) are operationally critical, since these are the views most likely to be presented to the observer when his position is under attack. Since these views receive little generalization from other views, they will have to be densely represented in training.

A surprisingly high level of performance was obtained on the receding views included in the ETT of the third study (Figure 9). The overall receding view mean of 72% was obtained without using any receding views in training.

Similarity, clearly, had a marked effect on ETT performance. There are several possible sources for such an effect. First, it may have occurred during training, with the ETT simply displaying differences in initial achievement between similarity levels. However, the differences between similarity levels on the last achievement test given during training were not statistically significant (Table 5) within each group. Since no differences existed among similarity levels at the end of training, the differences displayed in the ETT could not have originated during training.

Second, the effect may have occurred as a differential loss in retention during the period between the end-of-training and the administration of the ETT. On the average, this period was shorter for the noncriterion group (approximately 15 minutes) than for the criterion group (approximately two hours), since individual trainees were released from training as soon as they attained the 90% achievement criterion. Generally, it would be expected that retention would be enhanced by shorter periods of elapsed time; however, the criterion and noncriterion groups show the same loss from the last achievement test to the ETT (Tables 5 and 6). Thus, it does not appear likely that the effect occurred during the delay period from training to testing.

Finally, a differential effect may have occurred only during the ETT as a result of the addition of the previously unseen nontraining views. However, a more plausible explanation presents itself in the results of the subsequent exposure duration study. This explanation is considered in the general discussion at the end of this report.

Few, if any, existing aircraft recognition training materials contain systematic sets of views of each aircraft. The views are generally selected unsystematically from existing images of real aircraft. This series of studies clearly demonstrates that training views must be selected to provide uniform generalization across the view domain of interest. Observers trained with nonsystematic materials may have serious gaps in their recognition proficiency.

Chapter 3

THE EXPOSURE DURATION STUDY

PROCEDURE

The preceding series of generalization studies showed that recognition performance tends to be poorer on similar than on dissimilar aircraft. Therefore, it seems reasonable to expect that recognition of similar aircraft would be affected more by differences in exposure duration than would recognition of dissimilar aircraft.

The same six aircraft, representing three levels of similarity, were used in this study as in the first and third generalization studies. Seven views of each aircraft were used during training:

	Heading	Climb
(1)	0°	0°
(2)	340°	15°
(3)	315°	35°
(4)	0°	90°
(5)	90°	0°
(6)	0°	35°
(7)	45°	15°

The following seven views of each aircraft were added to the training views to constitute the end-of-training test:

	Heading	Climb
(1)	315°	0°
(2)	340°	35°
(3)	340°	0°
(4)	10°	0°
(5)	0°	15°
(6)	10°	35°
(7)	10°	15°

The seven training views were selected to produce the most uniform performance possible over the entire set of 14 views. This selection was based on the results of the preceding generalization studies (Chapter 2).

Two groups of 20 enlisted men (a total of 40 trainees) were trained to recognize the seven training views of each of the six aircraft. Training proceeded in 50-minute sessions. Trainees responded as a group, orally, to each slide image presented during practice. The instructor determined how long to show each slide (about 1 to 20 seconds), and he provided feedback, orally, to the class. If the group answer was predominantly wrong, he would tell them the designation of the aircraft and review its recognition features. If he deemed it necessary, he would distinguish it from the wrong aircraft named by the trainees. This procedure does require a highly skilled instructor, but the training is more efficiently conducted.

During the last five minutes of each training session, the trainees were tested on the seven training views of each of the six aircraft. The images were presented in a different random order on each test. As soon as an individual trainee scored 90% on one of these tests, he was released from training.

At the end of training each day, all trainees were administered an end-of-training test (ETT). This test consisted of the seven training views plus the seven nontraining views of each of the six aircraft presented in random order. The trainees for each day were divided into three groups matched with respect to the session in which they achieved the 90% training criterion. Each of these groups was administered the ETT at one of three image exposure conditions: one second, three seconds, and five seconds. A five-second blank period was provided between image exposures so that all trainees would have the same amount of time in which to write their responses to each image. Each group was shown the same images in the same random order.

RESULTS

Thirty of the 40 trainees attained the 90% achievement criterion in the time available for training. Eighteen attained the criterion on the first day, and 12 attained it on the second day.

The primary analysis was accomplished by means of a 3 (Exposure) x 3 (Similarity) x 2 (Views) analysis of variance with repeated measurements on the last two factors (summarized in Table 7). The main effect of exposure duration was not statistically significant at the .05 level. The main effects of similarity and view were statistically significant. The only significant interaction is the one between similarity and view (BC).

Table 7

Analysis of Variance for Exposure Duration, Similarity, and View

Source	*df*	Mean Square	*F*	*p*
Between Subjects				
Exposure (A)	2	9.09	1.98	NS
Error A	27	4.58		
Within Subjects				
Similarity (B)	2	50.14	17.17	<.01
AB	4	5.70	1.95	NS
Error B	54	2.92		
Views (C)	1	12.79	7.66	<.01
AC	2	0.20	<1	NS
Error C	27	1.67		
BC	2	23.72	17.06	<.01
ABC	4	0.76	<1	NS
Error BC	54	1.39		

The differences among the overall means at each level of similarity were evaluated by means of the Newman-Keuls procedure. The test performance on the high-similarity aircraft (79.9%) is significantly poorer ($p < .05$) than the test performance on the moderate- and low-similarity aircraft (91.0% and 91.8%), but the latter two groups do not differ significantly from each other. Thus, the significant main effect on the similarity factor is due to the inclusion of the high-similarity aircraft on which performance was poorest.

An analysis of the simple effects underlying the similarity-by-view interaction was conducted. As shown in Table 8, simple effects at all levels of both factors were statistically significant at, or beyond, the .05 level (plotted in Figure 11). As was expected, trainees performed better on the training views than on the nontraining views on both high- and low-similarity aircraft. On the moderate-similarity aircraft, however, the relationship is inverted; that is, trainees performed better on the nontraining views than on the training views. Inspection of the total scores on each view of each of the two moderate-similarity aircraft indicates that this difference is well distributed among all the nontraining views and also among the three exposure durations.

Table 8

Analyses of Interaction Between Similarity and View

Similarity	View		*F*	*df*	*p*
	Training	Nontraining			
Low	97.4	86.2	24.89	(1,60)	<.01
Moderate	87.9	94.1	7.61	(1,60)	<.01
High	83.1	76.7	8.21	(1,60)	<.01

ETT Performance at Each Similarity Level for Training and Nontraining Views, Exposure Duration Study

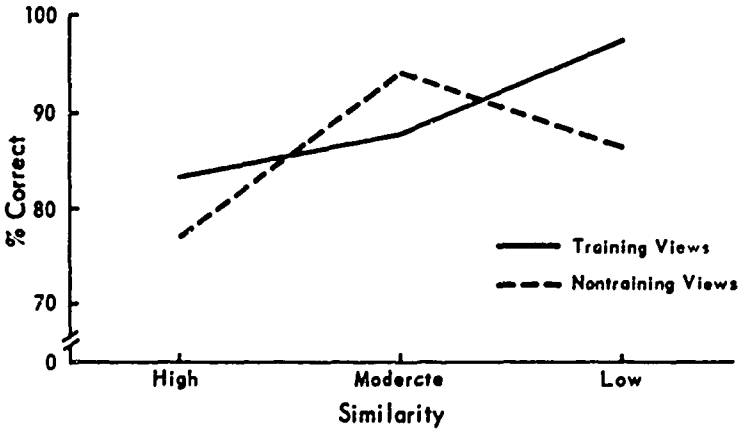


Figure 11

Overall performance on the high-similarity aircraft was significantly lower than overall performance on either the low- or moderate-similarity aircraft. There was no difference in overall performance on the latter two types of aircraft. An analysis of the simple effects underlying the similarity-by-exposure interaction was conducted (Table 9). Examination of Figure 12 suggests that performance on the low- and moderate-similarity aircraft would not improve by increasing exposure duration beyond five seconds. However, it would appear that performance on the high-similarity aircraft might continue to improve if exposure durations were increased beyond five seconds. Fortunately, the degree of similarity represented by the two high-similarity aircraft (Fishbed and Fishpot) is relatively uncommon among the aircraft of the world. Furthermore, this high degree of similarity is most likely to occur among aircraft produced in the same country. Consequently, two aircraft having this high degree of similarity are likely to be either both friendly or both hostile.

Relationships Between Similarity and Exposure Duration

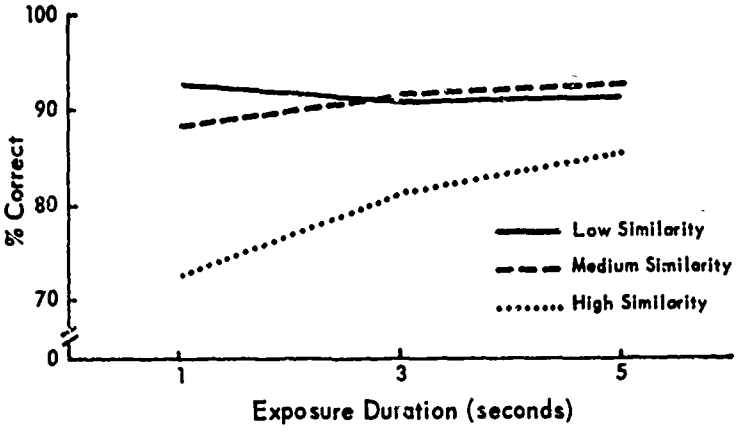


Figure 12

Table 9

Analysis of Interaction Between Exposure and Similarity

Similarity	Exposure			*F*	*df*	*p*
	1 Sec.	2 Sec.	3 Sec.			
Low	92.9	91.1	91.4	<1	(2,81)	NS
Moderate	88.2	91.8	92.9	1.34	(2,81)	NS
High	72.5	81.4	85.7	298.46	(2,81)	<.01

The relatively high performance obtained on the nontraining views of the moderate-similarity aircraft (Figure 11) is at variance with the results obtained in the nine-view generalization study (Figure 10). Both studies used the same aircraft. The difference in the results is interpreted as due to the differences in allocation of views for training and nontraining purposes between the studies; however, examination of the scores on each view of the moderate-similarity aircraft in each study does not show variations of sufficient magnitude to account for this reversal of results.

Chapter 4

STUDIES OF DIFFERENTIAL REPRESENTATION OF FRIENDLY AND HOSTILE AIRCRAFT IN TRAINING

STUDY 1

This study was conducted to evaluate the effectiveness of limiting instruction in aircraft recognition to either (a) friendly or (b) hostile aircraft. An approximately equal number of enlisted men were given recognition training on either six U.S. or six non-U.S. aircraft, but neither group of trainees were shown any other aircraft during training. Both groups were tested on all 12 aircraft at the end of training.

Procedure

One class (Group HT) of 22 enlisted men was taught to recognize the following six Soviet aircraft:

(1) Fishbed (Mig-21)
(2) Fishpot
(3) Fitter
(4) Farmer (Mig-19)
(5) Flashlight (Yak-25)
(6) Fagot (Mig-15)

A second class (Group FT) of 21 enlisted men was taught to recognize the following six American aircraft:

(1) F-102 (Delta Dagger)
(2) F-106 (Delta Dart)
(3) F-100 (Super Sabre)
(4) F-86 (Sabrejet)
(5) F-4 (Phantom)
(6) F-8 (Crusader)

To keep training time within practical limits, the same seven views of each aircraft were used during training as had been used in the exposure duration study, as follows:

	Heading	Climb
(1)	0°	0°
(2)	0°	35°
(3)	0°	90°
(4)	45°	15°
(5)	90°	0°
(6)	315°	35°
(7)	340°	15°

Thus, each class of trainees was required to learn seven views of each of six aircraft, or a total of 42 images.

Training was accomplished in 50-minute sessions. The first session began with an introduction to the six aircraft to be learned. This was followed by a paired-comparison

presentation in which views of two different aircraft, which were the most difficult to discriminate, were shown together and the critical recognition features were called to the attention of the class. Following the paired comparisons, the instructor presented each of the 42 images, one at a time, in random order. Sometimes he had individual trainees name the aircraft in each image; at other times, he had the class respond aloud as a group. In either case, if the class response was in error he told them the correct name of the aircraft and pointed out the critical recognition features. Name designations were used for Soviet aircraft, and alphanumeric designations were used for American aircraft. Image exposures were determined by the instructor and varied from a second or two to perhaps ten seconds. This procedure continued throughout subsequent sessions.

The last 10 minutes of each 50-minute session were used for achievement testing. The same 42 images were shown to the trainees, but in a different random order than used during the training session. Each image was exposed for five seconds. Trainees wrote their recognition responses on prepared answer sheets.

A 10- to 15-minute break was given between training sessions, during which time the instructor and assistant instructors scored the achievement test answer sheets from the preceding session. A trainee was released from training as soon as he scored 90% on one of the achievement tests. Training ranged from one to four sessions for different trainees.

All trainees in each group were reassembled at the end of training and administered an end-of-training test (ETT). This test contained the seven views of the six aircraft on which they were trained (1-7), plus the following seven views (8-14), which they had not seen before:

	Heading	Climb
(8)	0°	15°
(9)	10°	0°
(10)	10°	15°
(11)	10°	35°
(12)	315°	0°
(13)	340°	0°
(14)	340°	35°

This test also contained all 14 views of six aircraft which they had not seen before, or a total of 168 images (14 views of each of 12 aircraft). The images in the ETT can be schematized as follows:

		AIRCRAFT	
		Soviet	American
VIEWS	Training (1-7)	Cell A 42 images	Cell B 42 images
	Nontraining (8-14)	Cell C 42 images	Cell D 42 images

Each group was trained on only one-fourth of the images, either Cell A or Cell B. The images were presented in random order, and each image was exposed for five seconds. Trainees were instructed to identify each image as either hostile or friendly by marking the appropriate symbol on a prepared answer sheet. If they were trained to recognize Soviet aircraft, they were instructed to identify the aircraft on which they had been trained as hostile and all others as friendly. Converse instructions were given to the group trained to recognize American aircraft.

Results

The analyses are based upon the ETT results of the 16 trainees who attained the 90% achievement criterion in each group. The mean training time for Group HT was 2.5 sessions, with a standard deviation of 0.7 session. The mean training time for group FT was 2.0 sessions, with a standard deviation of 0.7 session. The difference between these means approaches statistical significance ($p < .10$).

The effects of training conditions (hostile trained vs. friendly trained), aircraft (familiar vs. unfamiliar), and views (training vs. nontraining) on ETT performance were analyzed using a 2 x 2 x 2 analysis of variance with repeated measures on two factors (summarized in Table 10). The main effect of the aircraft factor and the interaction between aircraft and training condition are statistically signficant ($p < .05$).

Table 10

Analysis of Variance for Training Conditions, Aircraft Category, and View Category

Source	*df*	Mean Square	*F*	*p*
Between Subjects				
Training Condition (A)	1	78.13	2.36	NS
Error A	30	33.13		
Within Subjects				
Aircraft Category (B)	1	1326.13	34.96	<.01
AB	1	276.12	7.28	<.05
Error B	30	37.93		
View Category (C)	1	2.00	<1	NS
AC	1	2.00	<1	NS
Error C	30	5.43		
BC	1	21.12	3.65	<.10
ABC	1	4.50	<1	NS
Error BC	30	5.78		

The average identification accuracy of each class was determined separately for each subset of aircraft and views included in the ETT. The average percentage correct for each of the four subsets of images is given in Table 11.

The specific comparisons are described below:

(1) HT versus FT training. The average percentage of correct identifications for the HT and FT classes for all aircraft was 81.0% and 77.3%, respectively. Since the difference in average accuracy was not statistically significant, it could be concluded that the training emphasis was equally effective in identifying all aircraft, irrespective of nationality.

(2) Familiar versus unfamiliar aircraft. As shown in Table 11, the FT class identified fewer unfamiliar aircraft (non-U.S. aircraft in the case of FT students) than did the HT class (U.S. aircraft, in this case). Statistical analyses (i.e., nonsignificant *F*-ratio) indicated that the two training conditions produced comparable proficiency in identifying the familiar aircraft. However, the two programs were not equally effective in producing accurate identification of the unfamiliar aircraft ($F = 18.24$, df 1,30, $p < .05$). As shown in Table 11, the HT class properly classified 76.9% of the unfamiliar aircraft as friendly (U.S.), while the FT class properly classified only 66.1% of the unfamiliar aircraft as hostile (non-U.S.). In other words, those trained on only hostile aircraft incorrectly classified 33.9% of the Soviet aircraft as friendly. Although the identification accuracy

Table 11

Identification Accuracy for Familiar and Unfamiliar Aircraft

(Percent)

Type of Training	Test Aircraft		Total
	Familiar	Unfamiliar	
Hostile Training (Non-U.S. Aircraft)	85.2 (Non-U.S.)	76.9 (U.S.)	81.0
Friendly Training (U.S. Aircraft)	88.5 (U.S.)	66.1 (Non-U.S.)	77.3

levels for unfamiliar aircraft were reliably less than for both training conditions, the decrement was significantly greater for the FT class than for the HT class.

(3) Familiar versus unfamiliar views. When the identification scores were averaged over both classes of aircraft, there were no differences between the accuracy of identifying unfamiliar and familiar views for both the familiar and unfamiliar aircraft ($F < 1.00$, $df = 1,30$, $p < .05$). For those aircraft included in training, the average accuracy scores were 87.5% and 86.1% for the familiar and unfamiliar views, respectively. For those aircraft not included in training, the corresponding average accuracies were 70.2% and 72.8%.

(4) Differences among aircraft. An additional analysis was made to compare the accuracy achieved by the two training conditions in identifying each aircraft. The percentage of correct identification for each aircraft in both training conditions, is given in Table 12.

Table 12

Mean Percent Identification Accuracy for Each Aircraft

Aircraft	Training Condition		
	Hostile	Friendly	*p*
United States			
F-86	52.7	80.7	<.01
F-100	59.8	87.5	<.01
F-102	94.2	88.4	NS
F-106	94.2	94.2	NS
F-4	95.6	91.1	NS
F-8	64.7	88.9	<.01
Non-United States			
Fishbed	77.2	57.6	<.01
Fishpot	87.5	48.6	<.01
Flashlight	81.7	89.2	NS
Fagot	89.7	75.4	<.01
Farmer	85.2	64.7	<.01
Fitter	98.7	61.1	<.01

(a) U.S. aircraft. Three of the six U.S. aircraft (the F-86, F-100, and F-8) were identified more often by the FT class than the HT class. In contrast, the identification accuracy for the two classes was essentially the same for the remaining three U.S. aircraft (F-102, F-106, and F-4).

(b) Non-U.S. aircraft. Five of the six non-U.S. aircraft were correctly identified more frequently by the HT class. Only in the case of Flashlight were the two average scores not different.

STUDY 2

Previous research (HumRRO Technical Report 68-1, January 1968) had already established that a two-category approach in which students are required to learn to differentiate equally among all aircraft, in both the friendly and hostile categories, is effective. The study described above established that single-category approach is not acceptably effective.

The training method used in Study 1 did not include simultaneous presentation of pairs of similar U.S. and non-U.S. aircraft in the same program. After Study 1 was completed, it was hypothesized that the effectiveness of the single category approach to training could be increased, or bolstered, by providing paired comparisons between similar U.S. and non-U.S. aircraft during training.

Procedure

In Study 2, the trainees were required during training to learn the designation of only the U.S. aircraft. In addition, the trainees were also given paired-comparison training between similar U.S. and non-U.S. aircraft. Two kinds of bolstered single-category training were tested. The paired-comparison training given to two classes consisted of 42 paired presentations of U.S. and non-U.S. aircraft on repeated occasions during training. For Class A, the trainees were told only that the non-U.S. aircraft was a hostile. For Class B, the trainees were also told the type designation of each of the hostiles. A third group of trainees, Class C, received paired-comparison training involving only U.S. aircraft. This control group did not observe any non-U.S. aircraft during the training. All three groups were instructed to learn the type designations of the U.S. aircraft.

Each type of training was completed within one day, including administration of the ETT. A second day of training was given to Class B, the class that had been told the type designation of the non-U.S. aircraft. On the second training day, Class B was given additional instruction on the non-U.S. aircraft. This training included presentation of 25 pairs of views of the six non-U.S. aircraft and single-image practice in recognizing these aircraft. This extra training was given to determine the amount of additional recognition accuracy that would occur when the ETT was readministered to these students.

Training proceeded in 30- to 50-minute sessions. Each session began with five to 15 minutes of paired comparisons. During paired-comparison training, the instructor displayed pairs of images of different aircraft, one pair at a time. For each pair, he stated and denoted the recognition features that differentiated between the two images, gave the designation of the friendly aircraft, and stated whatever information was appropriate for the hostile aircraft.

The paired-comparison activity was followed by 20 to 30 minutes of single-image practice on friendly aircraft only. Seven views of each aircraft were presented in random order. Sometimes the instructor would have individual trainees name the aircraft in each image. At other times he would have the class respond aloud as a group. In either case, if the class was in error he would tell them the correct name of the aircraft and point out

the critical recognition features. Image exposures were determined by the instructor, who varied them between one and ten seconds.

The last five minutes of each session were used for achievement testing. The images in each test were presented in a different order from the immediately preceding single-image practice activity. Seven views of each aircraft were used. Each image was exposed for five seconds during the test. Trainees wrote their recognition responses on prepared answer sheets.

Each trainee was released from training as soon as he scored 85% on one of the achievement tests. The achievement criterion was lowered from the level of 90% used in Study 1, because these trainees were particularly slow learners.

All trainees who attained the 85% criterion level in each group were reassembled at the end of training and administered the ETT. This test contained the seven views of each of the six friendly aircraft that had been used in training, plus seven views that had not been used in training. It also contained 14 views of each of six hostile aircraft. It was the same test used in Study 1.

Results

Ten of the 14 students assigned to Class A achieved the 85% criterion level on the hourly tests. For Class B, 11 of the 19 students made the criterion on the first training day; however, two of these 11 did not complete the ETT. Two of the remaining nine failed to appear for the second day of training given Class B on non-U.S. aircraft, but all seven remaining men achieved the 85% level on the second day. For Class C, only nine of the 14 trainees made the 85% criterion. In summary, the results for the ETT were based on the following numbers of students: ten men for Class A, nine for Class B on the first ETT, seven for Class B on the second ETT, and nine men for Class C.

The ETT was scored on a friendly or hostile basis only. Confusions among friendly aircraft were not scored as errors. The average percent correct identifications for each training condition for the friendly and hostile aircraft are presented in Table 13. The average accuracies of the three groups for friendly aircraft ranged between 86.2 and 89,7%. with no reliable variation occurring among the three training conditions. The average percent correct identifications for hostile aircraft ranged between 47.1 and 50.7. No reliable accuracy differences for hostile aircraft were evident.

Table 13

End-of-Training Test Friendly or Hostile Percent Correct for Each Class Trained on Friendly Aircraft Only

Aircraft	Class A (N=10)	Class B (N=9)	Class C (N=9)	All Classes
Friendly				
Mean	86.2	87.3	89.7	87.7
Standard Deviation	7.0	5.4	7.6	
Hostile				
Mean	50.7	47.1	50.4	49.4
Standard Deviation	10.2	9.5	9.7	

Class B was given additional training on the hostile aircraft on a second day and the ETT was readministered. Seven of the trainees achieved the 85% achievement criterion during training on both days. The average percent correct obtained on the ETT for each day and for each class of aircraft is given in Table 14. For friendly aircraft, the seven men were approximately equally accurate on both days. For the hostile aircraft, the average accuracy increased from 49.5% on the first day to 81.5% on the second day. This increment was statistically reliable ($p < .01$). The difference in average accuracy on the second day between friendly (86.2%) and hostile (81.5%) aircraft was also reliable ($p = < .05$).

Table 14

End-of-Training Test Friendly or Hostile Percent Correct for Class B at the End of Each Day of Training
(N = 7)

Aircraft	Class B	
	Day 1	Day 2
Friendly		
Mean	89.2	86.2
Standard Deviation	3.5	5.1
Hostile		
Mean	49.5	81.5
Standard Deviation	8.0	7.3

DISCUSSION

The results of these studies indicated that when only one group of aircraft (either friendly or hostile) was included in training, the accuracy of identifying the unfamiliar aircraft was significantly lower than for the familiar aircraft. This result persisted even when training on one group was bolstered by paired comparisons that included images from the other group. In addition, the results indicated that the men trained on the six U.S. aircraft had significantly lower accuracy for the six unfamiliar non-U.S. aircraft than was characteristic of the accuracy scores obtained from men trained on non-U.S. aircraft.

The average identification scores obtained over all aircraft in these studies were low relative to the accuracy levels desired for gunners. The reduced accuracy was particularly low for the unfamiliar aircraft. In the case of the FT Group in Study 1, the average identification score for non-U.S. aircraft was only 66%; that is, 34% of the non-U.S. aircraft views were incorrectly classified as U.S. aircraft by the FT Group. In Study 2, half of the non-U.S. aircraft were incorrectly classified as U.S. aircraft.

At the present time, the engagement doctrine for visually sighted air defense weapons defines two weapon control statuses:

(1) Weapon tight: The gunner engages only those aircraft that are positively identified as hostile. In the studies described here, the men receiving training on the non-U.S. aircraft would be expected to perform this task more effectively than the men trained on U.S. aircraft.

(2) Weapon free: The gunner engages aircraft *not* positively identified as friendly. In these studies, the men given the training on non-U.S. aircraft should have performed this task less effectively than those trained only on U.S. aircraft.

It is significant that neither the friendly-only nor the hostile-only training programs produced the identification accuracies needed to satisfy both the weapons tight and weapons free engagement rules. Each type of training favored either one or the other of the two rules of engagement, but neither the FT nor the HT Class was equally proficient in satisfying the requirements of both rules.

Chapter 5

GENERAL DISCUSSION OF ALL THE STUDIES

The results of the various studies support four broad generalizations regarding aircraft recognition performance and training.

First, generalization from training to nontraining views did occur in a systematic manner. Generalization tended to decrease as the distance between the training and nontraining views increased along either a heading or climb-angle dimension.

Second, the degree of similarity among aircraft is a powerful determiner of ease of recognition of each aircraft. The difficulty of recognizing a particular aircraft is largely a function of its similarity to other aircraft familiar to, and perhaps also expected by, the observer. The results of the friendly or hostile studies support this generalization.

Third, the duration of test exposure from one to five seconds does affect recognition performance, but only for the most highly similar aircraft. This did not prove to be so powerful a factor as might have been expected.

Fourth, the recognition of aircraft occurs in a relative rather than in an absolute sense; the trainee does not learn to recognize a single aircraft, he learns to discriminate among several aircraft in a set of aircraft. This was an overwhelming conclusion arising from (a) the differential effects of similarity in both the third view-generalization study and the exposure duration study, and (b) the failure to obtain adequate identification performance when only one category of aircraft (i.e., friendly or hostile) was emphasized in training.

The set of aircraft, which are germane to the conduct of the training, are not simply those selected as the objectives for a particular training program. Those aircraft that the trainee has previously learned to recognize must also be considered. For instance, the program may include the F-4, but not the A-4. The trainee would not have to choose between these two responses during training, if he knew that images of the A-4 would not be shown. However, if he had been previously trained to recognize the A-4, then he may have to make such a choice in an operational setting. In such a setting, the probability of trainee error could be intolerably large, if he had not been specifically trained to discriminate between these two moderately similar aircraft. At the very least, such specific discrimination training would require that the trainee be presented with images of both aircraft in the training criterion test.

If the aircraft recognition skills of a group of observers are to be updated by teaching them to recognize some number of new aircraft in addition to those they have previously learned to recognize, then the criterion test used for the updating training should contain not only the new aircraft, but also all previously learned aircraft that are at least moderately similar to one or more of the new aircraft. Restricting the aircraft on which the observers are tested to something less than the total number of aircraft in their recognition repertory may also restrict their opportunity for error and produce an overestimate of their recognition accuracy.

The results of the exposure duration study can be interpreted from a different point of view than that used previously. Rather than emphasizing the amount of time during which each image was available for observation, emphasis can be placed instead on the total time available for observing and responding to each image. Instead of each condition being defined by values of one, three, and five seconds, respectively, it would be defined

by values of six, eight, or 10 seconds, in terms of total time, since five seconds was uniformly provided following each image exposure. When the task is viewed in this manner, a statistically significant degradation in performance occurred only for the six-second total time condition for the highly similar aircraft.

The interpretation of the results of the exposure duration study in terms of total time can also be applied to the results of the third view-generalization study in which it was found that the drop from the last achievement test to the ETT on the high-similarity aircraft was significantly greater by a factor of two or three than the same drop for the moderate- and low-similarity aircraft. Each image was exposed for a total of five seconds in the ETT. However, no time was allowed between images. Consequently, the total time available to observe and respond to each image was also only five seconds (one second less than the total time available in the briefest condition of the exposure duration study). The achievement tests given during training provided a five-second exposure of each image plus a five-second blank between images, yielding a total time of 10 seconds to observe and respond to each image. Thus, it seems tenable that the exceptionally low performance on the high-similarity aircraft in the ETT of the third view-generalization study was due to the high degree of similarity between the two aircraft and to the restricted total time available for observing and responding to each image. Highly similar aircraft may not be so much more difficult to learn to recognize, but the act of recognizing them may require more time than the act of recognizing less similar aircraft.

Appendix A

PRODUCTION AND EVALUATION OF THE PROTOTYPE GOAR KIT

The following 18 aircraft were selected fc nclusion in the prototype kit:

(1) Fishbed	(10) F-4
(2) Fishpot	(11) F-5
(3) Fitter	(12) F-8
(4) Farmer	(13) F-100
(5) Fagot	(14) F-101
(6) Flashlight	(15) F-102
(7) A-4	(16) F-104
(8) A-5	(17) F-105
(9) A-6	(18) F-106

Aircraft models were used for producing the photographic images. Most of the models were at a 1:72 scale, but a few were at a 1:48 scale. These models were sorted into groups according to absolute size of the models. Shorter camera-to-model distances were used for the smaller models than for the larger models to compensate for differences in the model sizes. The camera-to-model distance for each model was selected to produce an image at the 90° heading - 0° climb that was about one-sixth as long as the long dimension of film format. The same camera-to-model distance was used for all views of each aircraft, but the distance was different for different group o aircraft models. This procedure minimized variations in image size so that size could not be used as an incidental recognition cue. It also resulted in an image size that was more suitable for training the recognition of aircraft at a distance; that is, it allowed the projection of much smaller images in typical Army classrooms.

It should be noted, however, that images projected on generally available screens do not adequately simulate natural world images. A projected image that subtends the same visual angle for a given screeen-to-observer distance as a natural world image at a given target-to-observer distance, presents less perceptual information to the observer than the natural world image does. There is a considerable loss of resolution on the screen (particularly beaded screens) so that the projected image is blurred in comparison to the natural image. Increasing the size of the projected image will not necessarily lead to a match with respect to perceptual information. The difficulty in discriminating some features of the aircraft is primarily dependent upon image *size*, and the difficulty in discriminating others is primarily dependent upon image *sharpness* and *internal contrasts*. Thus, an enlarged projected image may allow equal discrimination of the size and shape of an air intake, but a much easier discrimination of wing position than the smaller natural world image.

It is not yet possible to establish a direct correspondence between the characteristics of images as projected on a screen and as seen in the natural world. Consequently, it was decided to train with small projected images, recognizing that they could not be interpreted in terms of simulated target-to-observer distances in the natural world. As projected in the classroom, such images are considerably larger than their natural counterparts at a distance of 3,000 or more meters. However, they will subtend less than 7.5° as seen from the first row of trainees. In contrast, images produced by projecting

slides from the 5-QQ-8 (SLARK #1) kit in a typical classroom situation will often subtend more than 45° as seen by trainees in the first row.

In the original test of the improved classroom program,[1] two projectors were used to present each pair of images for paired-comparison training. All possible pairs were presented within each group of aircraft. In many instances, however, the configuration differences between two aircraft at a given view were so marked as to be apparent without presenting the images simultaneously. The presence of such pairs prevented paired-comparisons from being maximally efficient. Hence, it was decided to select for paired comparisons only those views of those aircraft that were most difficult to differentiate from each other. The selection was made in the following manner:

(1) Nine of the 45 views in the view matrices were selected for paired comparisons: six were selected from the approaching matrix and three from the receding matrix. Contact prints were made of each image.

(2) Ten judges were selected from the members of the research staff. Each judge sorted the images for each view into five piles of four[2] on the basis of similarity.

(3) Those pairs of aircraft that were placed in the same similarity pile for a given view by at least half the judges were selected for paired comparisons. There were 169 such pairs.

The aircraft were arranged in five groups of three to five aircraft so as to minimize the number of groups in which paired comparisons for any given aircraft might occur. The aircraft in each group were as follows:

Group 1	Group 2	Group 3	Group 4	Group 5
Fishbed	Fitter	F-4	F-101	Fagot
Fishpot	Farmer	A-4	F-104	Flashlight
F-102	F-100	F-105	F-5	A-6
F-106		F-8		
		A-5		

Ten views were set aside for use in achievement testing only. Five tests were constructed. Each test contained two views of each aircraft for a total of 36 slides per test. Within this restriction, slides were assigned randomly to the tests. The views used in the tests did not appear in any other part of the kit.

The pairs of stimulus-feedback (SF) slides were arranged into cumulative aircraft groups parallel to the paired-comparison (PC) groups. The first SF group contained the same aircraft as in the first PC group and no more; the second SF group contained the aircraft in both the first and second PC groups and no more; the third SF group contained the aircraft in the first, second, and third PC groups, and so forth. SF slide pairs were assigned to each SF group as follows:

(1) Because 10 views were reserved for testing, only 35 views were available for single-image recognition practice and review. On the average, there was one SF pair left for each of the nine paired-comparison views and two SF pairs for the remaining 26 views, for a total of 57 SF pairs for each aircraft.

(2) One SF pair for each of the 26 views of each aircraft not used in paired comparisons was assigned to the group in which the aircraft was introduced. Thus, Group 1 was assigned 104 pairs (26 views x 4 aircraft), Group 2 was assigned 78 pairs (26 views x 3 aircraft), Group 3 was assigned 130 pairs (26 views x 5 aircraft), and Group 4

[1] Paul G., Whitmore, John A., Cox, and Don J. Friel. *A Classroom Method of Training Aircraft Recognition*, HumRRO Technical Report 68-1, January 1968.

[2] The F-8 and F-84 were included at this time to give a total of 20 aircraft. They were subsequently dropped from the kit as obsolete.

was assigned 78 pairs (26 views x 3 aircraft). Two copies of each pair of non-PC views were assigned to Group 5 because it was the last group, for a total of 156 pairs (26 views x 2 copies x 3 aircraft).

(3) The remaining pairs of SF slides were randomly divided into equal piles—one pile for each subsequent group of aircraft. Thus, in theory, the remaining slide pairs from Group 1 were divided into four piles of about 35 each ([35 views ÷ 4 groups] x 4 aircraft), the remaining slide pairs from Group 2 were divided into three piles of about 35 each ([35 views ÷ 3 groups] x 3 aircraft), the remaining slide pairs from Group 3 were divided into two piles of about 90 each ([35 views ÷ 2 groups] x 5 aircraft), and the remaining slide pairs from Group 4 were left intact. One pair for each PC view was added to each Group 5 aircraft for a total of 27 pairs (9 views x 3 aircraft).

(4) In summary, slide pairs for aircraft in each PC group were assigned, in theory, to each SF group as in Table A-1. The above procedures describe the model that was devised for assigning SF slide pairs to SF groups. Actually, however, there was almost a 25% loss in the slides. On the average, there were four slides available for each view of each aircraft. Some views of some aircraft were used more than twice in the PC groups, thus depleting the number available for use in the SF groups. And some slides were exceedingly poor in quality and were discarded.

Table A-1

Assignment of Slide Pairs to Groups

	Paired-Comparison (PC)					Total
	1	2	3	4	5	
Number of Aircraft	(4)	(3)	(5)	(3)	(3)	
Stimulus-Feedback (SF) Group						
1	104					104
2	35	78				113
3	35	35	130			200
4	35	35	90	78		238
5	35	35	90	105	183	448
Tota.	244	183	310	183	183	1103
Mean/Aircraft	61	61	62	61	61	

(5) SF slide pairs were arranged in a random order within each SF group, with the restriction that pairs exhibiting either the same view or the same aircraft not be placed adjacent to each other.

A set of nomenclature familiarization slides and a set of aircraft familiarization slides were added to the kit. The former consisted of 10 slides showing different views of a fictitious aircraft—the Caped Crusader's BATPLANE—and were included to familiarize trainees with the names and locations of various aircraft structures. The latter consisted of one slide for each aircraft displaying the view of the aircraft that best displayed the structures most critical to its recognition. The aircraft familiarization slides were to be used in the first training session to introduce all the aircraft in the program.

Except for the test slides, the various sets of slides were arranged in the kit in the same order as they were to be used in training. The five sets of test slides were placed at the end of the order. The arrangement of the kit was as follows:

Nomenclature Familiarization Slides
Aircraft Familiarization Slides
Group 1 PC Slides
Group 1 SF Slides
Group 2 PC Slides
Group 2 SF Slides
Group 3 PC Slides
Group 3 SF Slides
Group 4 PC Slides
Group 4 SF Slides
Group 5 PC Slides
Group 5 SF Slides
Achievement Test No. 1 Slides
Achievement Test No. 2 Slides
Achievement Test No. 3 Slides
Achievement Test No. 4 Slides
Achievement Test No. 5 Slides

The complete kit contained approximately 2100 slides.

A manual was prepared for using the prototype GOAR kit to conduct training in accord with the improved classroom method previously developed.[1] This manual described the GOAR Slide Kit and the supplementary training materials, answer sheets, and record sheets required for conducting the training. It told the instructor how to use these materials, how to determine the appropriate image size to project on the screen, and how to conduct the training.

A troop test was conducted in May 1967 to evaluate the manual and the prototype slide kit. Two instructors, two assistant instructors, and 38 trainees were provided by Battery G, 68th Artillery at Fort Bliss. Virtually all of these men were quad-fifty gunners. They were all in their late teens or early twenties.

The test was divided into two phases. The first phase was concerned with evaluating the ability of the instructors to set up the classroom and conduct a training session as prescribed in the manual, using the manual as their sole source of information. The research staff met with the instructor teams for the first time for a half day on a Friday. Each team consisted of one instructor and one assistant instructor. At this time, the purpose of the test was explained to the instructor teams. They were provided with copies of the manual, a screen, two projectors, the printed materials required for training, the prototype GOAR Slide Kit, and approximately 30 slide trays each having a capacity of 84 slides. The instructors read the section of the manual describing the GOAR Slide Kit. The members of the research staff then helped them arrange the slides in the slide trays.

The research staff and the instructor teams met again for a half day on the following Monday. On this occasion, each of the instructors talked his way through each of the instructional procedures specified in the manual. During the Tuesday and Wednesday sessions, each instructor performed each of the instructional procedures while the rest of the group acted as a class of trainees. The performance of the instructors in these sessions provided the research staff with information regarding defects in the manual.

[1] Whitmore *et al., op. cit.*

If the instructors erred in the explanation or performance of a procedure, the description of the procedure in the manual was to be reviewed with the instructors to determine what revisions would be necessary to prevent future instructors from making the same error. Only one such error occurred. One section of the manual contained some simplified terms to be used in describing aircraft structures, and contrasted these terms to the traditional and more complex terms. Both instructors interpreted this section to imply that they should teach both sets of terms to their trainees. The intent of the manual was that only the simplified terms be used during training.

The Thursday, Friday, and Saturday sessions were spent by the instructor teams going over the slides and learning to recognize the aircraft. The training phase of the test began on the following Monday.

The 36 trainees were divided into two classes matched on GT. However, the battery baseball team unexpectedly won the Air Defense Center championship the day before the second phase of the troop test was to begin and, consequently, was scheduled to participate in the Fourth Army play-offs. Two-thirds of the men in one of the classes turned out to be members of the battery baseball team. Although the battery replaced them, there was not enough time to form two classes again matched on GT. Nineteen men were assigned to each class for a total of 38 men. Five men were subsequently dropped from each class for excessive absences. he means and standard deviations on GT for the remaining 14 men in each class were:

	N	Mean	SD
Class A	14	96.9	16.2
Class B	14	95.9	10.9

The differences between the classes with respect to either their means or standard deviations are not statistically reliable. One instructor team was assigned to each class.

In the original test of the improved classroom program,[1] 16 aircraft were taught to an average of 95% achievement in 16 training hours. However, 18 aircraft were included in the Prototype GOAR Slide Kit, but the allotted training time remained at 16 hours. Consequently, it was decided to seek an average of 90% achievement rather than 95%.

Figure A-1 shows the course of achievement for each of the two classes throughout the 16 sessions. During testing at the end of the fifth session, it was discovered that the trainees in both classes were engaged in massive cheating. The remaining tests were therefore closely monitored by the instructor and assistant instructor.

Two trainees were dropped from Class B for continued cheating. As can be seen in Figure A-1, scores dropped from the fifth to the sixth session because of monitoring. As a result of the cheating, the fourth and fifth sessions probably contributed little to the overall progress and achievement of the classes. Despite this, Class A reached 88.5% in the 16th session—just 1.5% short of the 90% criterion. Class B reached only 72.9% in the 16th session.

Examination of Figure A-1 shows that the two classes progressed similarly until the 13th session at which point Class A moved ahead and kept increasing its lead to the last session. From the 10th session to the last session, the rate of progress of Class B was sufficiently slight as to suggest that it could not achieve the 90% criterion before the 22nd session—that is, six more than were scheduled. The markedly different behavior of the two classes was induced largely by the two instructors. Class A was noisily talkative, but the talk of the class was aircraft; Class B was quiet. Both instructors were eager to teach and enjoyed the assignment, but they made different disciplinary demands upon their classes.

[1] Whitmore *et al., op. cit.*

Progression of Overall Achievement for Each Class

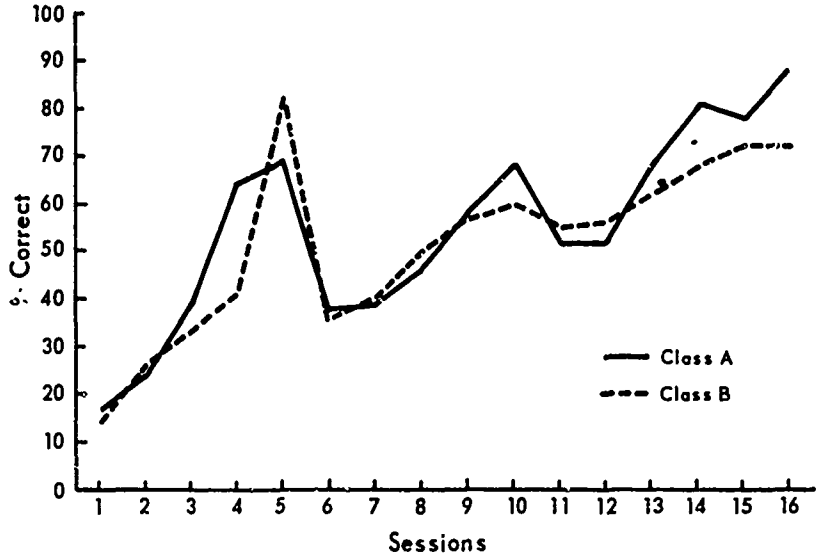


Figure A-1