DOCUMENT RESUME

ED 129 338 IR 004 125

AUTHOR Hatch, Richard S.
TITLE An Evaluation of the Effectiveness of a Self-Tutoring Approach Applied to Pilot Training.
INSTITUTION Air Force Human Resources Lab., Wright-Patterson AFB, Ohio.
REPORT NO F-1710-T-77535; WADC-TR-59-320
PUB DATE Jul 59
NOTE 25p.; Archival document

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Comparative Analysis; Conventional Instruction; Equipment Evaluation; *Flight Training; Self Pacing Machines; *Teaching Machines

ABSTRACT

One of the critical aspects of the pilot's job is the requirement for accurate and ready recall of a large body of flight information. The effectiveness of a voluntary self-tutoring device to facilitate information acquisition and retention was examined. The device was a card machine which presented and scored responses to multiple choice information questions. Two matched groups of Air Force pilots were pretested on their knowledge of instrument flying information. The device was then installed in the crew lounge of one of the groups. No device was available to the other group. After a two month period both groups were posttested. Despite the fact that minimal exposure to the machine occurred, players improved significantly on the criterion tests while non-players did not improve. Factors pertinent to the interpretation of the results and implications for further research are discussed. (KB)

WADC TECHNICAL REPORT 59-320

AN EVALUATION OF THE EFFECTIVENESS OF A SELF-TUTORING APPROACH APPLIED TO PILOT TRAINING

Richard S. Hatch, 1st Lt., USAF

JULY 1959

Project No. 1710
Task No. 77535

AERO MEDICAL LABORATORY
WRIGHT AIR DEVELOPMENT CENTER
AIR RESEARCH AND DEVELOPMENT COMMAND
UNITED STATES AIR FORCE
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

1,000 — December 1959 — 15-397

FOREWORD

The research described in this report was conducted entirely under Project 7721, Task 47020-7, held by the Operator Laboratory of the defunct Air Force Personnel and Training Research Center. The author of this report, Lt. Richard S. Hatch, had completed an essentially final draft of the report in September 1957. Due to cessation of publication activities preceding the then imminent dissolution of AFPTRC the report was not given the final processing necessary for publication, and was transferred with the files of the Operator Laboratory to the Training Psychology Branch, Aero Medical Laboratory, Wright Air Development Center.

Publication of this report was initiated at this time as an opportune contribution to the recently begun program in the automation of training under Project 1710, "Human Factors in the Design of Training Equipment", Dr. Marty R. Rockway, Project Scientist, Task 77535, "Automation of Training Systems", Mr. Felix F. Kopstein, Task Scientist.

The author wishes to express his appreciation to Mr. Irving Cohen for his efforts in initiating interest in, and subsequent support of, the project. Captain Truman D. Salyer compiled and edited the item pool in addition to contributing many valuable suggestions for the application of the technique to pilots. Dr. Paul Hood assisted in the design of the experiment and Dr. Benjamin Bernstein assisted in interpreting the analysis of the data. Lt. Reginald Young assisted in the collection and analysis of the data. Finally, the study was facilitated through the efforts and cooperation of Lt/Col J. E. Bagley and Lt/Col M. W. Rogers, Commanders of the 320th and 9th Air Refueling Squadrons, Strategic Air Command.

ABSTRACT

This study concerns the problem of insuring ready recall of a large body of in-flight job information for Air Force pilots. The effectiveness of a voluntary self-tutoring approach utilizing one type of "game appeal" device was examined. Two matched groups of Air Force pilots were pre-tested on their knowledge of instrument flying information. The device was then installed in the crew lounge of one of the groups. No device was available to the other group. After a 2-month period both groups were post-tested. Despite the fact that minimal exposure to the machine occurred, players improved significantly on the criterion tests while non-players did not improve. Factors pertinent to the interpretation of the results and implications for further research were discussed.

PUBLICATION REVIEW

This report has been reviewed and is approved.

FOR THE COMMANDER:

Walter F. Grether

WALTER F. GRETHER
Director of Operations
Aero Medical Laboratory

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

One of the more critical aspects of the pilot's job is the requirement for accurate and ready recall of a massive body of flight information. A considerable amount of factual matter is taught to pilots during their initial flying training and in subsequent transition programs throughout their careers. To illustrate the problem, a B-47 pilot is expected to know all of the information in the B-47 Flight Handbook, a 607-page volume. In addition, he must know a great many STRATEGIC AIR COMMAND STANDARD OPERATING PROCEDURES verbatim. He must know all current instrument procedures and regulations. Moreover, much of this information is undergoing continual revision so that learning of original subject matter is not sufficient. A substantial portion of the B-52 pilots' Flight Handbook is devoted to Safety of Flight revisions - and the B-52 is a comparatively new aircraft. The effort expended on the part of the pilot to retain a high degree of familiarity with his job information is reflected in his flight proficiency and accident record.

Complicating the learning problem associated with the acquisition and retention of such a massive body of information are several factors related to the nature of the subject matter itself. Much of the factual material which the pilot must have at his command is of an isolated nature; e.g., temperature ranges associated with clear icing, or required minimums associated with varying flight conditions. Consequently, most of the material does not lend itself to organization around a central concept permitting greater rapidity in the original learning situation. Retention of isolated information is correspondingly difficult.

There is also a requirement for the maintenance of seldom used job information in a state of ready recall. In flying, the most obscure fact may attain great importance under certain conditions. It may, in fact, be the one item of knowledge which makes the difference between a successful flight and disaster. It is well known that factual material which is seldom used is poorly retained, yet seldom used information makes up much that the pilot should know.

The Air Force realizes the importance of requiring pilots to maintain a high order of job-knowledge. Mandatory Annual Instrument Flying Examinations represent one program intended to insure this requirement. The average pilot studies the subject-matter once a year and is able to pass the objective-type quiz. Between the periods of examination, however, most pilots forget a good deal of this detailed information.

A final consideration concerns the techniques utilized for imparting new information to pilots. These techniques are the same as those used for reviewing pilots' knowledge of old information. The typical approach involves either a series of lectures or the issuance of lengthy and uninteresting publications. As much of the material is unrelated, little used, and often of a highly technical nature, a student motivation problem exists. Daily study, home reading, lectures and refresher courses are often unpopular with the operational pilot. In addition, most of the techniques do not require active student participation; hence, the conditions under which learning takes place are less than ideal.

APPROACH TO THE PROBLEM

One possible approach to the training problems just discussed is the use of "self-tutoring" devices. It seems likely that a practical solution to the problem of maintaining job-information in a state of ready recall lies with frequent practice, and a self-tutoring device offering some sort of game appeal may provide a painless medium for achieving a high rate of exposure to the desired learning materials. Several types of devices have been developed, including paper-and-pencil techniques as well as machines of varying complexity. In the study reported, the effectiveness of one type of self-tutoring device as a technique for improving pilot training and retention of in-flight information has been evaluated.

Description of the Self-Tutoring Device

The device employed was a card machine developed by the U. S. Navy's Training Devices Center (see Fig. 1). The device presents multiple-choice information items to the player. Scoring is based on elapsed time and accuracy of response. A score of 20 points may be attained if the button corresponding to the correct alternative is pressed within 5 seconds following the appearance of the item. For each additional second utilized thereafter, a point is deducted from the 20 points possible. No points are awarded a correct response requiring more than 25 seconds, or for an incorrect response. Following the subject's response, the correct answer to the item is exposed by the machine to give the player immediate knowledge of results. The machine cumulates the player's scores for 10 items. It is important to note that the cards fall into a drawer after each play and may be shuffled prior to being restacked, allowing a randomized presentation.

Item Selection

Over 600 instrument flying information items were compiled. An item pre-test was conducted using 104 pilots assigned to the Flying Support Squadrons at Mather Air Force Base. Item-analysis was performed to

Figure 1. Self-Tutoring Device Employed in the Study

determine difficulty level of all items. Several criteria were utilized in the rejection of items from the original item pool:

1. All items correctly answered by over 80% of the Mather item pre-test group were rejected.

2. Items judged obsolete as a result of changes in regulations were discarded.

3. Items characterized as ambiguous or poorly written by the Mather item pre-test group were discarded.

4. Lengthy items, or items with accompanying diagrams which would not lend themselves to typing in the restricted space available on the card, were discarded.

The above criteria were used in arriving at the required 320 item pool.

Criterion Test Design and Reliability

The study was designed with two objectives: the primary objective was to examine the effectiveness of a self-tutoring device as a technique for imparting job information to pilots; the secondary objective was to examine delayed retention of self-tutored information after a reasonably long period of time had elapsed.* The design called for three equivalent test forms. Two forms (A, B) were used to pre and post-test; a third form (C) was to be administered from 6 to 12 months after the post-test as the delayed retention measure. The forms were constructed, and items corresponding to all three forms were prepared for use in the machine by a lamination process (see Figure 2).

It was deemed desirable to obtain some measure of frequency of exposure to the machine and its effect upon learning and retention. One method of arriving at a measure of frequency of exposure to the machine consisted of analysis of questionnaire data. The questionnaire administered with the post-test would elicit information with respect to "remembered number of plays" over the 2-month period; however, these data would be somewhat unreliable due to error in memory on the part of the pilots. Consequently, an alternative approach to the problem of measuring exposure to the machine was devised. The plan called for a differential frequency of item exposure in the machine; that is, some items would be repeated in the machine pool while others would not. No laminations for the machine were made for one-fourth of the item pool, one lamination was made for each item in the second fourth of the pool, two laminations were made for each item in the third fourth of the pool, and three laminations were made for each item in the remaining fourth of the pool. This produced four classes of item representing a ratio of item exposure in the machine of 0: 1: 2: 3. The classes are labeled: "non-machine items", "once items", "twice items", and "three times items". Introduction of the latter three classes of items into the machine permitted manipulation of the relative frequency of machine item exposure. Play on the machine would result in a differential frequency of exposure to the three classes of machine items. Finally, the effect of this differential machine item exposure could be measured by the criterion tests employed.

In designing the experiment careful consideration was given to the control of motivational factors. Under certain conditions, including knowledge of their role and importance in an experiment, subjects may become highly motivated to succeed in the experimental situation studied.

---

* This study was terminated before data could be collected on delayed retention; consequently, only the primary objective is considered in this paper.

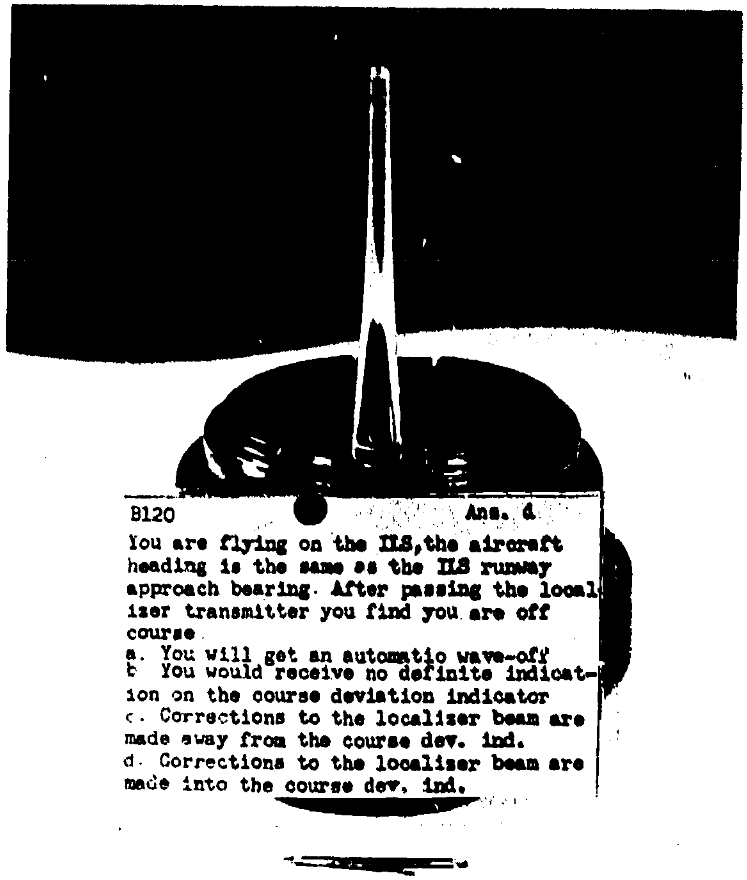

Figure 2. Typical Item Laminated for Use in the Machine

Such a phenomenon has been termed the "Hawthorne" effect after the studies carried out by Roethlisberger and Dickson at the Hawthorne Plant of the Western Electric Company (4). A "Hawthorne" effect, if it were to occur in this study, could introduce a source of spurious improvement as measured by the criterion tests. There was no feasible way to prevent the possible existence of a slight Hawthorne effect <u>during</u> the post-test, as subjects immediately recognized the items from the pre-test and from the machine. Consequently, it was necessary to "brief" them on the study at this point in order to obtain their co-operation. Increased motivation in the post-test situation could result in measured improvements independent of play on the machine. However, improvements attributable to this effect would be random with respect to the relative frequency of machine item exposure. A significant

within-group test of differential learning taking place as a function of the relative frequency of machine item exposure would substantially reduce the probability of the presence of a Hawthorne effect, and, in fact, eliminate the possibility of its being the lone contributor to the improvements measured. The improvement would stem from inordinate interest in the machine and in the learning of machine items during the experimental period. Spurious improvement due to a "Hawthorne" effect was prevented by withholding from the subjects, knowledge that an experiment was being conducted and/or that a post-test would be administered. This will be discussed in detail in a later section of this report.

The criterion forms can be considered as consisting of four scales, each scale consisting of 40 items from each of the four classes of items described above. Gains on the non-machine items scale would provide a base line against which gains on the three machine item scales might be evaluated within the experimental group. In evaluating such a comparison, however, the possible influence of the machine on the behavior of the experimental group with respect to instrument job-knowledge "in general" would have to be considered. Play on the machine might stimulate controversy over items, resulting in discussions over coffee and references to official source materials. "Incidental learning" might take place under these conditions (3). Such a phenomenon, if it were to occur, would be evidenced by a gain on the "non-machine items" scale.

The effects discussed above, combined with the characteristic "practice effect" expected in a test-retest design, dictated the necessity of a control group for the study. The groups were named the Device Group and the No Device Group. It is most probable that the No Device Group would be subject to the "practice effect" alone. A significant difference between the two groups in gains on the non-machine items scale would indicate the presence of either the "Hawthorne" or incidental learning effect, or both, in the experiment. It would then be necessary to take these effects into account in the interpretation of the results.

The assignment of items to the various scales within the three test forms and to the machine is shown in table 1. The forms were composed of 80 items common to all three forms and 80 items specific to each form. Items were assigned to the scales by a method of matched item difficulty indices. This technique assured equal means and variances with respect to item difficulty as determined by the Mather item pre-test.

Construction of the criterion test forms provided for two types of retest situation. From table 1 it may be seen that half of the items in each of the four frequency-of-exposure scales were identical from form to form; the remaining items were nonidentical form to form but of equivalent difficulty level. Greater practice effect could be expected to occur for the identical item retest situation than for the equivalent item retest

TABLE 1

ITEM ASSIGNMENT TO THE CRITERION TESTS AND MACHINE

Scale	Assignment of Items to the Criterion Tests			Assignment of Items to the Machine
	Form A	Form B	Form C	
Non Machine Items	This sub-scale identical for all forms 20			0
	20	20	20	0
Once Items	This sub-scale identical for all forms 20			20
	20	20	20	60
Twice Items	This sub-scale identical for all forms 20			20 X 2
	20	20	20	60 X 2
Three Times Items	This sub-scale identical for all forms 20			20 X 3
	20	20	20	60 X 3
TOTAL	160	160	160	

situation. Also, a higher reliability coefficient could be expected to obtain for the identical item subscales than for the equivalent item subscales. Three reliability measures were necessary to evaluate the criterion tests. Data provided by the No Device Group produced: a test-retest reliability coefficient for the 80 identical items ($r = .82$, $n = 36$), an equivalent form reliability coefficient for the 80 non-identical items ($r = .71$, $n = 36$), and a measure of form-to-form "equivalent and stability" arrived at by simply computing a product-moment correlation between total pre- and post-test scores ($r = .84$, $n = 36$). These r's were considered to be satisfactory.

Questionnaire

A questionnaire was developed for administration to the Device Group with the post-test. It was believed that attitudinal data, in addition to reported behavioral information, might provide additional insight in evaluating the effectiveness of the self-tutoring experience examined in this study.

EXPERIMENTAL PROCEDURES

Selection of the Experimental Groups

Pilots assigned to two Air Refueling Squadrons in the Strategic Air Command were utilized in this study. The Device Group was located at March Air Force Base, California. The No-Device Group was located at Mountain Home Air Force Base, Idaho. In addition to ascertaining that the assignment of pilots to the two squadrons was random with respect to factors considered relevant to the experimental variables to be examined, the squadron missions, flying duties, type of aircraft flown and training procedures were determined to be identical. Flying experience and distribution of grade did not differ significantly for the two squadrons. Subsequent analysis of the pre-test scores revealed no statistically significant difference between the two squadrons in knowledge of the items utilized in the study (see Results).

Table 2 presents information concerning the N's obtained for the study. A loss of subjects occurred due to permanent change of station, flying duty, temporary duty away from the station, leave, or other official reasons for absence during either of the testing periods. Responses on the post-test questionnaire indicated that 15 subjects in the Device Group did not play the machine. This Device Non-Player Group was treated separately from the Device Player Group in the analysis of the data.

TABLE 2

SCHEDULE OF SUBJECT PARTICIPATION
IN THE TEST ADMINISTRATIONS

SCHEDULE	DEVICE GROUP	NO DEVICE GROUP	TOTAL
Pre-Tested Form A	31	27	58
Pre-Tested Form B	31	28	59
Total Pre-Tested	62	55	117
Post-Tested Form A	28	19	47
Post-Tested Form B	30	17	47
Total Post-Tested*	58	36	94

* These figures include pilots present at both test administrations only. Twelve pilots not present at the pre-test were post-tested; however, these data were not used in the study. The questionnaire data reflects the total Device Group post-test N of 62.

Procedures

Pilots assigned to both experimental groups were pre-tested prior to the installation of the self-tutoring device. The equivalent forms (A and B) were administered alternately to all subjects. The device was then installed in the crew lounge of the Device Group for two months. No device was available to the control group. Both groups were then post-tested with the equivalent form of the test administered in the pre-test, i.e., one-half of the pilots in each group were pre-tested on Form A and post-tested on Form B, the other half of each group were pre-tested on Form B and post-tested on Form A. The questionnaire was administered to the Device Group only. Form C was not administered in this study; however, the machine items corresponding to this form comprised one-third of the machine pool.

Conditions Surrounding Voluntary Participation in the Study

Prior knowledge of any test administration might result in some form of subject preparation and it was imperative that both groups be totally unaware of their participation in the evaluation of the device. An important

aspect of the evaluation of the self-tutoring experience examined in this study concerned the nature and frequency of voluntary exposure to the device on the part of the pilots. Knowledge of an impending post-test might influence not only the frequency with which pilots would volunteer to play the machine, but also the nature of the learning situation taking place during the experience. Prior to the post-test, a form of "cramming" on the machine could develop, destroying the objective of the evaluation. The briefings delivered at the pre-test administrations were carefully prepared so as to prevent the arousal of suspicion of any kind. The pilots were told that all 15th Air Force bases would receive machines (a true statement) and that the purpose of the test was to provide data for item analysis of the large pool of items to be developed for the machines. They were told that many of the items were unsuitable, and that each should note on the test booklet those items that were poorly written, obsolete, contained no correct answer, contained more than one correct answer, etc., in addition to answering all items. The fact that 15th Air Force was actively procuring 65 machines for its 17 bases made the instructions plausible in addition to preventing the arousal of suspicion over the installation of a machine the week following the pre-test. No briefing or incentives to play of any kind, other than those supplied by the game appeal of the device itself, were operative during the study. All pilots in the Device Group were asked in the post-test questionnaire whether they suspected they would have to take a second test, and 100% of the group answered this item "No." Two individuals in the squadron were briefed on the study: the squadron commander and the sergeant assigned the responsibility of maintaining the machine and keeping a record of the total number of cards restacked daily.

RESULTS AND DISCUSSION

The error-score means and SDs for the experimental groups on the criterion tests are shown in table 3. The test of the difference between the raw error-score means of the Device Player Group and the No-Device Group on the pre-test was not significant ($t = .971$, $df = .78$). Thus, there appears to be no bias in the selection of the experimental groups with respect to knowledge of instrument flying information as measured by the criterion tests. The difference between the pre-test error-score and post-test error-score was derived for each subject in the study. The resulting gain-scores, used in all further analysis of the data, were independent of initial difference levels on the pre-test.

TABLE 3

ERROR-SCORE MEANS AND STANDARD DEVIATIONS FOR THE EXPERIMENTAL GROUPS ON THE CRITERION TESTS

Group	Pre-Test			Post-Test	
	N	M	SD	M	SD
Device Players Group	43	60.00	11.43	53.27	12.61
Device Non-Player Group	15	58.20	13.10	57.46	13.30
No Device Group	36	62.80	13.41	62.33	15.20

TABLE 4

MEANS AND SD's OF GAIN-SCORES FOR SUBGROUPS
IN THE ANALYSIS OF VARIANCE

	Device Player Group				No-Device Group			
	Identical Item Mean	Identical Item SD	Equivalent Item Mean	Equivalent Item SD	Identical Item Mean	Identical Item SD	Equivalent Item Mean	Equivalent Item SD
Non-Machine Items	.74	2.58	-.14	3.03	0.58	2.47	-.28	2.76
Once Items	1.07	2.41	0.60	2.65	1.39	2.97	-.78	2.70
Twice Items	1.28	2.70	0.65	2.78	-.44	2.00	0.06	3.39
Three Times Items	1.44	2.66	1.12	2.67	0.53	1.93	-.19	2.74

TABLE 5

ANALYSIS OF VARIANCE OF
GAIN-SCORES ON THE CRITERION TESTS

Source of Variance	df	Sums of Squares	Mean Squares	F	P
Between Subjects	78	847.19			
Experimental Groups (E)	1	85.45	85.45	8.64	$<.01$
error (b)	77	761.75	9.89		
Within Subjects	553	3844.75			
Frequency (F)	3	27.42	9.14	1.23	
Practice Effect (P)*	1	73.82	73.82	10.71	.01
F x P Interaction	3	25.12	8.37	1.37	
F x E Interaction	3	25.15	8.38	1.13	
P x E Interaction	1	2.20	2.20	.32	
F x P x E Interaction	3	43.13	14.38	2.36	
Error (w)	539	3647.92	6.77		
$error_1$	77	530.67	6.89		
$error_2$	231	1711.63	7.41		
$error_3$	231	1405.61	6.09		
TOTAL	631	4691.94			

* This effect refers to the difference between gains due to identical-item subscales and gains due to equivalent-item subscales.

Since this experiment corresponded to complex design of the "mixed" type described by Lindquist (2, p. 292), his suggestions for appropriate analysis of variance procedures were followed in comparing the gains of the Device Player and No-Device Groups. Because there were 8 gain-scores for each of the 79 subjects (representing each scale frequency for identical and equivalent subscales), there is a total of 632 scores in the analysis. Table 4 summarizes the design and presents the means and SD's of the subgroups; and table 5 presents the analysis of variance with the appropriate error terms.

The Experimental Groups main effect tested by the Between-Subjects error term yielded an F-ratio significant beyond the .01 level of confidence ($F = 8.64$, $df = 1$ and 77). The results of this test support the major hypothesis that an improvement in knowledge of instrument flying information would result from the use of a self-tutoring device.

All other tests of significance were part of the within-subjects analysis. The following tests were made: Frequency and P x E against $error_1$ (w), Practice and Effect and P x E against $error_2$ (w), and F x P and F x P x E, against $error_3$ (w). The only significant within-subjects comparison was that between gains attributable to identical item subscales and gains attributable to equivalent item subscales. Practice Effect mean square tested by the within-subject $error_2$ term was significant at the .01 level ($F = 10.71$, $df = 3$ and 77).

Trend Analysis

Because frequency of item exposure is a "reguler increasing" independent variable, a more appropriate analysis of this variable is provided by a trend test. The technique utilized, following Lindquist's discussion (2, p. 342), tests for the presence of a significant within-group trend over the extremes of the expected sequence. Using this technique, a matched t test of the difference between mean gains achieved on the non-machine items scale and mean gains achieved on the "three times items" scale by the Device Player Group was significant beyond the .01 level (matched $t = 3.06$, $df = 85$). Figure 3 presents the mean gains on the various exposure frequency scales for three groups. The Device Non-Player Group was not included in the analysis of variance but is presented in Figure 3 and in the section on questionnaire results for comparative purposes. It may be safely inferred from the trend test reported, and from inspection of the increasing monotonic sequence obtained for the Device Group, that learning took place as a function of the relative frequency of machine item exposure. Further, figure 3 shows that no significant differences in improvement on the "non-machine items" scale resulted for the experimental groups. This comparison may be interpreted as a test for the presence of uncontrolled effects. No significant Hawthorne or incidental learning effects were demonstrated by this comparison; thus the possibility of any bias in the results due to these effects may be dismissed for lack of evidence.

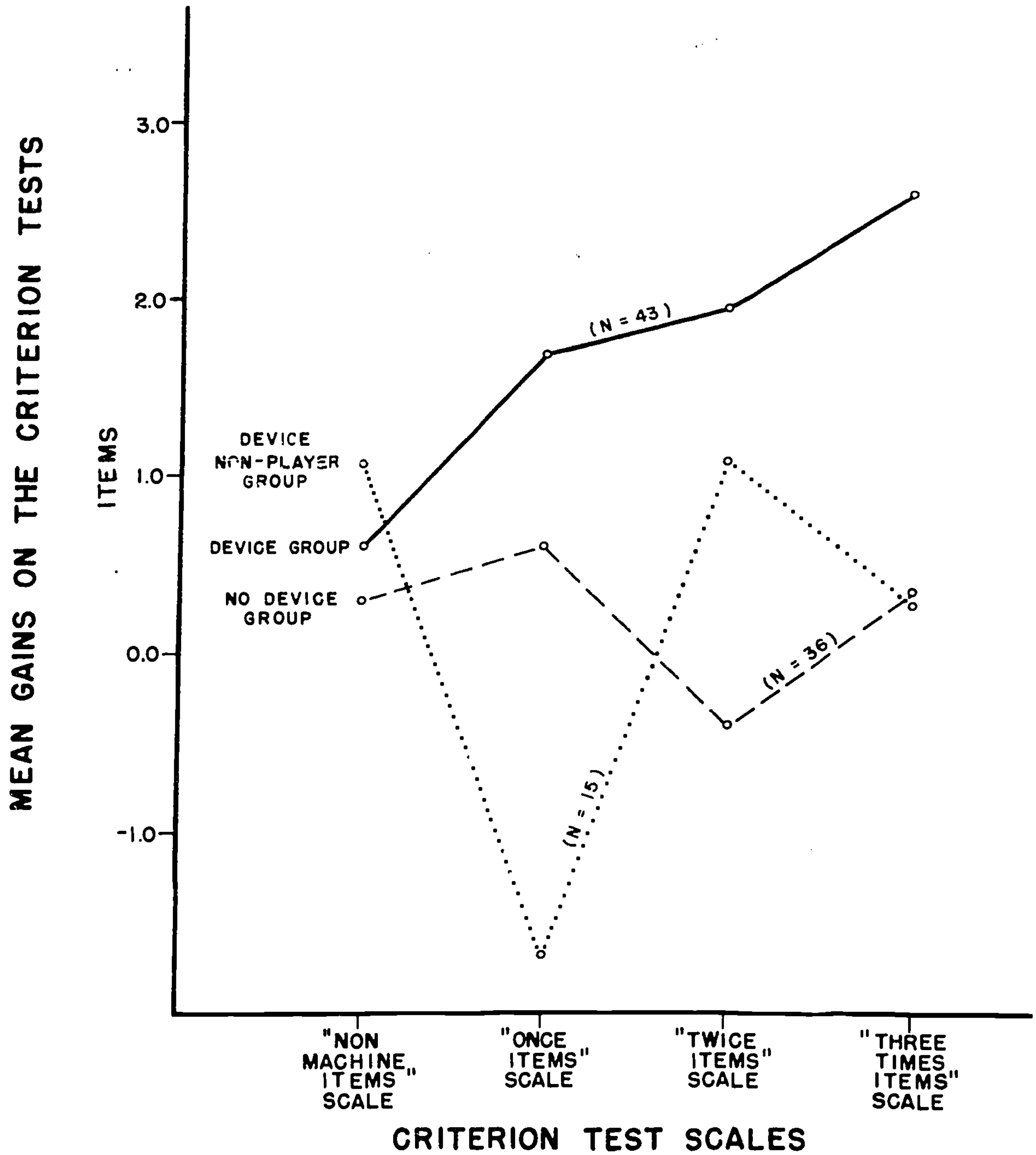

Figure 3. Relationship between Item Exposure on the Machine and Group Performance

Questionnaire Results

The most important finding from questionnaire data is depicted by a bar diagram (see figure 4). The reported frequency of play on the

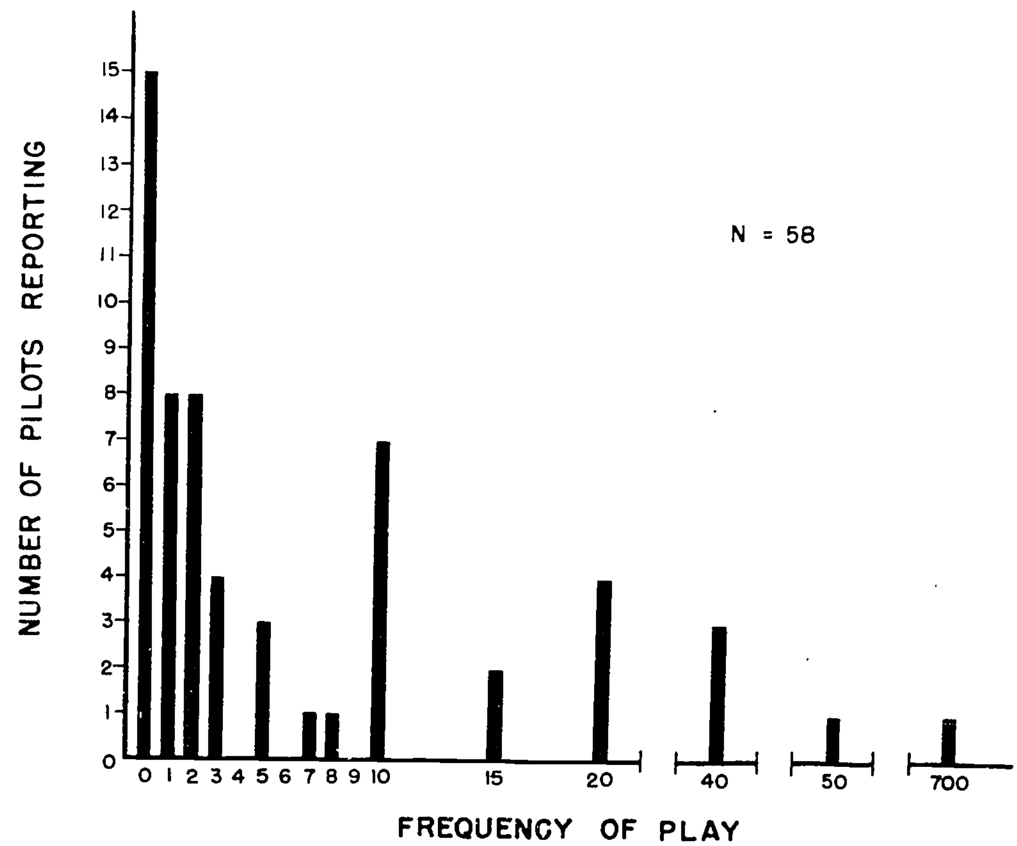

Figure 4. Frequency of Reported Play on the Self-Tutoring Device

machine should be noted. Total reported play was checked against the actual number of items restacked during the study. Although airmen and non-pilot officers were observed to have played the machine at least as often as the pilots assigned to the study, the play "reported" by the Device Player Group accounts for about 95% of the total play recorded. From this observation, the information in figure 4 would appear to be exaggerated and should be discounted considerably. The amount of play taking place during the study was small by any standards. The average Device Player Group pilot played less than 30 minutes during the entire 2-month period.

The investigator did not consider the questionnaire findings to be of sufficient importance for detailed examination in this report; however, a summary of results is presented in table 6.

TABLE 6

SUMMARY OF QUESTIONNAIRE RESULTS *

	Items (Condensed)		Device Players (n=46) %	Device Non-Players (n=16) %	Device Group Total (N=62) %
5.	"By watching others play... able to read additional items."	Yes	63	31	55
		No	33	50	37
8.	"Controversial nature of some items resulted in:"				
	a. arguments with other pilots		.54	19	45
	b. references to official sources		39	13	32
	c. loss of interest in machine		7	13	8
	d. other		9	0	6
10.	"Greatest incentive to play:"				
	a. gambling opportunity		0	0	0
	b. improve self on items		37	31	35
	c. challenge of items themselves		50	19	42
	d. nothing better to do		17	13	16
11.	Would have played more if items more interesting:	Yes	74	--	--
		No	22	--	--
12.	"Other pilots attitudes toward machine were:"				
	a. extremely favorable		4	13	6
	b. very favorable		48	19	40
	c. neutral		44	25	39
	d. negative		2	0	1
	e. felt it a waste of time		0	0	0

TABLE 6 (continued)

Items (Condensed)		Device Players (n=46) %	Device Non-Players (n=16) %	Device Group Total (N=62) %
14. "I personally:"				
a. learned new facts from machine		24	--	--
b. boned up on...forgotten facts		67	--	--
c. derived no benefit from machine		20	--	--
15. "SAC should have a machine in each squadron."	Yes	80	50	72
	No	17	25	19
17. "I was able to play as often as I liked."	Yes	46	31	42
	No	52	44	50

* Percentages are based on an N of 62 as data from four pilots available for the post-test only were included. Percentages do not add to 100 due to omissions or responses to more than one alternative for some items.

CONCLUSIONS

The self-tutoring experience examined in this study resulted in statistically significant gains in job-information for pilots voluntarily playing the device. The statistical assurance that success of the self-tutoring approach applied to pilot training and retention by them was not accidental constitutes an important finding. The fact that improvements resulting from the application of the self-tutoring technique employed in this particular study were not large in no way detracts from it, as conditions affecting pilot participation on the device were far from ideal. The crew lounge, in addition to being undesirably situated, was unpopular with the pilots. Many pilots did not visit this lounge more than once or twice during the entire experimental period. No external incentives of any kind, such as contests or briefings, were employed to promote interest in the device. Probably the most severe handicap to a fair evaluation of the self-tutoring technique examined involved the

specific device employed. The device required almost continual maintenance due to malfunctions directly attributable to the ineffectiveness of its design. As it was impossible to arrange for full-time monitorship of the device, it was out of order for a considerable amount of time, which discouraged play.

The stimulus situation presented to the player by the device, when operating correctly, was considered challenging and appropriate in the self-tutoring experience examined. It should be relatively easy to design a device, incorporating the functions and characteristics desired, with mechanical and electrical reliability vastly superior to that of the device employed.* It is believed that a "troublefree" device of this type, in the presence of play-motivating incentives over and above those supplied by the game appeal of the device itself would produce worthwhile effects in practice. The self-tutoring approach to pilot training in, and retention of, in-flight job information appears profitable; the questions remaining concern methods of employing the technique to the greatest advantage.

---

* In the course of the study a highly reliable device was designed for construction. This device would employ a 16mm projector in place of the laminated card presentation. The major source of malfunction in the Navy device was inability of the equipment to handle the cards. They warped, split and were otherwise damaged by the mechanical knife used to transport them within the machine.

REFERENCES

1. Johnson, P. O., Statistical Methods in Research, New York: Prentice Hall, Inc., 1950

2. Lindquist, E. F., Design and Analysis of Experiments in Psychology and Education, New York: Houghton Mifflin Company, 1953

3. McGeoch, John A., The Psychology of Human Learning, New York: Longmans, Green and Co., Inc., 1946 (pp. 299-308)

4. Ryan, Thomas A., Work and Effort, The Psychology of Production, New York: The Ronald Press Co., 1947 (pp. 179-82)